



Evaluation of the Efficiency of Artificial Neural Network and Random Forest Models in Predicting Groundwater Quality Parameters of Yazd-Ardakan Plain

Hassan Fathizad^{1*} and Mohammad Ali Hakimzadeh Ardakani²

1- Ph.D, Combating Desertification, Department of Management Arid and Desert Regions, Faculty of Natural Resources and Desert, Yazd University, Yazd, Iran

2- Associate Professor, Department of Arid and Desert Regions Management, Faculty of Natural Resources and Desert, Yazd University, Iran

*Corresponding author: h.fathizad@stu.yazd.ac.ir

(Received: 15 December 2021

Revise: 8 February 2022

Accepted: 10 February 2022)

Extended Abstract

Introduction: Groundwater resources are one of the main natural resources that support socio-economic development of countries (Siebert *et al.*, 2010). Management and protection of groundwater resources is of great importance in countries such as Iran, which are located in arid and semi-arid regions without surface water resources. Therefore, resource conservation, in addition to short-term and long-term planning, is essential to optimizing their productivity (Abu-Khalaf *et al.*, 2013). These natural resources face a variety of issues that threaten their sustainability, such as the effects of climate change, human activities and natural processes (Alabjah *et al.*, 2018; Baghvand *et al.*, 2010; Burri *et al.*, 2019; El Asri *et al.*, 2019; Houemenou *et al.*, 2020; Mountadar *et al.*, 2018). Considering the great importance of recognizing the quality characteristics of groundwater in desert areas and the need for its proper management, predicting groundwater quality for the management and exploitation of water resources in the Yazd-Ardakan plain seems urgent. Therefore, the aim of this study was to evaluate and compare the efficiency of artificial neural network and random forest models in predicting EC, SAR, SO₄⁻ and TDS values. In this research, modeling will be done based on the relationship between environmental (auxiliary) data and groundwater quality parameters.

Materials and Methods: The study area is spread over 482900 ha of land, which in terms of position is 53° 08' 36" and 54° 85' 32" E longitude, is located in the central plateau of Iran and the central part of Yazd province. The maximum height of the area is 2677 m and the minimum height is 997 m from the sea level. To evaluate the quality of groundwater sample, 201 wells of the monitoring network of Yazd Regional Water Company in 2016, the measured parameters EC, SAR, SO₄⁻ and TDS were used. The factors of environmental (auxiliary) data in this research include geological data, land use, vegetation indices, soil salinity indices, derivation of digital elevation model, distance from mines, distance from road, distance from river, distance from residential areas, rainfall and population.

After preparing the groundwater quality parameters and environmental data, the values of EC, SAR, SO₄⁻ and TDS were predicted using artificial neural network and random forest models. In order to validate the random forest model, the cross-validation method (10-fold) was used in R statistical software. In this method, the data is divided into 10 parts. 9 parts of the data are used for modeling and the remaining part is used to validate the obtained model. MAE, RMSE and R² statistical indices were used to evaluate the efficiency of artificial neural network and random forest models for groundwater parameters.

Results and Discussion: In this research, in simulating EC, SAR, SO₄⁻ and TDS parameters, the best structure obtained from 100 repetitions of artificial neural network learning has 2 hidden layers and 10 hidden neurons in each layer. The results of random forest sensitivity analysis show the high importance of the parameters extracted from the DEM. The use of satellite image data to investigate the groundwater quality parameters is also a convenient and cost-effective method. Furthermore, combining satellite data with DEM to investigate groundwater quality parameters and their zoning makes the results more efficient and increases its accuracy. A very important point in the analysis of parameters is the important role of parameters such as distance from the road and distance from the mines. This indicates the direct effect of man-made environmental factors on the groundwater quality of the study area. The results of modeling the EC, SAR, SO₄⁻ and TDS parameters using artificial neural network and random forest models show the appropriate accuracy of the artificial neural network model in predicting these parameters. Accordingly, the neural network model has predicted the parameters of EC, SAR, SO₄⁻ and TDS of groundwater with coefficient of determination of 0.82, 0.92, 0.92 and 0.97, respectively, while the model Stochastic forest has predicted EC, SAR, SO₄⁻ and TDS parameters of groundwater with coefficients of determination of 0.37, 0.65, 0.80 and 0.68, respectively. Based on the obtained results, it can be concluded that the artificial neural network model has a higher accuracy than the random forest model in predicting groundwater quality parameters. EC, SAR, SO₄⁻ and TDS parameters have the highest values in the north, center and southwest, and the lowest values in the southeast and southwest.

Conclusion: It can be said that the technique is highly efficient for estimating EC, SAR, SO₄⁻ and TDS parameters in Yazd-Ardakan plain as long as a proper and a sufficient number of input elements, a proper and compatible artificial neural network, and an appropriate calibration are used.

Keywords: Calibration, Coefficient of determination, Environmental data, Protection and planning, Yazd-Ardakan plain.

Citation: Fathizad, H. & Hakimzadeh Ardakani, M. A. (2022). Evaluation of the efficiency of artificial neural network and random forest models in predicting groundwater quality parameters of Yazd-Ardakan plain. *Integrated Watershed Management*, 1(2), 1-19. doi: 10.22034/iwm.2022.250822

Copyrights:

Copyright for this article is retained by the author(s), with publication rights granted to Integrated Watershed Management. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>).



بررسی کارایی مدل‌های شبکه عصبی مصنوعی و جنگل تصادفی در پیش‌بینی پارامترهای کیفی آب زیرزمینی دشت یزد- اردکان

حسن فتحی‌زاد^{۱*} و محمد علی حکیم‌زاده اردکانی^۲

- ۱- دانش‌آموخته دکتری بیابان‌زدایی، دانشکده منابع طبیعی و کویرشناسی، دانشگاه یزد، یزد، ایران
۲- دانشیار گروه مدیریت مناطق خشک و بیابانی، دانشکده منابع طبیعی و کویرشناسی، دانشگاه یزد، یزد، ایران

*نویسنده مسئول: h.fathizad@stu.yazd.ac.ir

تاریخ پذیرش: ۱۴۰۰/۱۱/۲۱

تاریخ بازنگری: ۱۴۰۰/۱۱/۱۹

تاریخ دریافت: ۱۴۰۰/۰۹/۲۴

چکیده

آب‌های زیرزمینی یکی از مهمترین منابع آبی هستند که مطالعه کیفی آن‌ها برای حفاظت و برنامه‌ریزی منابع آب بسیار مهم است. همچنین اندازه‌گیری پارامترهای کیفی آب زیرزمینی بر هزینه و زمان بر است. مدل‌ها هزینه برآوردهای کیفی آب را کاهش داده و بانک اطلاعاتی برای مدیریت منابع آب ارائه می‌دهند. در این پژوهش کارایی مدل‌های شبکه عصبی مصنوعی و جنگل تصادفی در پیش‌بینی مقادیر هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول آب زیرزمینی مورد بررسی قرار گرفته است. برای بررسی کیفی نمونه‌ی آب زیرزمینی، ۲۰۱ حلقه چاه (داده‌های وابسته) شبکه پایش شرکت آب منطقه‌ی استان یزد در سال ۱۳۹۵ و داده‌های محیطی (داده‌های مستقل) شامل: داده‌های زمین‌شناسی، کاربری اراضی، شاخص‌های پوشش گیاهی، شاخص‌های شوری خاک، مشتقات حاصل از مدل رقومی ارتفاعی، فاصله از معادن، فاصله از جاده، فاصله از رودخانه، فاصله از نقاط مسکونی، بارندگی و جمعیت استفاده شده است. نتایج مدل‌سازی پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول با استفاده از مدل‌های شبکه عصبی مصنوعی و جنگل تصادفی نشان‌دهنده دقت مناسب مدل شبکه عصبی مصنوعی در پیش‌بینی این پارامترها است به طوری که نقشه نهایی پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول آب زیرزمینی را به ترتیب با ضریب تبیین ۰/۸۲، ۰/۹۲، ۰/۹۲ و ۰/۹۷ و پیش‌بینی کرده است و این در حالی است که مدل جنگل تصادفی پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول آب زیرزمینی را به ترتیب با ضریب تبیین ۰/۳۷، ۰/۶۵، ۰/۸۰ و ۰/۶۸ پیش‌بینی نموده است. پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول در شمال، مرکز و جنوب‌غربی دارای بالاترین مقدار و در جنوب شرقی و جنوب‌غربی دارای پایین‌ترین مقدار می‌باشند. با انتخاب نوع و تعداد مناسب عوامل ورودی و نیز استفاده از نوع مناسب و سازگار شبکه عصبی مصنوعی و نیز کالیبره کردن مناسب آن می‌توان گفت که این تکنیک ابزاری بسیار کارا و مناسب برای برآورد پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول در دشت یزد- اردکان می‌باشد.

واژه‌های کلیدی: حفاظت و برنامه‌ریزی، داده‌های محیطی، دشت یزد- اردکان، ضریب تبیین، کالیبره.

استناد: فتحی‌زاد، ح. و حکیم‌زاده اردکانی، م. ع. (۱۴۰۰). بررسی کارایی مدل‌های شبکه عصبی مصنوعی و جنگل تصادفی در پیش‌بینی پارامترهای کیفی آب زیرزمینی دشت یزد- اردکان. مدیریت جامع حوزه‌های آبخیز، ۱(۲)، ۱۹-۱.

حق چاپ:



حق چاپ برای نویسنده (گان) این مقاله محفوظ است. بر اساس قوانین انتشارات با دسترسی آزاد، تمام مطالعات چاپ شده در این مجله به صورت آزاد در وب سایت مجله برای عموم بدون پرداخت هزینه قابل دسترس است.

مقدمه

al., 2020; Tyagi & Sharma, 2013; Wu *et al.*, 2020).

در طول دو دهه اخیر، کاربرد تکنیک‌های هوش مصنوعی، ماشین یادگیری و داده‌کاوی در بسیاری از زمینه‌ها افزایش یافته است. به‌عنوان مثال، در زمینه‌های پیش‌بینی هیدرولوژیکی، مدل‌های ماشین یادگیری در پیش‌بینی جریان رودخانه بسیار دقیق هستند (Yaseen *et al.*, 2019). برای مدل‌سازی آب‌های زیرزمینی، تکنیک‌های هوش مصنوعی با استفاده از مدل‌های مختلف ماشین یادگیری در پیش‌بینی سطح آب‌های زیرزمینی، قوی و رضایت‌بخش هستند (Rajaei *et al.*, 2019).

