

Modeling Soil Available Water Using Neuro-fuzzy, Gene Expression Programming and Random Forest Methods

M. Hasanpour Kashani^{1*}, Sh. Asghari²

1- Department of Water Engineering, Water Management Research Center, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran

(*- Corresponding Author Email: m.hkashani@uma.ac.ir)

2- Department of Soil Sciences and Engineering, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran

Received: 25 February 2025
Revised: 09 April 2025
Accepted: 14 May 2025
Available Online: 14 May 2025

How to cite this article:

Hasanpour Kashani, M., & Asghari, Sh. (2025). Modeling soil available water using neuro-fuzzy, gene expression programming and random forest methods. *Journal of Water and Soil*, 39(2), 139-154. (In Persian with English abstract).
<https://doi.org/10.22067/jsw.2025.92342.1470>

Introduction

Soil available water (SAW) is defined as the difference between field capacity (FC) and permanent wilting point (PWP). FC is the amount of soil water content held by the soil after the gravitational water was drained from the soil. PWP is defined as a minimum water content of a soil which is needed for the crop survival and if the water content decreases lower than PWP, a plant wilts and can no longer recover itself. The direct measurement of FC and PWP soil water contents is very costly and time consuming; therefore, it is useful the use of different intelligent models such as neuro-fuzzy (NF), gene expression programming (GEP) and random forest (RF) to estimate FC, PWP and SAW through easily accessible and low-cost soil characteristics. The objectives of this research were: (1) to obtain NF, GEP and RF models for estimating SAW from the easily accessible soil variables in the cultivated lands of Ardabil plain, and (2) to compare the accuracy of the mentioned models in estimating SAW using the coefficient of determination (R^2), root mean square error (RMSE), mean error (ME) and Nash-Sutcliffe coefficient (NS) criteria.

Materials and Methods

The measured data from 102 soil samples taken from 0-10 cm soil depth of the cultivated lands of Ardabil plain, northwest of Iran, were used in this study. Sand, clay, mean geometric diameter (d_g) and geometric standard deviation (σ_g) of soil particles, bulk density (BD) and organic carbon (OC) were introduced as input variables to the applied three intelligent models for estimating soil available water (SAW). Data randomly were divided in two series as 82 data for training and 20 data for testing of models. In all models, six different input variables combinations were used; SPSS 22 software with stepwise method was applied to select the input variables. MATLAB, Gene Xpro Tools 4.0 and Weka softwares were used to derive neuro-fuzzy (NF), gene expression programming (GEP) and random forest (RF) models, respectively. One of the important steps by using NF method is selecting the appropriate membership functions (MFs) and its numbers. Based on a trial and error procedure, 3 numbers of MFs and 50 to 100 optimum replications were found for the NF modeling. Also, the input MFs were chosen as “triangular”, “trapezoid”, “generalized bell” and “pi” and the output MF was selected as “constant”. A set of optimal parameters were chosen before developing a best GEP model. The number of chromosomes and genes, head size and linking function were selected by the trial and error method, and they are 30, 3, 8, and +,



©2025 The author(s). This is an open access article distributed under [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

<https://doi.org/10.22067/jsw.2025.92342.1470>

respectively. The rates of genetic operators were chosen according to literature studies. Various tree numbers were analyzed for choosing the best random forest (RF) method. Increasing the tree numbers beyond 100 made lower variations in the average squared error values for the SAW estimation cases. The accuracy of NF, GEP and RF models in estimating SAW was evaluated by coefficient of determination (R^2), root mean square error (RMSE), mean error (ME) and Nash-Sutcliffe coefficient (NS) statistics.

Results and Discussion

The studied soils were loam ($n=53$), clay loam ($n=26$), sandy loam ($n=15$), silt loam ($n=6$) and clay ($n=2$) textural classes. The values of sand (24.40 to 68.00 %), clay (3.80 to 42.90 %), dg (0.02 to 0.26 mm), σ_g (7.48 to 19.41), BD (1.04 to 1.70 g cm⁻³), OC (0.31 to 1.52 %) and SAW (5.10 to 25.10 % g g⁻¹) indicated good variations in the soils of studied region. Significant correlations were found between SAW and BD ($r = -0.59$), clay ($r = 0.56^{**}$), OC ($r = 0.45^{**}$), and sand ($r = -0.44^{**}$). NF, GEP and RF models were applied to estimate SAW using six different combinations of input soil variables (sand, clay, dg, σ_g , BD and OC). The results of the best NF, GEP and RF models indicated that the most appropriate input variables to predict SAW were OC and BD. The values of R^2 , RMSE, ME and NS criteria were obtained equal 0.73, 2.51 % g g⁻¹, 0.09 % g g⁻¹ and 0.71, and 0.76, 3.10 % g g⁻¹, -1.41 % g g⁻¹ and 0.56, 0.68, 3.30 % g g⁻¹, -1.45 % g g⁻¹, 0.50 for the best NF, GEP and RF models in the testing data set, respectively. Numerous investigations also showed that there is significant negative correlation between SAW with BD and sand and positive correlation between SAW with OC and clay.