در مورد پیش‌بینی کیفیت آب، مطالعات متعددی با استفاده از تکنیک‌های هوش مصنوعی انجام شده است که می‌توان به Tung و Yaseen (۲۰۲۰) اشاره کرد. Lu و Ma (۲۰۲۰) از مدل‌های تقویت گرادیان شدید و جنگل تصادفی برای پیش‌بینی شش شاخص کیفیت آب در رودخانه Tualatin استفاده کردند. Castrillo و Garcia (۲۰۲۰) از مدل‌های خطی و جنگل تصادفی برای تخمین غلظت بالای عناصر غذایی در رودخانه Thames استفاده کردند. Bilali و همکاران (۲۰۲۰) از مدل‌های شبکه عصبی مصنوعی برای پیش‌بینی کیفیت شیمیایی آب زیرزمینی برای اهداف آشامیدنی استفاده کردند. Fijani و همکاران (۲۰۱۹) روش‌های هوش مصنوعی را برای پشتیبانی از نظارت بر کیفیت آب مخزن در زمان واقعی طراحی و اجرا نمودند. Di و همکاران (۲۰۱۹) مدل‌های یادگیری ماشین طبقه‌بندی را برای ارزیابی کیفیت آب در رودخانه Yangtze در چین اعمال کردند. Chou و همکاران (۲۰۱۸) با استفاده از مدل‌های یادگیری ماشین، شاخص وضعیت تغذیه‌ای (CTSI)^۱ را تعیین کردند. Meyers و همکاران (۲۰۱۷) از مدل‌های شبکه عصبی مصنوعی، ماشین بردار پشتیبان و جنگل تصادفی برای پیش‌بینی کدورت آب در بریتانیا استفاده کردند. Wagh و همکاران (۲۰۱۶) از

منابع آب زیرزمینی یکی از اصلی‌ترین منابع طبیعی است که از توسعه اجتماعی-اقتصادی کشورها حمایت می‌کند (Siebert *et al.*, 2010). مدیریت و حفاظت از منابع آب زیرزمینی در کشورهایی مانند ایران که در مناطق خشک و نیمه‌خشک فاقد منابع آب سطحی قرار دارند، از اهمیت بالایی برخوردار است. بنابراین، حفظ و نگهداری منابع علاوه بر برنامه‌ریزی کوتاه‌مدت و بلندمدت، برای بهینه‌سازی بهره‌وری آن‌ها ضروری است (Abu-Khalaf *et al.*, 2013). این منابع طبیعی با مسائل مختلفی مانند اثرات تغییرات آب و هوا، فعالیت‌های انسانی و فرآیندهای طبیعی مواجه هستند که پایداری آن‌ها را تهدید می‌کنند (Alabjah *et al.*, 2018; Baghvand *et al.*, 2010; Burri *et al.*, 2019; El Asri *et al.*, 2019; Houemenou *et al.*, 2020; Mountadar *et al.*, 2018). به‌طور کلی، این مشکلات ترکیب شیمیایی آب را کاهش داده و آن را برای مصارف آشامیدنی و یا آبیاری نامناسب می‌کنند. با این حال، نظارت و ارزیابی رویکردهای نوآورانه ممکن است برای مدیریت کیفیت آب و کاهش آلودگی و خطرات سلامتی ارزشمند باشد (Chowdury *et al.*, 2019; Krupkova *et al.*, 2019; Sakai *et al.*, 2018; Zhu *et al.*, 2019). عدم وجود داده‌های کافی برای توسعه مدل‌های ریاضی به‌علت محدودیت‌های زمانی و هزینه‌ای جهت اجرای مکرر این مدل‌ها به‌خصوص در شرایط ترکیب با مدل‌های بهینه‌سازی به‌منظور یافتن سناریوهای بهینه مدیریت آب زیرزمینی، کاربرد مدل‌های ریاضی را با محدودیت مواجه می‌کند (Plumb *et al.*, 2005). در چنین شرایطی، مدل‌های مبتنی بر هوش مصنوعی، ماشین یادگیری و داده‌کاوی می‌توانند جایگزین مناسبی به‌شمار آیند که در ارزیابی و نقشه‌برداری کیفیت آب‌های زیرزمینی استفاده می‌شوند (Adimalla *et al.*, 2019; Adimalla & Taloor, 2020; Das *et al.*, 2020; El Mountassir *et al.*, 2020; Heib *et al.*, 2020; Jahin *et al.*, 2020; Jamaa *et al.*, 2020; Taloor *et*

منطقه به ترتیب ۲۶۷۷ و ۹۹۷ متر می‌باشد. بارش‌های این منطقه اندک و نامنظم و به میزان ۵۵ میلی‌متر در سال و میزان تبخیر آن بین ۲۲۰۰ تا ۳۲۰۰ میلی‌متر در سال می‌باشد. منطقه دارای قدیمی‌ترین تشکیلات زمین‌شناسی (پرکامبرین) تا جوان‌ترین آن (هولوسن) است. سازند ذخایر تراسی و مخروط‌افکنه‌های کوهپایه‌ای جدید کم ارتفاع بیشترین مساحت منطقه را در بر گرفته است. منطقه مورد مطالعه فاقد رودخانه دائمی بوده و جریانات عمده رودخانه‌های این دشت منحصر به جریانات سیلابی است. سیمای پوشش گیاهی در محدوده مطالعاتی گیاهان درختچه‌ای و بوته‌ای می‌باشد و طبق طبقه‌بندی تاکسونومی خاک‌های منطقه در دو رده آنتی‌سول و اریدی‌سول قرار می‌گیرند.

پارامترهای کیفی آب زیرزمینی

برای بررسی کیفی نمونه‌ی آب زیرزمینی، ۲۰۱ حلقه چاه شبکه پایش شرکت آب منطقه‌ی استان یزد در سال ۱۳۹۵، از پارامترهای اندازه‌گیری شده هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول استفاده شد. شکل ۱ موقعیت چاه‌های منطقه مورد مطالعه در دشت یزد- اردکان را نشان می‌دهد.

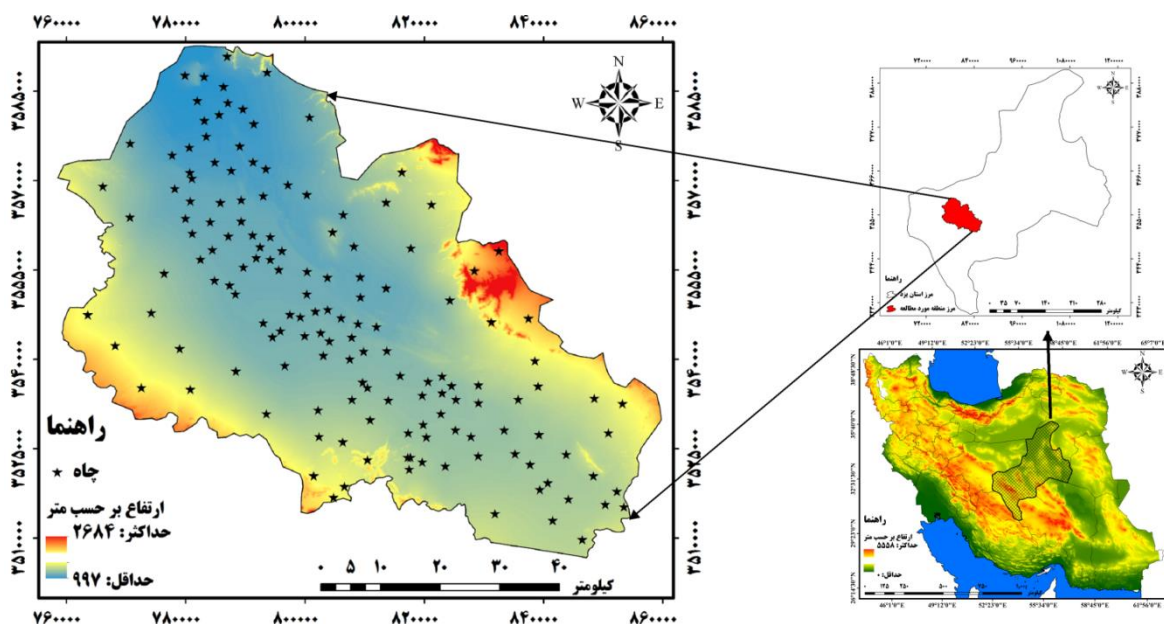
مدل جنگل تصادفی برای تعیین تناسب آب زیرزمینی برای اهداف آبیاری استفاده کردند.

با توجه به اهمیت فراوان شناخت خصوصیات کیفی آب زیرزمینی در مناطق بیابانی لزوم مدیریت صحیح آن، پیش‌بینی کیفیت آب زیرزمینی برای مدیریت و بهره‌برداری از منابع آب در دشت یزد- اردکان امری لازم و ضروری به نظر می‌رسد. بنابراین هدف از انجام این تحقیق بررسی و مقایسه کارایی مدل‌های شبکه عصبی مصنوعی و جنگل تصادفی در پیش‌بینی مقادیر هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول می‌باشد. در این تحقیق بر اساس ایجاد ارتباط بین داده‌های محیطی (کمکی) و پارامترهای کیفی آب زیرزمینی مدل‌سازی انجام خواهد گرفت.

مواد و روش‌ها

معرفی منطقه مورد مطالعه

منطقه مورد مطالعه با مساحت ۴۸۲۹۰۰ هکتار که از نظر موقعیت $31^{\circ} 45' 14''$ تا $31^{\circ} 48' 18''$ طول شرقی و $31^{\circ} 06' 14''$ تا $31^{\circ} 26' 37''$ عرض شمالی در فلات مرکزی ایران و بخش مرکزی استان یزد واقع شده است و حداکثر و حداقل ارتفاع از سطح دریای



شکل ۱- موقعیت چاه‌های منطقه مورد مطالعه در دشت یزد- اردکان

Figure 1- Location of wells in the study area in Yazd-Ardakan plain

تهیه داده‌های محیطی یا کمکی

داده‌های محیطی برای بهبود تخمین مکانی به‌ویژه در مواقع کمبود اطلاعات اصلی مفید هستند. بر اساس تحقیقات انجام شده، بیشترین ترکیب مربوط به فاکتورهای خصوصیات خاک و توپوگرافی بوده و اغلب موارد، سنجش از دور و لایه‌های اطلاعات خاک و به‌دنبال آن مدل رقومی ارتفاعی را به‌عنوان داده کمکی استفاده نموده‌اند (Fathizad *et al.*, 2020). از جمله فاکتورهای داده‌های محیطی (داده‌های کمکی) در این

تحقیق داده‌های زمین‌شناسی، کاربری اراضی، شاخص‌های پوشش گیاهی، شاخص‌های شوری خاک، مشتقات حاصل از مدل رقومی ارتفاعی، فاصله از معادن، فاصله از جاده، فاصله از رودخانه، فاصله از نقاط مسکونی، بارندگی و جمعیت می‌باشند. لازم به ذکر است که بعضی از لایه‌های اطلاعاتی موجود بوده (مثل نقشه زمین‌شناسی) و تنها در محیط GIS رقومی شده و از آن‌ها استفاده خواهند شد (جدول ۱ و شکل ۲).