Conclusion

The results from the three investigated intelligent models indicated that organic carbon (OC) and bulk density (BD) were the most important and readily available soil variables for predicting soil available water (SAW) in the study area. Among the models, the Neuro-Fuzzy (NF) approach demonstrated the highest accuracy, as evidenced by the lowest root mean square error (RMSE) and the highest Nash–Sutcliffe efficiency (NS) values. In contrast, the Random Forest (RF) model provided the least accurate estimates of SAW, performing worse than both the NF and Gene Expression Programming (GEP) models.

Keywords: Intelligent models, Prediction, Readily variables, Soil water

مدل سازی آب قابل استفاده خاک با استفاده از روش های نروفازی، برنامه ریزی بیان ژن و جنگل تصادفی

مهسا حسینیور کاشانی^{۱*} - شکراله اصغری^۲

تاریخ دریافت: ۱۴۰۳/۱۲/۰۷

تاریخ پذیرش: ۱۴۰۴/۰۲/۲۴

چکیده

آب قابل استفاده خاک (SAW) از تفاضل دو نقطه مهم و کاربردی منحنی رطوبتی خاک (رطوبت های ظرفیت مزرعه ای و پژمردگی دائم) به دست می آید؛ روش مستقیم اندازه گیری این نقاط رطوبتی خاک مستلزم صرف وقت و هزینه زیادی می باشد. بنابراین اخیراً از روش های غیر مستقیم مانند توابع انتقالی رگرسیونی و مدل های هوشمند برای تخمین این متغیرهای دیربافت خاک استفاده گردیده است. هدف از این پژوهش، مقایسه عملکرد سه مدل هوشمند نروفازی (NF)، برنامه ریزی بیان ژن (GEP) و جنگل تصادفی (RF) در برآورد SAW از روی ویژگی های زودبافت خاک بود. در این تحقیق، داده های اندازه گیری شده مربوط به تعداد ۱۰۲ نمونه خاک سطحی دشت اردبیل شامل مقادیر ذرات شن، سیلت، رس، میانگین و انحراف معیار هندسی قطر ذرات، جرم مخصوص ظاهری، کربن آلی و SAW مورد استفاده قرار گرفت. داده ها به طور تصادفی به دو گروه آموزشی (۸۲ نمونه) و آزمونی (۲۰ نمونه) تقسیم شدند. شش تابع SAW با ترکیب متفاوت از متغیرهای زودبافت خاک به عنوان ورودی مدل انتخاب و مدل های NF، GEP و RF برای هر کدام اجرا گردید. نتایج مدل ها بیانگر آن بود که کربن آلی و جرم مخصوص ظاهری خاک، دو متغیر ورودی مهم و اثرگذار در برآورد SAW بودند. مقادیر آماره های ضریب تبیین (R^2)، مجذور میانگین مربعات خطا (RMSE)، میانگین خطا (ME) و ضریب نش-ساتکیلف (NS) براساس داده های آزمونی برابر 0.73 ، 0.51 ، 0.09 ، 0.71 ، 0.76 ، 3.10 ، 1.41 ، 0.56 ، 0.68 ، 0.30 ، 3.30 ، 1.45 و 0.50 به ترتیب برای بهترین مدل NF، GEP و RF تعیین شد. مقایسه نتایج آماره های ارزیابی مدل ها نشان داد که مدل NF به علت داشتن مقدار کم RMSE، مقدار نزدیک به صفر ME و مقدار زیاد NS نسبت به دو مدل دیگر از خطای پایین و دقت بالا در برآورد SAW برخوردار بود.

واژه های کلیدی: آب خاک، برآورد، متغیر زودبافت، مدل های هوشمند

(2017; Cheema, Shah, Farooque, et al., 2024).