جدول ۱- لیست پارامترهای محیطی استفاده شده

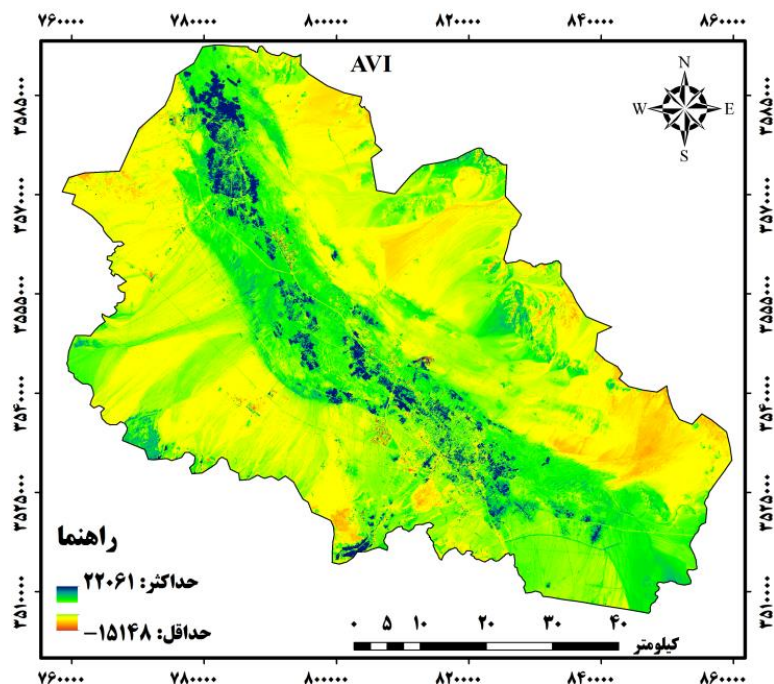
Table 1- List of environmental parameters used

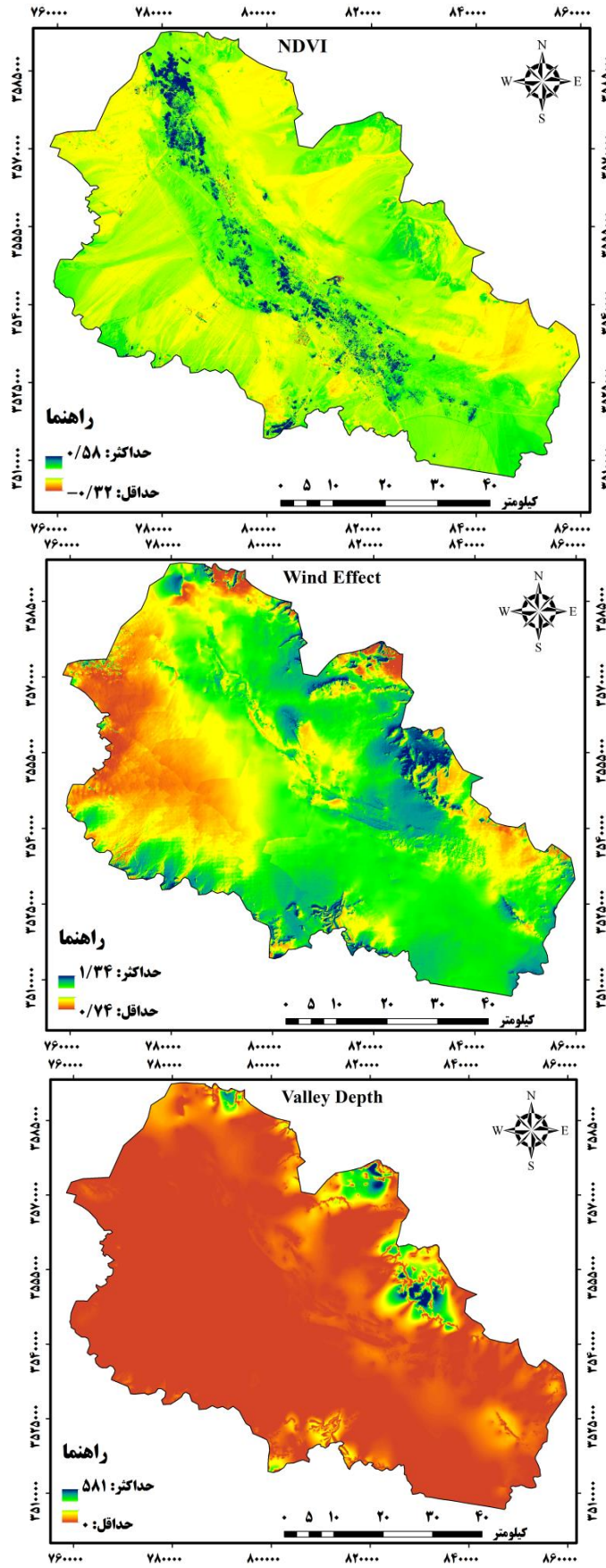
منبع	تعریف	داده محیطی
Conrad <i>et al.</i> , 2015	زاویه بین سطح و پرتوهای نور دریافتی	Analytical hillshade
Conrad <i>et al.</i> , 2015	تعادل بین توده خاک رسوب کرده و فرسایش یافته	Mass balance index
Conrad <i>et al.</i> , 2015	فاصله عمودی تا سطح پایه شبکه کانال را محاسبه می‌کند	Vertical distance to channel network
Conrad <i>et al.</i> , 2015	یک تقریب ساده از گرمای ناهمسانگرد روزانه را محاسبه می‌کند	Diurnal anisotropic heating
Conrad <i>et al.</i> , 2015	جریان انباشته شده را محاسبه می‌کند	Flow accumulation
Conrad <i>et al.</i> , 2015	ارتفاع مؤثر جریان هوا را محاسبه می‌کند	Effective air flow heights
Conrad <i>et al.</i> , 2015	گرادیان متوسط بالاتر از مسیر جریان	Slope gradient
Conrad <i>et al.</i> , 2015	موقعیت نسبی دره	Valley depth
Conrad <i>et al.</i> , 2015	ناهمواری زمین را اندازه‌گیری می‌کند	Terrain ruggedness index
Conrad <i>et al.</i> , 2015	پارامترهای سطح زمین مخصوص توپو اقلیم‌شناسی	Average view distance
Conrad <i>et al.</i> , 2015	شاخص بدون بعد مناطق در معرض باد را نشان می‌دهد	Wind effect
Conrad <i>et al.</i> , 2015	شاخص بی‌بعدی که پیکسل‌های در معرض باد را برجسته می‌کند	Wind exposition
Huete, 1988	NIR- Red	Difference vegetation index (DVI)
Huete, 1988	$G \times (NIR - Red) / (NIR + c1 \times Red - c2 \times Blue + L)$	Enhanced vegetation index (EVI)
Huete, 1988	$-0.29 (G) - 0.56 (Red) + 0.6 (IR) + 0.49 (NIR)$	Global vegetation index (GVI)
Crippen, 1990	$NIR / (NIR + Red)$	Infrared percentage vegetation index (IPVI)
Foody <i>et al.</i> , 2001	$(Red - NIR) / (Red + NIR)$	Normalized difference vegetation index (NDVI)
تصویر ماهواره لندست	مقدار بازتاب باند ماهواره لندست	Blue
تصویر ماهواره لندست	مقدار بازتاب باند ماهواره لندست	Green
تصویر ماهواره لندست	مقدار بازتاب باند ماهواره لندست	Red
تصویر ماهواره لندست	مقدار بازتاب باند ماهواره لندست	Near infrared (NIR)
تصویر ماهواره لندست	مقدار بازتاب باند ماهواره لندست	Shortwave infrared (SWIR)
Nield <i>et al.</i> , 2007	PC1, PC2, PC3, and PC4	Principal components of Landsat bands
Foody <i>et al.</i> , 2001	$(NIR - (TM1 + Green)) / (NIR + (TM1 + Green))$	Normalized-NDVI
Rondeaux <i>et al.</i> , 1996	$(NIR - Red) / (NIR + Red + 0.16)$	Optimized soil-adjusted vegetation index (OSAVI)
Arzani & King, 2008	Red - TM1	PD 311
Arzani & King, 2008	$(Red - Blue) / (Red + TM1)$	PD 312

مشتقات حاصل از مدل رقومی ارتفاعی (DEM)

شاخص‌های پوشش گیاهی

Arzani & King, 2008	Red – Green	PD 321	
Arzani & King, 2008	(Red– Green) / (Red + Green)	PD 322	
Arzani & King, 2008	NIR/ (Blue + Green)	Ratio-Based	
Foody <i>et al.</i> , 2001	(NIR/ Red)	Ratio vegetation index (RVI)	
Jordan, 1969	$[(NIR- Red) / (NIR + Red + L)] \times (1 + L)$	Soil-adjusted vegetation index (SAVI)	
Pettorelli <i>et al.</i> , 2005	(TM1× Green) / Red	Stress- related	
Kullberg <i>et al.</i> , 2017	(SWIR– Red)/ (SWIR + Red)	Transformed vegetation index (TVI)	
Kullberg <i>et al.</i> , 2017	Red/ Thermal	VIT01	
Kullberg <i>et al.</i> , 2017	Thermal/ (Red + SWIR)	VTI02	
Kullberg <i>et al.</i> , 2017	Thermal / Red	VIT03	
Kullberg <i>et al.</i> , 2017	Thermal / (SWIR + Green)	VIT04	
Khan <i>et al.</i> , 2005	$BI = ((Red \times Red) + (NIR \times NIR))^{0.5}$	Brightness index	
Wilson & Sader, 2002	$NDMI = (NIR- SWIR) / (NIR + SWIR)$	Normalized difference moisture index (NDMI)	
Major <i>et al.</i> , 1990	$NDSI = (Red- NIR) / (Red + NIR)$	Normalized difference snow index (NDSI)	
Nicolas & Walter, 2006	$S1 = Blue / Red$	Salinity index ₁ (S ₁)	شاخص‌های شوری خاک
Nicolas & Walter, 2006	$S2 = (Blue - Red) / (Blue + Red)$	Salinity index ₂ (S ₂)	
Nicolas & Walter, 2006	$S3 = (Green \times Red) / Blue$	Salinity index ₃ (S ₃)	
Nicolas & Walter, 2006	$S4 = (Blue \times Red) / Green$	Salinity index ₄ (S ₄)	
Nicolas & Walter, 2006	$S5 = (Red \times NIR) / Green$	Salinity index ₅ (S ₅)	
Nicolas & Walter, 2006	$S6 = (Blue \times Red)^{0.5}$	Salinity index ₆ (S ₆)	
Nicolas & Walter, 2006	$S7 = (Green \times Red)^{0.5}$	Salinity index ₇ (S ₇)	
Nicolas & Walter, 2006	$S8 = ((Blue)^2 \times (Green)^2 \times (Red)^2)^{0.5}$	Salinity index ₈ (S ₈)	
تصویر ماهواره لندست	نشان‌دهنده کاربری‌های یک "واحد" زمین	کاربری اراضی	سایر پارامترهای محیطی
Danielsson, 1980	فاصله اقلیدسی تا مناطق معدن	فاصله از معدن	
Danielsson, 1980	فاصله اقلیدسی تا مناطق شهری	فاصله مناطق مسکونی	
Danielsson, 1980	فاصله اقلیدسی تا مراکز جمعیتی	فاصله از جمعیت	
Danielsson, 1980	فاصله اقلیدسی تا آبراهه‌ها	فاصله از آبراهه	
Danielsson, 1980	فاصله اقلیدسی تا جاده‌ها	فاصله از جاده	
Fathizad <i>et al.</i> , 2020	نشان‌دهنده ویژگی‌های مختلف زمین‌شناسی	زمین‌شناسی	





شکل ۲- نقشه برخی از پارامترهای محیطی مورد استفاده

Figure 2- Map of some of the environmental parameters used

مدل‌سازی

بعد از آماده‌سازی پارامترهای کیفی آب زیرزمینی و داده‌های محیطی، اقدام پیش‌بینی مقادیر هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول با استفاده از مدل‌های شبکه عصبی مصنوعی و جنگل تصادفی گردید.

مدل شبکه عصبی مصنوعی (ANN^۱)

شبکه عصبی مصنوعی روشی محاسباتی است که جهت مدل‌سازی عملکردهای مغز انسان طراحی گردید. این روش جهت مدل‌سازی سیستم‌های عصبی درون رابطه‌ای توسعه یافت، به طوری که کامپیوترها بتوانند توانایی مغز را در طبقه‌بندی الگوها تقلید کنند و با آزمون و خطا یاد بگیرند، بنابراین می‌توانند روابط را در داده‌ها مشاهده کنند (Pijanowski *et al.*, 2002). ANN یک شبکه اتصال داخلی بزرگ از سلول‌های محاسباتی ساده که به نام «عصب» یا «واحدهای پردازش» نامیده می‌شوند، به کار می‌گیرد که توانایی نقشه‌برداری روابط درونی- بیرونی برای مسائل طبیعی می‌باشند (Haykin & Network, 1999). از آنجایی که ANN در کلاس مدل‌های جعبه سیاه قرار می‌گیرد، به جزئیات اطلاعات توابع درونی سیستم به منظور شناسایی روابط بین ورودی‌ها و خروجی‌ها نیازی ندارد (Ha & Stenstrom, 2003; Sabahattin *et al.*, 2013).

جنگل تصادفی (RF^۲)

جنگل تصادفی یک طبقه‌بندی مجموعه‌های متشکل از طبقه‌بندی‌های درخت تصمیم است. هر طبقه‌بندی برای هر نمونه ورودی به صورت $h(x, \Theta_k)$ است، که x یک نمونه ورودی و مجموعه آموزش برای درخت k ام است. Θ ها مستقل از یکدیگر ولی با توزیع یکسان هستند. برای هر نمونه x ، هر درخت یک پیش‌بینی را برای رده نمونه x ارائه می‌دهد و در نهایت رده‌ای با بیشترین تعداد رأی درختان روی ورودی x به عنوان رده نمونه انتخاب

می‌شود. این فرآیند را جنگل تصادفی می‌نامند (Breiman, 2001).

الگوریتم ساخت یک جنگل تصادفی با T درخت از یک مجموعه داده با n مشاهده و p متغیر بدین صورت است:

۱- با روش بوت‌استرپ یک نمونه تصادفی با جایگذاری به حجم n از مشاهدات انتخاب می‌شود.

۲- برای نمونه بوت‌استرپ انتخاب شده یک درخت کلاس‌بندی با استفاده از الگوریتم افزایشی بازگشتی، رشد می‌کند. در هر گره‌افراز بر اساس یک نمونه تصادفی m تایی از p متغیر پیشگو انجام می‌شود.

۳- الگوریتم افزایشی بازگشتی آنقدر ادامه می‌یابد تا درخت به بزرگ‌ترین اندازه خود (یعنی برای هر مشاهده یک گره نهایی)، برسد بدون آنکه درخت هرس شود.

۴- مراحل یک تا سه، T بار تکرار می‌شوند تا یک جنگل تصادفی ساخته شود (Breiman, 2001).

انتخاب‌های رایج برای T ، 1000 درخت و برای m ، \sqrt{p} و $\text{Log}(p)$ هستند (Genuer, 2008).

الگوریتم جنگل تصادفی می‌تواند دقت پیش‌بینی را نسبت به درخت طبقه‌بندی فردی افزایش دهد. در درخت فردی با تغییرات کوچک در مجموعه آموزش بی‌ثباتی به وجود می‌آید که باعث اختلال در دقت پیش‌بینی در نمونه آزمایشی می‌شود. اما گروهی بودن الگوریتم جنگل تصادفی باعث سازگاری با تغییرات می‌شود و بی‌ثباتی را از بین می‌برد.

به منظور اعتبارسنجی مدل جنگل تصادفی، از روش اعتبارسنجی متقابل (10-fold) در نرم‌افزار آماری R استفاده شد. در این روش داده‌ها به ۱۰ قسمت تقسیم می‌شود، نه قسمت از داده‌ها برای مدل‌سازی و یک قسمت باقیمانده برای اعتبارسنجی مدل به دست آمده استفاده می‌شود. به این ترتیب ۱۰ مرتبه این روش تکرار می‌شود. یک جنگل تصادفی آنقدر بزرگ است که تفسیر آن کار بسیار دشواری است، بنابراین نیازمند خلاصه

آن‌ها را نشان می‌دهد.

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

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

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

که X_i, Y_i, X_i^*, Y_i^* به ترتیب مربوط به مقادیر اندازه‌گیری و برآورد شده، میانگین مقادیر اندازه‌گیری و برآورد شده و n تعداد کل مشاهدات می‌باشد.

نتایج

نتایج آمار توصیفی پارامترهای اندازه‌گیری شده در جدول ۲ آورده شده است. اگر ضریب تغییرات کم‌تر از ۱۰ درصد باشد، تغییرپذیری کم، اگر ضریب تغییرات بیشتر از ۱۰ درصد باشد تغییرپذیری متوسط و اگر بیشتر از ۱۰۰ درصد باشد تغییرپذیری بسیار زیاد است (Nielsen & Bouma, 1985). بنابراین همان‌طور که در جدول ۲ مشاهده می‌شود می‌توان شدت تغییرات تمامی ویژگی‌های مورد اندازه‌گیری شده در منطقه مورد مطالعه را متوسط در نظر گرفت.

کردن اطلاعات آن با استفاده از شاخص‌های کمی است. یکی از این شاخص‌ها اهمیت متغیر است (VI)^۱ است. VI شاخصی برای رتبه‌بندی متغیرها بر حسب اهمیت آن‌ها در اثرگذاری روی پاسخ است. در این تحقیق از شاخص‌های اهمیت میانگین مربع خطا (% Inc MSE) و خلوص گره (Inc Node Purity)^۲ استفاده شد. به‌منظور تعیین سهم هر یک از متغیرهای معنی‌دار در مدل از بسته varImp_0.2 در نرم‌افزار آماری R استفاده شد. در این تحقیق برای مدل پیش‌بینی نهایی از اعتبارسنجی^۳ متقابل ۲:۱ استفاده و با روش نمونه‌برداری خطی^۴ کردیم. برای کالیبراسیون^۵ از ۷۰ درصد داده‌ها و برای ارزیابی^۶ از ۳۰ درصد داده‌ها بدون جایگزینی استفاده شد.

معیارهای ارزیابی مدل

جهت ارزیابی کارایی مدل‌های شبکه عصبی مصنوعی و جنگل تصادفی برای پارامترهای آب زیرزمینی، از شاخص‌های آماری میانگین خطای مطلق، جذر میانگین مربعات خطا، ضریب تبیین استفاده شد. روابط ۱ تا ۳

جدول ۲- آمار توصیفی پارامترهای اندازه‌گیری شده

Table 2- Descriptive statistics of measured parameters

پارامتر Parameter	تعداد Number	انحراف معیار Standard deviation	چولگی Skewness	حداکثر Maximum	حداقل Minimum	میانگین Means	ضریب تغییرات CV
هدایت الکتریکی EC	201	999.7	1.3	278.0	6285.0	1754.6	57.0
نسبت جذب سدیم SAR	201	2.1	2.1	0.7	15.3	3.4	63.6
سولفات SO4	201	2	1.2	0.6	12.8	3.7	55.3

تخمین از یک شبکه عصبی MLP^۷ سه لایه استفاده شد. برای مدل‌سازی هر یک پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول شبکه دارای ۵۶ نورون در لایه ورودی (تعداد

نتایج مدل‌سازی شبکه عصبی مصنوعی

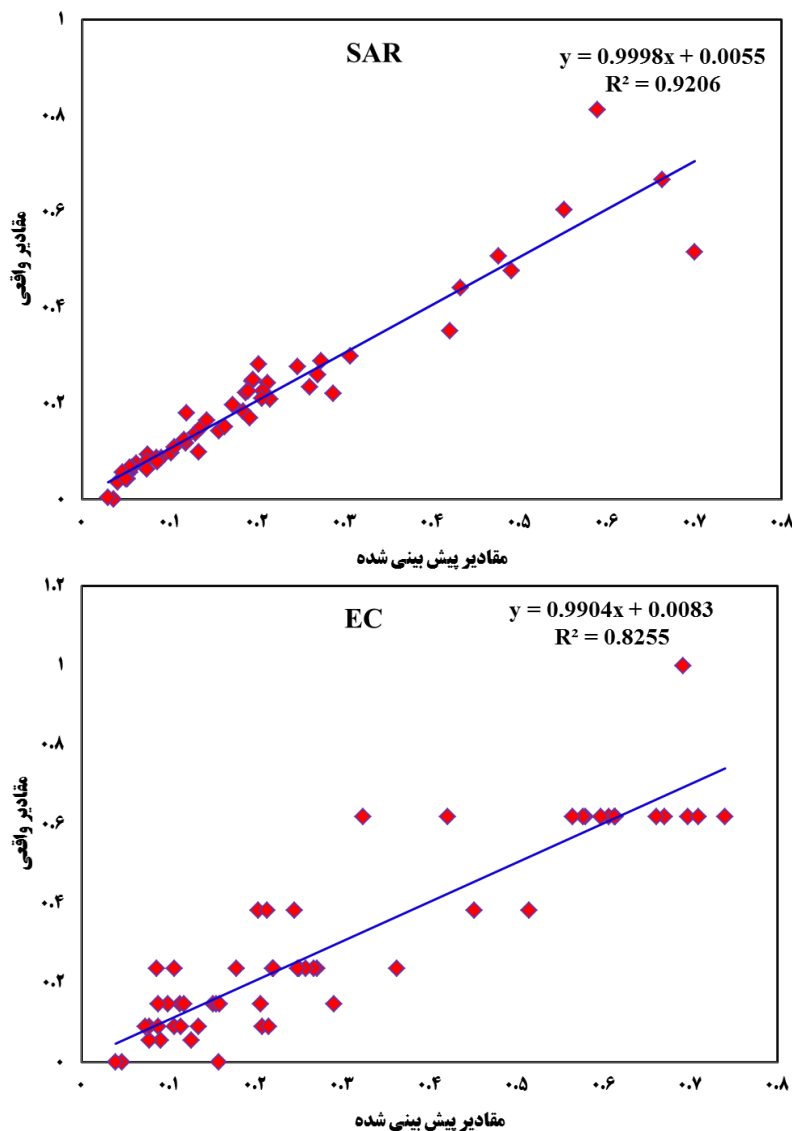
در این تحقیق تعداد لایه‌های مخفی و شمار نورون‌ها در لایه‌های میانی متناسب با تعداد نورون‌ها لایه‌ی میانی بر مبنای مقایسه‌ی عملکرد شبکه انتخاب شد. برای