مقدمه

آب قابل استفاده خاک (SAW) از اختلاف رطوبت های ظرفیت زراعی^۴ (FC) و نقطه پژمردگی دائم^۵ (PWP) به دست می آید. رطوبت FC عبارت از مقدار آبی است که پس از خروج آب ثقی در خاک نگهداری می شود. رطوبت PWP عبارت از مقدار آبی است که در مقادیر

اطلاع از وضعیت آب قابل استفاده خاک^۳ (SAW) از نظر برنامه ریزی آبیاری، زهکشی و احیای اراضی، تجزیه و تحلیل فعالیت زیستی خاک، شبیه سازی رواناب سطحی، نیاز آبخویی و شبیه سازی رشد گیاه حائز اهمیت فراوان است (Shiri, Keshavarzi, Kisi, & Karimi, 2024).

۱- گروه مهندسی آب، پژوهشکده مدیریت آب، دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران
(Email: m.hkashani@uma.ac.ir)

۲- گروه علوم و مهندسی خاک، دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران

با مدل شبکه‌های عصبی پرسپترون چند لایه و مدل GEP از دقت زیاد و خطای کم در برآورد منحنی رطوبتی خاک در ۱۰۸ نمونه خاک جنگلی منتخب از فنلاند برخوردار بودند. شجاعی و همکاران (Shojaei et al., 2023) در خاک‌های منتخب از منطقه روانسر کرمانشاه گزارش کردند مدل کیوبیست و مدل ANN در مقایسه با مدل‌های MLR، RF و SVM دارای دقت بیشتر و خطای کمتر در تخمین FC و PWP بودند. چیم و همکاران (Cheema et al., 2024) از مدل GEP برای برآورد منحنی رطوبتی در خاک‌های منتخب از کانادا با سه نوع کلاس بافت شن لومی، لوم و لوم شنی استفاده کردند. برای این منظور، داده‌های شن، سیلت، رس، جرم مخصوص ظاهری، کربن آلی و مکش خاک به عنوان متغیرهای ورودی مدل در نظر گرفته شد؛ نتایج نشان داد مدل GEP توانست منحنی رطوبتی خاک را در محدوده خشک (۲۰ درصد حجمی) تا اشباع (۷۰ درصد حجمی) با دقت بالا و خطای پایین تخمین بزند. کیانی و بیات (Kiani & Bayat, 2024) گزارش نمودند که مدل RF در مقایسه با مدل MLR توانست باعث افزایش صحت تخمین منحنی رطوبتی خاک به میزان ۲۲ تا ۴۶ درصد در ۱۲۰ نمونه خاک منتخب از استان‌های تهران و همدان گردد.

بررسی منابع انجام گرفته در این پژوهش نشان داد اگرچه توابع انتقالی رگرسیونی و شبکه عصبی مصنوعی برای برآورد FC، PWP و SAW در خاک‌های دشت اردبیل توسط امیرعابدی و همکاران (Amirabedi et al., 2013) ارائه شده است ولی بر اساس اطلاعات موجود، کارایی سایر مدل‌های هوشمند مانند نروفازی، برنامه‌ریزی بیان ژن و جنگل تصادفی در برآورد SAW در خاک‌های استان اردبیل تاکنون بررسی نشده است. از سوی دیگر، توابع انتقالی خاک روابط تجربی هستند و استفاده از مدل‌های هوشمند توسعه‌یافته توسط سایر محققان برای برآورد ضرایب رطوبتی خاک (FC، PWP و SAW) در خاک‌های دشت اردبیل با توجه به مواد مادری و منطقه اقلیمی متفاوت این خاک‌ها، نیاز به راستی‌آزمایی دارد. بنابراین، هدف اصلی پژوهش حاضر عبارت بود از: ارائه مدل‌های NF، GEP و RF برای برآورد SAW از روی متغیرهای زودیافت در خاک‌های دشت اردبیل و نیز مقایسه دقت مدل‌های مذکور در برآورد SAW با استفاده از معیارهای ضریب تبیین (R^2)، مجذور میانگین مربعات خطا ($RMSE$)، میانگین خطا (ME)^{۱۱} و ضریب نش-ساتکلیف (NS)^{۱۲}.

کمتر از آن رشد گیاه متوقف می‌شود (Jury & Horton, 2004). اندازه‌گیری مستقیم FC، PWP و SAW کاری بسیار پرهزینه و زمان‌بر می‌باشد بنابراین تلاش‌های زیادی انجام شده است تا این نقاط مهم و پرکاربرد منحنی رطوبتی خاک به جای اندازه‌گیری مستقیم، از روی ویژگی‌های زودیافت و کم‌هزینه خاک مانند شن، سیلت، رس، میانگین (d_g) و انحراف استاندارد (σ_g) هندسی قطر ذرات، جرم مخصوص ظاهری و کربن آلی تخمین زده شود (مانند Amirabedi, Asghari, Mesri, & Keivan behjo, 2013; Shiri et al., 2017; Jafari, Ojaghlou, & Karbasi, 2021; Shojaei, Farhadi, Bansoulehi, Fatehi, & Rahmani, 2023).