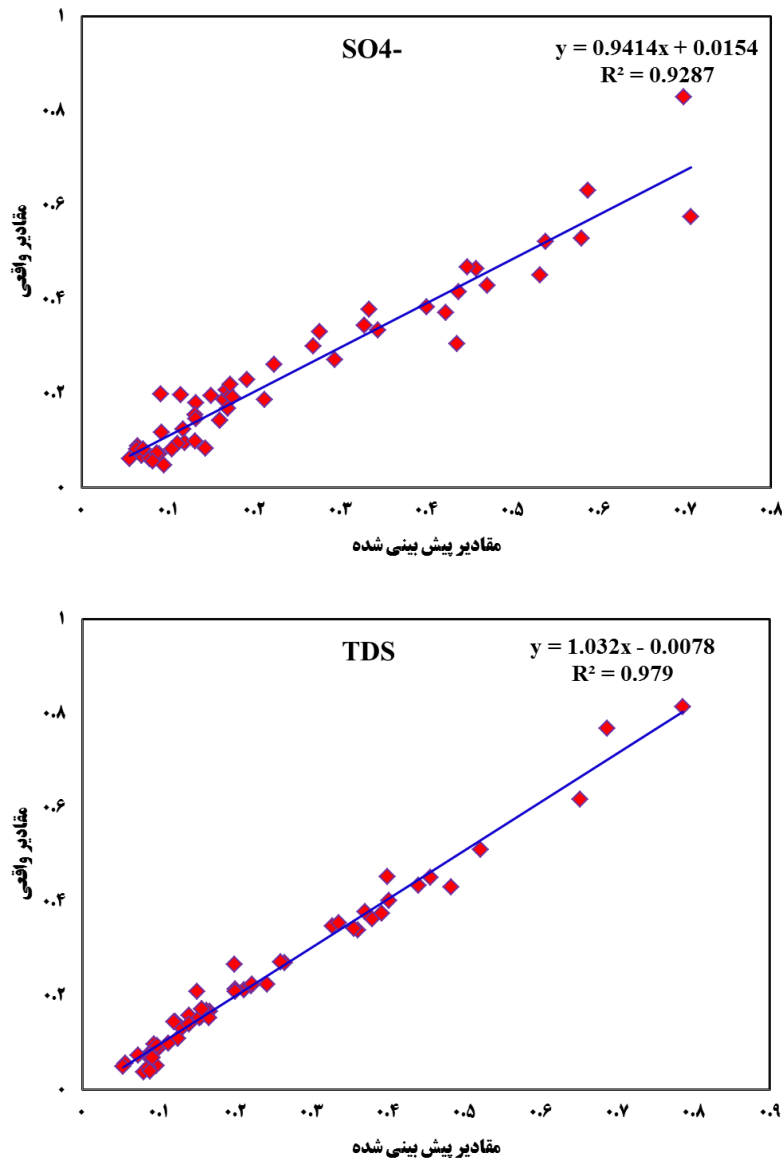
4. Linear sampling
5. Calibration
6. Validation
7. Multilayer perceptron artificial neural network

1. Variable importance
2. Mean square-error (% Inc MSE) and node purity (Inc Node Purity)
3. Cross-validation

انتخاب می‌گردد (Almasri & Kaluarachchi, 2005). در این تحقیق در شبیه‌سازی پارامترهای EC، SAR، TDS و SO_4^- بهترین ساختار به‌دست آمده از ۱۰۰ تکرار یادگیری شبکه عصبی مصنوعی، دارای دو لایه پنهان و ۱۰ نورون پنهان در هر لایه است. جهت ارزیابی دقت شبکه عصبی انتخاب شده، یک تحلیل رگرسیون بین X داده‌های پیش‌بینی شده و Y داده‌های مقادیر واقعی پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول انجام گرفت. نمودار پراکنش مقادیر واقعی و برآورده شده پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول در شکل ۳ ارائه شده است.

پارامترهای محیطی) و یک نورون در لایه خروجی (هر کدام از پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول) به‌کار برده شد. تعداد نورون‌های لایه پنهان، با توجه به اینکه هنوز از روش خاصی تبعیت نمی‌کنند، به‌صورت سعی و خطا تعیین شدند، به این صورت که شبکه با یک نورون آموزش داده شده و RMSE آن محاسبه می‌گردد. سپس به نورون‌ها یک واحد اضافه گردیده و مجدداً مقادیر RMSE محاسبه شده و این عمل آنقدر تکرار می‌گردد تا تعداد نورون مناسب تعیین گردد. در نهایت شبکه‌ای که بهترین جواب را در مرحله‌ی آموزش و آزمون به‌دست می‌آورد به‌عنوان مناسب‌ترین شبکه برای مدل‌سازی



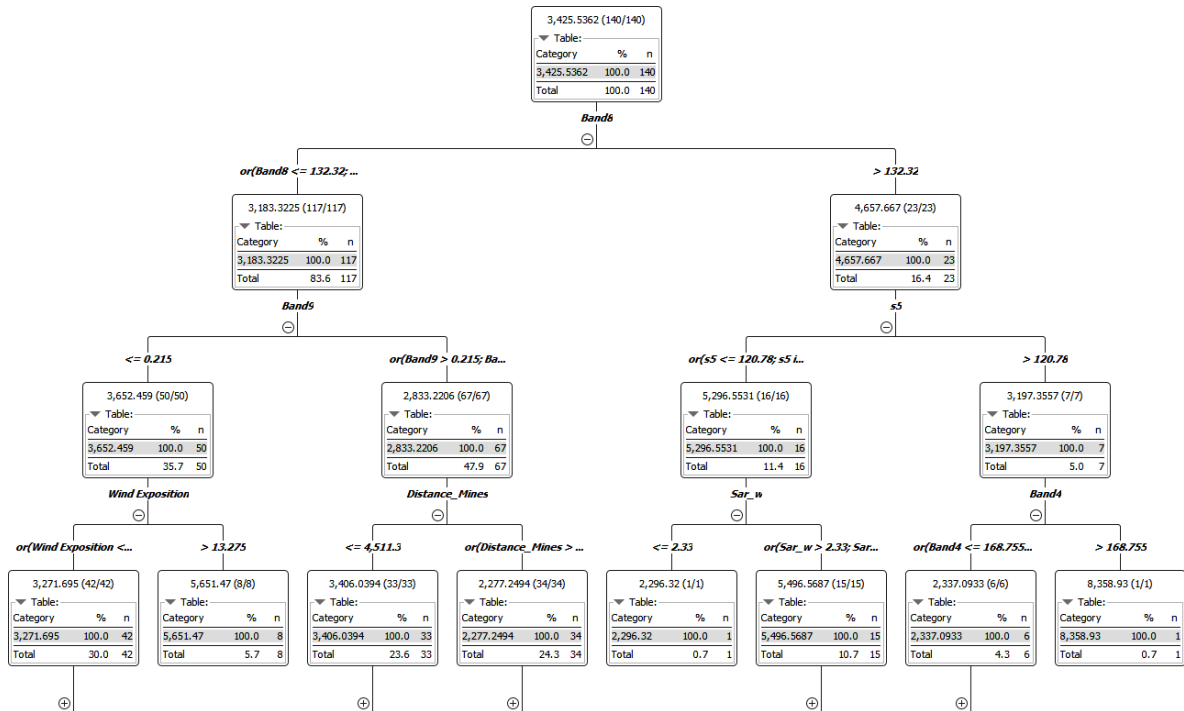


شکل ۳- نمودار پراکنش مقادیر واقعی و برآورده شده پارامترهای EC, SAR, SO_4^- و TDS با استفاده از مدل شبکه عصبی
 Figure 3- Distribution diagram of true and estimated values of EC, SAR, SO_4^- and TDS parameters using neural network model

مناسب و به‌صرفه است. همچنین تلفیق داده‌های ماهواره‌ای با DEM به‌منظور بررسی پارامترهای کیفی آب زیرزمینی و پهنه‌بندی آن‌ها باعث کارآمدتر شدن نتایج و بالا رفتن صحت و دقت آن می‌شود نکته بسیار مهم در آنالیز پارامترها، نقش مهم پارامترهای مانند فاصله از جاده و فاصله از معادن است. این نشان‌دهنده اثر مستقیم عوامل محیطی انسان ساخت در کیفیت آب زیرزمینی منطقه مورد مطالعه است.

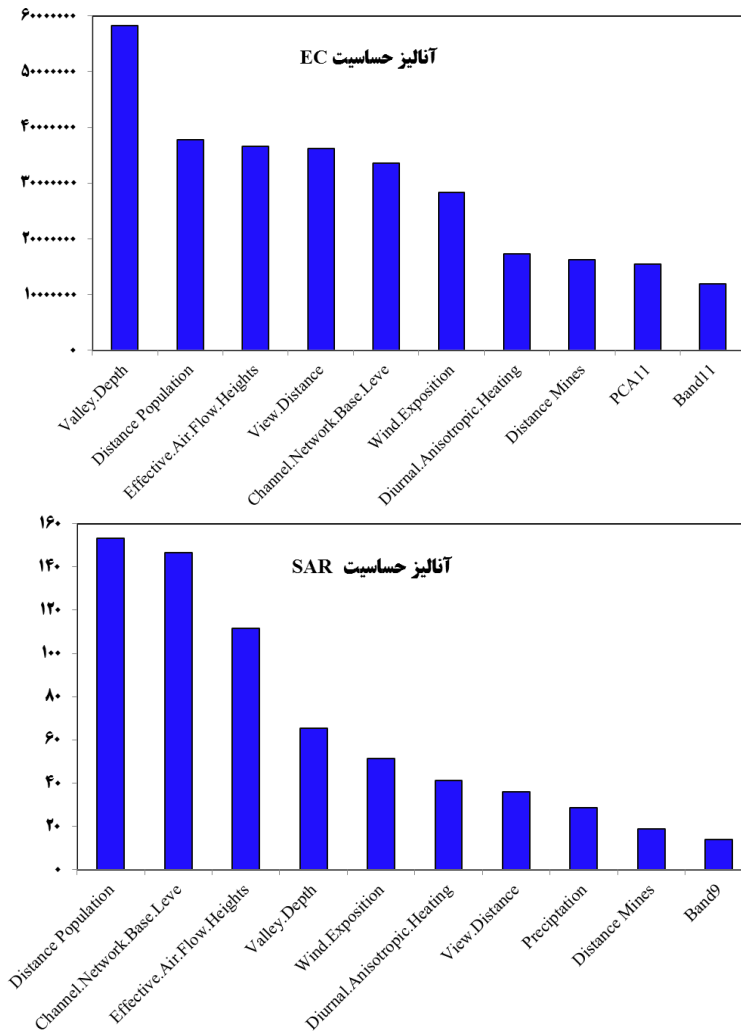
نتایج مدل‌سازی جنگل تصادفی

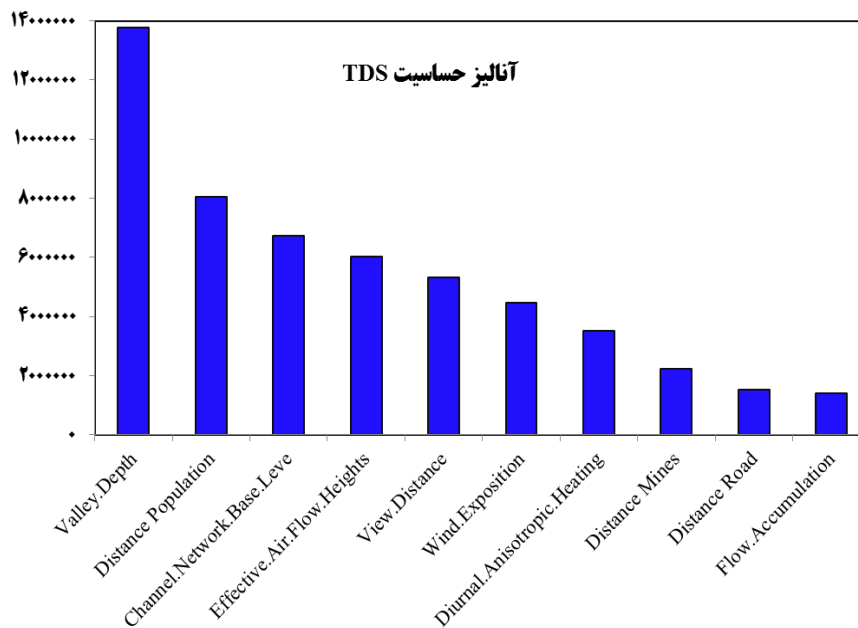
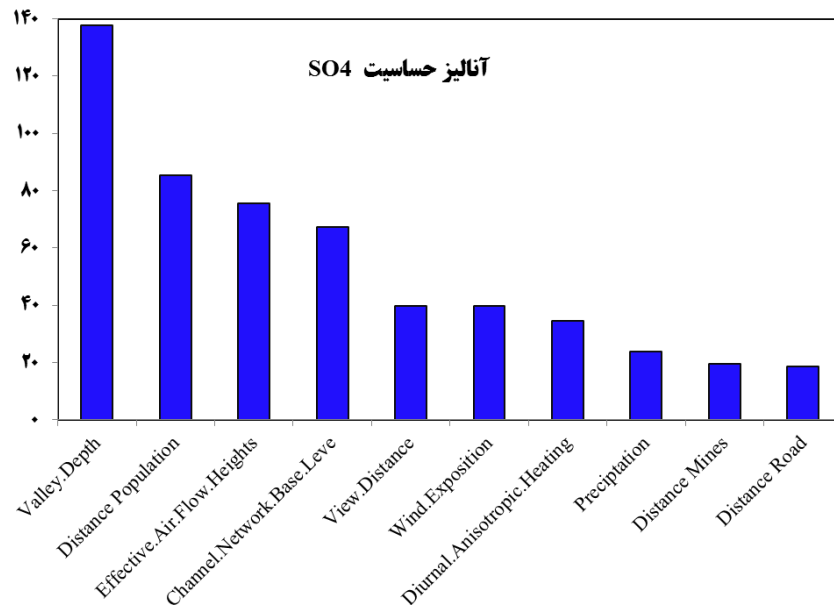
ساختار درختی مدل‌سازی جنگل تصادفی در شکل ۴ و آنالیز حساسیت مدل جنگل تصادفی پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول در شکل ۵ آورده شده است. نتایج نشان‌دهنده اهمیت بالای پارامترهای مستخرج شده از مدل رقومی ارتفاعی است و استفاده از داده‌های تصاویر ماهواره‌ای برای بررسی پارامترهای کیفی آب زیرزمینی روش



شکل ۴- ساختار درختی مدل‌سازی جنگل تصادفی

Figure 4- The tree structure of random forest modeling



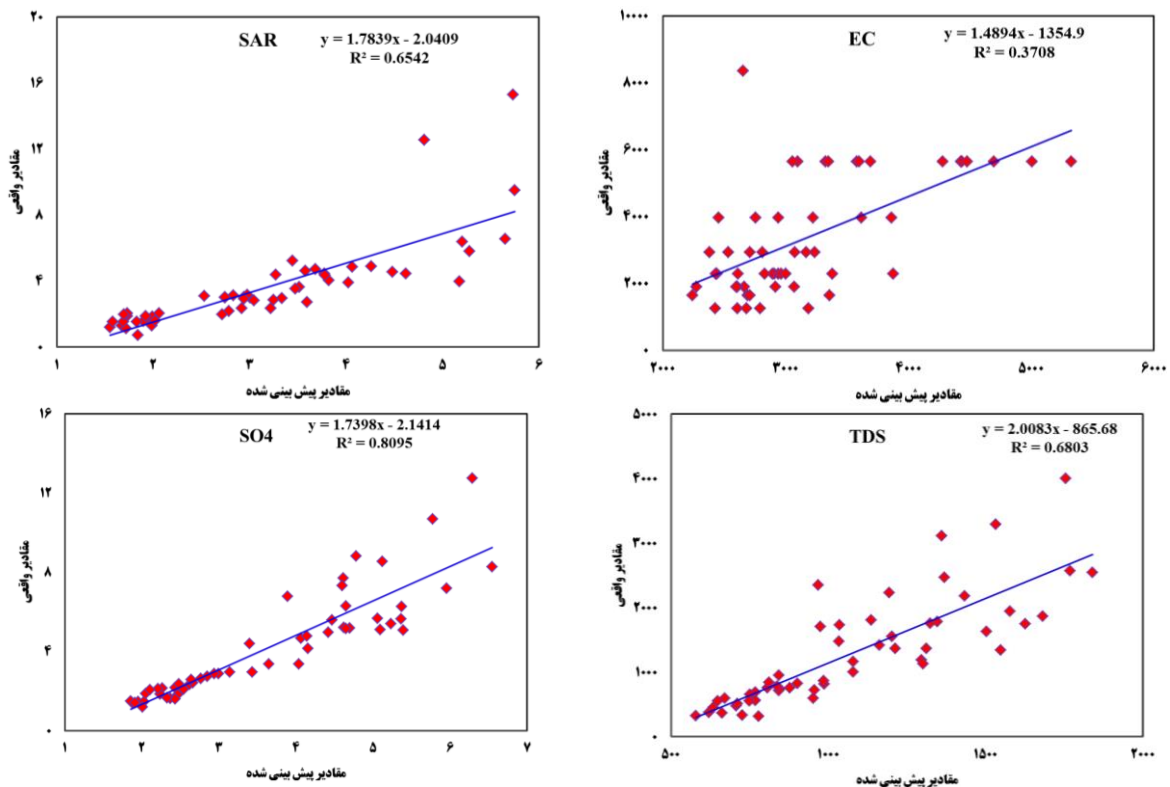


شکل ۵- نمودار آنالیز اهمیت پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول با استفاده از مدل جنگل تصادفی

Figure 5- Graph analysis of the importance of electrical conductivity parameters, adsorption ratio of sodium, sulfate and total soluble solids using a random forest model

محلول انجام گرفت. نمودار پراکنش مقادیر واقعی و برآورده شده پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول در شکل ۶ ارائه شده است.

جهت ارزیابی دقت مدل جنگل تصادفی انتخاب شده، یک تحلیل رگرسیون بین X داده‌های پیش‌بینی شده و Y داده‌های مقادیر واقعی پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد



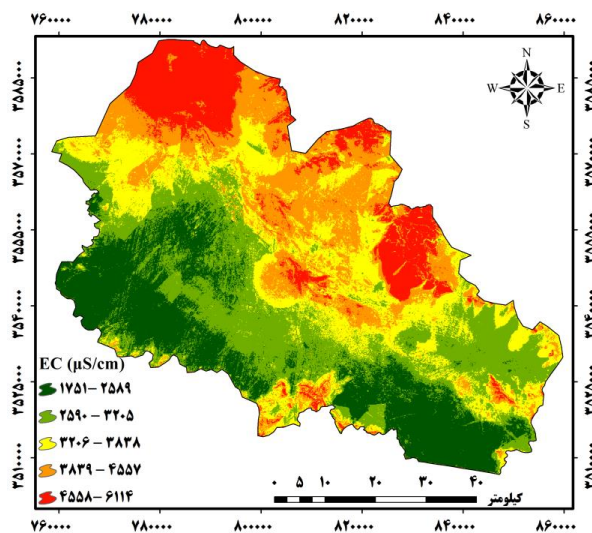
شکل ۶- نمودار پراکنش مقادیر واقعی و برآورده شده پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول با استفاده از مدل جنگل تصادفی