تاکنون توابع انتقالی متعددی مبتنی بر مدل‌های رگرسیون خطی چندگانه^۱ (MLR) و رگرسیون غیر خطی چندگانه^۲ (MNLR) و نیز مدل‌های هوشمند مانند شبکه عصبی مصنوعی^۳ (ANN)، نروفازی^۴ (NF)، برنامه‌ریزی بیان ژن^۵ (GEP) و جنگل تصادفی^۶ (RF) برای برآورد نقاط رطوبتی مهم خاک (FC، PWP و SAW) از طریق ویژگی‌های سهل‌الوصول و کم‌هزینه خاک توسعه یافته‌اند. مردون و همکاران (Merdun, Cinar, Meral, & Apan, 2006) از مدل‌های MLR و ANN برای برآورد FC و PWP در ۱۹۵ نمونه خاک استفاده کرده و دریافتند که جرم مخصوص ظاهری و d_g از متغیرهای ورودی مهم و اثرگذار بودند. امیرعابدی و همکاران (Amirabedi et al., 2013) دقت مدل‌های MLR و ANN را در برآورد FC، PWP و SAW در ۱۰۲ نمونه خاک منتخب از دشت اردبیل مقایسه و بررسی نمودند؛ نتایج بیانگر دقت بالا و خطای پایین مدل ANN در برآورد هر سه ضریب رطوبتی مذکور بود. استواری و همکاران (Ostovari, Asgari, & Cornelis, 2015) با به‌کارگیری مدل‌های MLR در ۲۵۵ نمونه خاک دریافتند رس و d_g از متغیرهای ورودی مهم در تخمین FC و PWP بودند. شیر و همکاران (Shiri et al., 2017) نتایج مدل‌های MLR، NF، GEP، RF، ماشین بردار پشتیبان^۷ (SVM)، اسپلین رگرسیون تطبیقی چند متغیره^۸ (MARS) و مدل درختی^۹ (MT) را در برآورد FC و PWP از روی متغیرهای ورودی خاک شامل رس، سیلت، جرم مخصوص ظاهری و d_g در ۱۹۲ نمونه خاک منتخب از اراضی مرتعی و زراعی استان فارس بررسی کردند؛ نتایج نشان داد که مدل NF دارای بالاترین و مدل MT دارای ضعیف‌ترین صحت تخمین در بین مدل‌های مورد مطالعه بود. نتایج پژوهش جعفری و همکاران (Jafari et al., 2021) بیانگر آن بود که مدل شبکه‌های عصبی آبخاری در مقایسه

7- Support vector machine
8- Multivariate adaptive regression spline
9- Model tree
10- Root mean square error
11- Mean error
12- Nash-Sutcliffe coefficient

1- Multiple linear regression
2- Multiple nonlinear regression
3- Artificial neural network
4- Neuro-fuzzy
5- Gene expression programming
6- Random forest