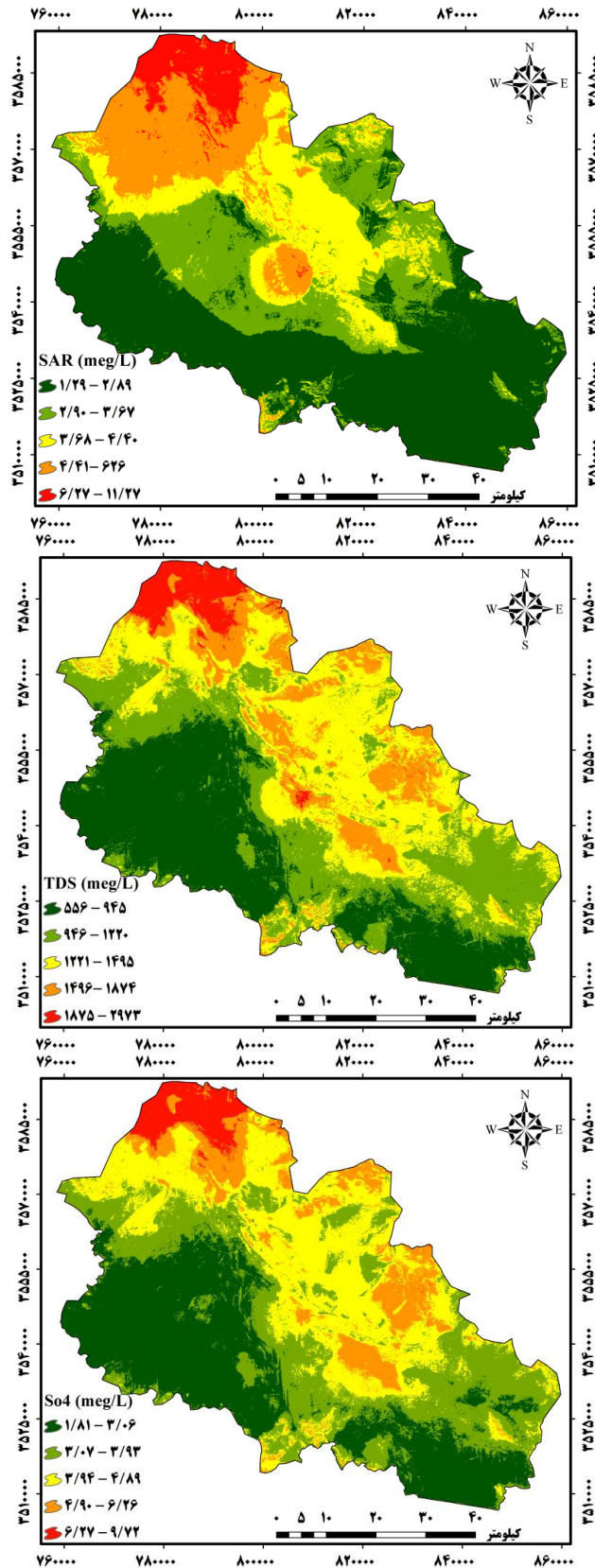
Figure 6- Distribution diagram of true and estimated values of EC, SAR, SO₄⁻ and TDS parameters using a random forest model

مصنوعی دقت پیش‌بینی بیشتری نسبت به مدل جنگل تصادفی دارد. به همین دلیل از مدل شبکه عصبی مصنوعی برای تهیه نقشه پارامترهای هدایت الکتریکی، نسبت جذب سدیم، مقدار سولفات و کل مواد جامد محلول آب زیرزمینی دشت یزد- اردکان استفاده شد (شکل ۷).

تهیه نقشه پارامترهای آب زیرزمینی

با مقایسه رگرسیون بین داده‌های مقادیر واقعی و مقادیر پیش‌بینی شده خصوصیات هدایت الکتریکی، نسبت جذب سدیم، مقدار سولفات و کل مواد جامد محلول حاصل از مدل‌های شبکه عصبی مصنوعی و جنگل تصادفی مشاهده می‌شود که مدل شبکه عصبی





شکل ۷- نقشه توزیع مکانی پارامترهای هیدروشیمیایی در دشت یزد-اردکان با استفاده از مدل شبکه عصبی مصنوعی

Figure 7- Spatial distribution map of hydrochemical parameters in Yazd-Ardakan plain using artificial neural network model

بحث

میکروزیمنس بر سانتی‌متر می‌باشد. قسمت مرکزی، شمال و شمال‌غربی منطقه مورد پژوهش بیشترین میزان هدایت الکتریکی را دارا می‌باشد و چون بافت اصلی خاک در این مناطق از نوع لومی، شنی-لومی، سیلت-لومی و سیلتي می‌باشد، نفوذ آب در این مناطق کم بوده و در نتیجه منابع زیرزمینی در این مناطق توسط بارش‌ها احیاء نمی‌شوند و در نهایت روزبه‌روز کیفیت آب در این نواحی رو به کاهش است. وجود آب‌های شور در این ناحیه نیز به همین دلیل می‌باشد (Derakhshan et al., 2015). از علت‌های کاهش کیفیت آب در این ناحیه، خشک‌سالی‌های اخیر و اراضی کشاورزی تحت کشت پسته در میبد و اردکان به‌علت آبیاری نامناسب می‌باشد که در کل روی کیفیت منابع آب تأثیر گذاشته و باعث کاهش کیفیت آب شده است. وضعیت سایر پارامترهای نسبت جذب سدیم، سولفات و کل مواد جامد محلول مشابه پارامتر هدایت الکتریکی می‌باشد. در شمال، مرکز و جنوب‌غربی دارای بالاترین مقدار و در جنوب‌شرقی و جنوب‌غربی دارای پایین‌ترین مقدار می‌باشند.

نتیجه‌گیری کلی

آنالیز حساسیت مدل‌سازی پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول آب زیرزمینی با استفاده از مدل جنگل تصادفی در منطقه مورد مطالعه نشان‌دهنده اهمیت بالای شاخص‌های پوشش گیاهی تصاویر ماهواره‌ای می‌باشد. به‌صورت کلی می‌توان بیان کرد که سنجش از دور در تهیه نقشه پارامترهای کیفی آب زیرزمینی ابزاری بسیار قوی بوده که در صورت موجود بودن تصاویر می‌توان با تعداد نقاط محدود و صرف هزینه کم نقشه‌های بسیار دقیقی تهیه نمود.

نتایج مدل‌سازی پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول با استفاده از مدل‌های شبکه عصبی مصنوعی و جنگل تصادفی نشان‌دهنده دقت مناسب مدل شبکه عصبی مصنوعی در پیش‌بینی این پارامترها است به‌طوری‌که نقشه نهایی پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول آب زیرزمینی را به‌ترتیب با ضریب تبیین ۰/۸۲، ۰/۹۲، ۰/۹۲ و ۰/۹۷ پیش‌بینی کرده است و این در حالی است که مدل جنگل تصادفی پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول آب زیرزمینی را به‌ترتیب با ضریب تبیین ۰/۳۷، ۰/۶۵، ۰/۸۰ و ۰/۶۸ پیش‌بینی نموده است. بر اساس نتایج به‌دست آمده می‌توان نتیجه گرفت که مدل شبکه عصبی مصنوعی از دقت بالاتری نسبت به مدل جنگل تصادفی در پیش‌بینی پارامترهای کیفی آب زیرزمینی داشته است. در خصوص کارایی شبکه عصبی مصنوعی در مدل‌سازی پارامترهای کیفی آب زیرزمینی واضح است که توانایی‌های این تکنیک با توجه به ساختارهای مختلف آن و نیز طبیعت مسئله مورد نظر، متفاوت می‌باشد. با انتخاب نوع و تعداد مناسب عوامل ورودی و نیز استفاده از نوع مناسب و سازگار شبکه عصبی مصنوعی و نیز کالیبره کردن مناسب آن می‌توان گفت که این تکنیک ابزاری بسیار کارا و مناسب برای برآورد پارامترهای هدایت الکتریکی، نسبت جذب سدیم، سولفات و کل مواد جامد محلول در دشت یزد- اردکان می‌باشد.

نقشه‌های پیش‌بینی پارامترهای کیفی آب زیرزمینی نشان می‌دهد که حداکثر مقدار هدایت الکتریکی در منطقه مورد مطالعه ۶۱۱۴ و حداقل آن ۱۷۵۱

References

- Abu-Khalaf, N., Khayat, S. & Natsheh, B. (2013). Multivariate data analysis to identify the groundwater pollution sources in Tulkarm area/Palestine. *Science and Technology*, 3(4), 99-104. Doi: 10.5923/j.scit.20130304.01.
- Adimalla, N. & Taloor, A. K. (2020). Hydrogeochemical investigation of groundwater quality in the hard rock terrain of South India using Geographic Information System (GIS) and

- groundwater quality index (GWQI) techniques. *Groundwater for Sustainable Development*, 10, 100288. Doi: 10.1016/j.gsd.2019.100288.
- Adimalla, N., Qian, H. & Li, P. (2019). Entropy water quality index and probabilistic health risk assessment from geochemistry of groundwaters in hard rock terrain of Nanganur County, South India. *Geochemistry*, 125544. Doi: 10.1016/j.chemer.2019.125544.
 - Alabjah, B., Amraoui, F., Chibout, M. & Slimani, M. (2018). Assessment of saltwater contamination extent in the coastal aquifers of Chaouia (Morocco) using the electric recognition. *Journal of Hydrology*, 566, 363-376. Doi: 10.1016/j.jhydrol.2018.09.003.
 - Almasri, M. N. & Kaluarachchi, J. J. (2005). Modular neural networks to predict the nitrate distribution in ground water using the on-ground nitrogen loading and recharge data. *Environmental Modelling & Software*, 20(7), 851-871. Doi: 10.1016/j.envsoft.2004.05.001.
 - Arzani, H. & King, G. W. (2008). Application of remote sensing (Landsat TM data) for vegetation parameters measurement in western division of NSW. In *International Grassland Congress. Hohhot, China. ID (No. 1083)*.
 - Baghvand, A., Nasrabadi, T., Bidhendi, G. N., Vosoogh, A., Karbassi, A. & Mehrdadi, N. (2010). Groundwater quality degradation of an aquifer in Iran central desert. *Desalination*, 260(1-3), 264-275. Doi: 10.1016/j.desal.2010.02.038.
 - Bilali, A. E., Taleb, A., Mazigh, N. & Mokhliss, M. (2020). Prediction of chemical water quality used for drinking purposes based on artificial neural networks. *Moroccan Journal of Chemistry*, 8(3), 8-3. Doi: 10.48317/IMIST.PRSM/morjchem-v8i3.19786.
 - Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. Doi: 10.1023/A:1010933404324.
 - Burri, N. M., Weatherl, R., Moeck, C. & Schirmer, M. (2019). A review of threats to groundwater quality in the anthropocene. *Science of the Total Environment*, 684, 136-154. Doi: 10.1016/j.scitotenv.2019.05.236.
 - Castrillo, M. & Garcia, A. L. (2020). Estimation of high frequency nutrient concentrations from water quality surrogates using machine learning methods. *Water Research*, 172, 115490. Doi: 10.1016/j.watres.2020.115490.
 - Chou, J. S., Ho, C. C. & Hoang, H. S. (2018). Determining quality of water in reservoir using machine learning. *Ecological Informatics*, 44, 57-75. Doi: 10.1016/j.ecoinf.2018.01.005.
 - Chowdury, M. S. U., Emran, T. B., Ghosh, S., Pathak, A., Alam, M. M., Absar, N. & Hossain, M. S. (2019). IoT based real-time river water quality monitoring system. *Procedia Computer Science*, 155, 161-168. Doi: 10.1016/j.procs.2019.08.025.
 - Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L. & Bohner, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8(7), 1991-2007. Doi: 10.5194/gmd-8-1991-2015.
 - Crippen, R. E. (1990). Calculating the vegetation index faster. *Remote Sensing of Environment*, 34(1), 71-73. Doi: 10.1016/0034-4257(90)90085-Z.
 - Danielsson, P. E. (1980). Euclidean distance mapping. *Computer Graphics and Image Processing*, 14(3), 227-248. Doi: 10.1016/0146-664X(80)90054-4.
 - Das, A., Das, S. S., Chowdhury, N. R., Joardar, M., Ghosh, B. & Roychowdhury, T. (2020). Quality and health risk evaluation for groundwater in Nadia district, West Bengal: An approach on its suitability for drinking and domestic purpose. *Groundwater for Sustainable Development*, 10, 100351. Doi: 10.1016/j.gsd.2020.100351.
 - Derakhshan, Z., Faramarzan, M., Miri, M. & Shokouhi, M. R. (2015). The zoning of groundwater quality for drinking purpose using scholler model and geographic information system (GIS). *Journal of Community Health Research*, 4(2), 138-147.
 - Di, Z., Chang, M. & Guo, P. (2019). Water quality evaluation of the Yangtze River in China using machine learning techniques and data monitoring on different time scales. *Water*, 11(2), 339. Doi: 10.3390/w11020339.
 - El Asri, H., Larabi, A. & Faouzi, M. (2019). Climate change projections in the Ghis-Nekkor region of Morocco and potential impact on groundwater recharge. *Theoretical and Applied Climatology*, 138(1), 713-727. Doi: 10.1007/s00704-019-02834-8.
 - El Mountassir, O., Bahir, M., Ouazar, D., Ouhamdouch, S., Chehbouni, A. & Ouarani, M. (2020). The use of GIS and water quality index to assess groundwater quality of krimat aquifer (Essaouira; Morocco). *SN Applied Sciences*, 2(5), 1-16. Doi: 10.1007/s42452-020-2653-z.

- Fathizad, H., Ardakani, M. A. H., Sodaiezhadeh, H., Kerry, R. & Taghizadeh-Mehrjardi, R. (2020). Investigation of the spatial and temporal variation of soil salinity using random forests in the central desert of Iran. *Geoderma*, 365, 114233. Doi: 10.1016/j.geoderma.2020.114233.
- Fijani, E., Barzegar, R., Deo, R., Tziritis, E. & Skordas, K. (2019). Design and implementation of a hybrid model based on two-layer decomposition method coupled with extreme learning machines to support real-time environmental monitoring of water quality parameters. *Science of the Total Environment*, 648, 839-853. Doi: 10.1016/j.scitotenv.2018.08.221.
- Foody, G. M., Cutler, M. E., McMorrow, J., Pelz, D., Tangki, H., Boyd, D. S. & Douglas, I. A. N. (2001). Mapping the biomass of Bornean tropical rain forest from remotely sensed data. *Global Ecology and Biogeography*, 10(4), 379-387. Doi: 10.1046/j.1466-822X.2001.00248.x.
- Genuer, R., Poggi, J. M. & Tuleau, C. (2008). Random forests: some methodological insights. *arXiv preprint arXiv:0811.3619*.
- Ha, H. & Stenstrom, M. K. (2003). Identification of land use with water quality data in stormwater using a neural network. *Water Research*, 37(17), 4222-4230. Doi: 10.1016/S0043-1354(03)00344-0.
- Haykin, S. & Network, N. (2004). A comprehensive foundation. *Neural Networks*, 2(2004), 41.
- Heib, L., Bouchaou, L., Tadoumant, S. & Reichert, B. (2020). Index-based groundwater vulnerability and water quality assessment in the arid region of Tata city (Morocco). *Groundwater for Sustainable Development*, 10, 100344. Doi: 10.1016/j.gsd.2020.100344.
- Houemenou, H., Tweed, S., Dobigny, G., Mama, D., Alassane, A., Silmer, R., Babic, M., Ruy, S., Chaigneau, A., Gauthier, P., Socohou, A., Dossou, H.J., Badou, S. & Leblanc, M. (2020). Degradation of groundwater quality in expanding cities in West Africa. A case study of the unregulated shallow aquifer in Cotonou. *Journal of Hydrology*, 582, 124438. Doi: 10.1016/j.jhydrol.2019.124438.
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295-309. Doi: 10.1016/0034-4257(88)90106-X.
- Jahin, H. S., Abuzaid, A. S. & Abdellatif, A. D. (2020). Using multivariate analysis to develop irrigation water quality index for surface water in Kafr El-Sheikh Governorate, Egypt. *Environmental Technology & Innovation*, 17, 100532. Doi: 10.1016/j.eti.2019.100532.
- Jamaa, H., El Achheb, A. & Namr, K. I. (2020). Spatial variation of groundwater quality and assessment of water table fluctuations in Plio-Quaternary aquifer formations in Doukkala Plain, Morocco. *Groundwater for Sustainable Development*, 11, 100398. Doi: 10.1016/j.gsd.2020.100398.
- Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 50(4), 663-666. Doi: 10.2307/1936256.
- Khan, N. M., Rastoskuev, V. V., Sato, Y. & Shiozawa, S. (2005). Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agricultural Water Management*, 77(1-3), 96-109. Doi: 10.1016/j.agwat.2004.09.038.
- Krupkova, L., Havrankova, K., Krejza, J., Sedlak, P. & Marek, M. V. (2019). Impact of water scarcity on spruce and beech forests. *Journal of Forestry Research*, 30(3), 899-909. Doi: 10.1007/s11676-018-0642-5.
- Kullberg, E. G., DeJonge, K. C. & Chavez, J. L. (2017). Evaluation of thermal remote sensing indices to estimate crop evapotranspiration coefficients. *Agricultural Water Management*, 179, 64-73. Doi: 10.1016/j.agwat.2016.07.007.
- Lu, H. & Ma, X. (2020). Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, 249, 126169. Doi: 10.1016/j.chemosphere.2020.126169.
- Major, D. J., Baret, F. & Guyot, G. (1990). A ratio vegetation index adjusted for soil brightness. *International journal of Remote Sensing*, 11(5), 727-740. Doi: 10.1080/01431169008955053.
- Meyers, G., Kapelan, Z. & Keedwell, E. (2017). Short-term forecasting of turbidity in trunk main networks. *Water Research*, 124, 67-76. Doi: 10.1016/j.watres.2017.07.035.
- Mountadar, S., Younsi, A., Hayani, A., Siniti, M. & Tahiri, S. (2018). Groundwater salinization process in the coastal aquifer sidi abed-ouled ghanem (province of el Jadida, Morocco). *Journal of African Earth Sciences*, 147, 169-177. Doi: 10.1016/j.jafrearsci.2018.06.025.
- Nicolas, H. & Walter, C. (2006). Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma*, 134(1-2), 217-230. Doi: 10.1016/j.geoderma.2005.10.009.

- Nield, S. J., Boettinger, J. L. & Ramsey, R. D. (2007). Digitally mapping gypsic and natric soil areas using Landsat ETM data. *Soil Science Society of America Journal*, 71(1), 245-252. Doi: 10.2136/sssaj2006-0049.
- Nielsen, D. R. & Bouma, J. (1985). *Soil spatial variability*. proceedings of a workshop of the ISSS (Int. Society of Soil Science) and the SSSA (Soil Science Society of America), Las Vegas (USA), 30 Nov-1 Dec 1984.
 - Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J. & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9), 503-510. Doi: 10.1016/j.tree.2005.05.011.
 - Pijanowski, B. C., Brown, D. G., Shellito, B. A. & Manik, G. A. (2002). Using neural networks and GIS to forecast land use changes: a land transformation model. *Computers, Environment and Urban Systems*, 26(6), 553-575. Doi: 10.1016/S0198-9715(01)00015-1.
 - Plumb, A. P., Rowe, R. C., York, P. & Brown, M. (2005). Optimisation of the predictive ability of artificial neural network (ANN) models: a comparison of three ANN programs and four classes of training algorithm. *European Journal of Pharmaceutical Sciences*, 25(4-5), 395-405. Doi: 10.1016/j.ejps.2005.04.010.
 - Rajaei, T., Ebrahimi, H. & Nourani, V. (2019). A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology*, 572, 336-351. Doi: 10.1016/j.jhydrol.2018.12.037.
 - Rondeaux, G., Steven, M. & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55(2), 95-107. Doi: 10.1016/0034-4257(95)00186-7.
 - Sabahattin I., Latif K., Jon, S., Puneet, S. & Graeme, L. (2013). Modeling effects of changing land use/cover on daily streamflow: An Artificial Neural Network and curve number based hybrid approach. *Journal of Hydrology*, 485, 103-112.
 - Sakai, N., Mohamad, Z. F., Nasaruddin, A., Abd Kadir, S. N., Salleh, M. S. A. M. & Sulaiman, A. H. (2018). Eco-Heart Index as a tool for community-based water quality monitoring and assessment. *Ecological Indicators*, 91, 38-46. Doi: 10.1016/j.ecolind.2018.03.079.
 - Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Doll, P. & Portmann, F. T. (2010). Groundwater use for irrigation—a global inventory. *Hydrology and Earth System Sciences*, 14(10), 1863-1880. Doi: 10.5194/hess-14-1863-2010.
 - Taloor, A. K., Pir, R. A., Adimalla, N., Ali, S., Manhas, D. S., Roy, S. & Singh, A. K. (2020). Spring water quality and discharge assessment in the Basantar watershed of Jammu Himalaya using geographic information system (GIS) and water quality Index (WQI). *Groundwater for Sustainable Development*, 10, 100364. Doi: 10.1016/j.gsd.2020.100364.
 - Tung, T. M. & Yaseen, Z. M. (2020). A survey on river water quality modelling using artificial intelligence models: 2000–2020. *Journal of Hydrology*, 585, 124670. Doi: 10.1016/j.jhydrol.2020.124670.
 - Tyagi, S., Sharma, B., Singh, P. & Dobhal, R. (2013). Water quality assessment in terms of water quality index. *American Journal of Water Resources*, 1, 34-38. Doi: 10.12691/ajwr-1-3-3.
 - Wagh, V. M., Panaskar, D. B., Muley, A. A., Mukate, S. V., Lolage, Y. P. & Aamalawar, M. L. (2016). Prediction of groundwater suitability for irrigation using artificial neural network model: a case study of Nanded tehsil, Maharashtra, India. *Modeling Earth Systems and Environment*, 2(4), 1-10. Doi: 10.1007/s40808-016-0250-3.
 - Wilson, E. H. & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, 80(3), 385-396. Doi: 10.1016/S0034-4257(01)00318-2.
 - Wu, J., Zhang, Y. & Zhou, H. (2020). Groundwater chemistry and groundwater quality index incorporating health risk weighting in Dingbian County, ordos basin of northwest China. *Geochemistry*, 80(4), 125607. Doi: 10.1016/j.chemer.2020.125607.
 - Yaseen, Z. M., Sulaiman, S. O., Deo, R. C. & Chau, K. W. (2019). An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. *Journal of Hydrology*, 569, 387-408. Doi: 10.1016/j.jhydrol.2018.11.069.
 - Zhu, J. J., Yu, L. Z., Xu, T. L., Wei, X. & Yang, K. (2019). Comparison of water quality in two catchments with different forest types in the headwater region of the Hun River, Northeast China. *Journal of Forestry Research*, 30(2), 565-576. Doi: 10.1007/s11676-018-0688-4.