مواد و روش ها

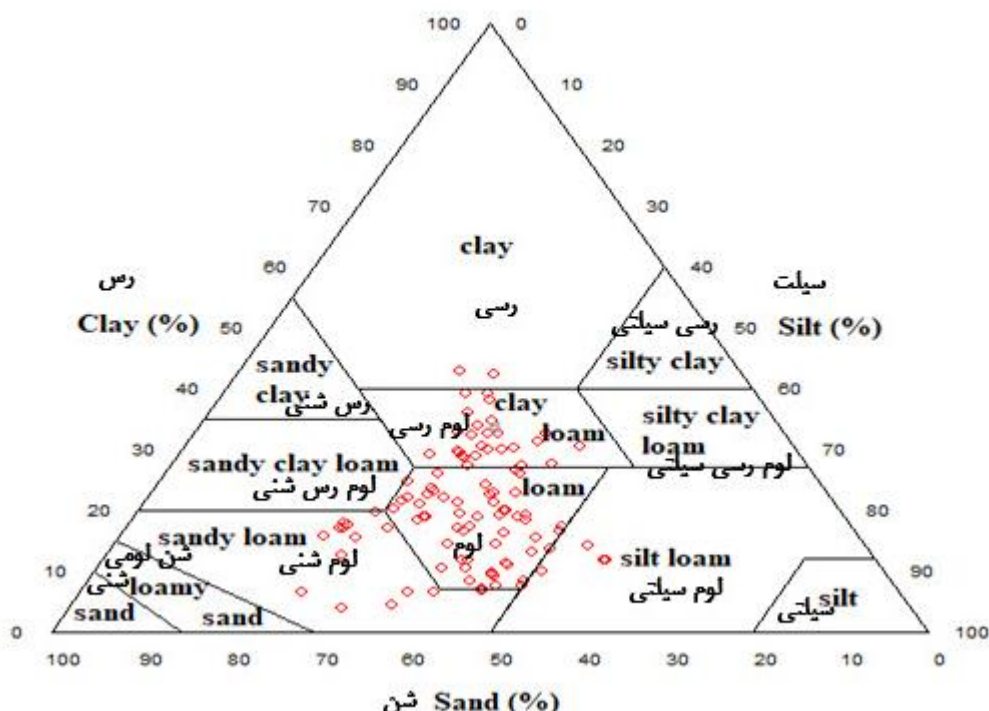
در این پژوهش، از داده های اندازه گیری شده توزیع اندازه ذرات، جرم مخصوص ظاهری، کربن آلی، رطوبت های ظرفیت زراعی (FC) و نقطه پژمردگی دائم (PWP) مربوط به تعداد ۱۰۲ نمونه خاک که از افق Ap اراضی زراعی دشت اردبیل با مساحت حدود ۲۲۰۰ کیلومتر مربع برداشته شده بود، استفاده گردید (Amirabedi et al., 2013). درصد های ذرات شن، سیلت و رس به روش هیدرومتری ۴ قرائته (Gee & Or, 2002) و کلاس بافت خاک (شکل ۱) بر اساس استاندارد وزارت کشاورزی آمریکا تعیین و پارامتر های میانگین (d_g) و انحراف استاندارد (σ_g) هندسی قطر ذرات خاک با استفاده از روابط ارائه شده توسط شیرازی و بورسما (Shirazi & Boersma, 1984) محاسبه گردید. کربن آلی به روش اکسیداسیون تر (Nelson & Sommers, 1982) و جرم مخصوص ظاهری (BD) در استوانه های استیل (Blake & Hartge, 1986) به قطر و ارتفاع ۵ سانتی متر اندازه گیری شد. اندازه گیری های رطوبت خاک در مکش ۰/۱ بار (معادل رطوبت FC در خاک های درشت بافت)، ۰/۲ بار (معادل رطوبت FC در خاک های متوسط بافت) و ۰/۳ بار (معادل رطوبت FC در خاک های ریزبافت) در استوانه های استیل و در مکش ۱۵ بار (معادل رطوبت PWP در همه خاک ها) در نمونه های خاک گذرانده شده از الک ۲ میلی متری با استفاده از دستگاه صفحات فشاری و پس از آون خشک کردن در دمای 105°C به روش وزنی انجام گرفت (Gardner, 1986). آب قابل استفاده خاک (SAW) از اختلاف رطوبت های FC و PWP محاسبه گردید (Bauer & Black, 1992). برخی ویژگی های نمونه های خاک مورد استفاده در پژوهش حاضر در جدول ۱ آورده شده است. با توجه به شکل ۱ مشاهده می شود که خاک ها از تنوع بافتی نسبتاً مناسب (۵۳ نمونه لوم، ۲۶ نمونه لوم رسی، ۱۵ نمونه لوم شنی، ۶ نمونه لوم سیلتی و ۲ نمونه رسی) برخوردار هستند.

نرمال بودن توزیع داده ها با آزمون کولموگروف - اسمیرنوف بررسی گردید؛ آزمون نرمال بودن نشان داد داده های همه متغیر های مورد مطالعه از توزیع نرمال برخوردار هستند. محاسبات مربوط به آمار توصیفی و ضریب همبستگی پیرسون (جدول ۲) بین متغیر ها و نیز اشتقاق توابع رگرسیونی خطی چندگانه (MLR) بین متغیر های زود یافت خاک با متغیر دیر یافت (SAW) به روش گام به گام^۱ در نرم افزار SPSS 22 انجام شد. بر اساس روش گام به گام، شش مدل رگرسیونی (جدول ۳) با ترکیب های متفاوتی از متغیر های زود یافت ورودی شامل شن، سیلت، رس، جرم مخصوص ظاهری، کربن آلی، d_g و σ_g برای برآورد متغیر دیر یافت (SAW) ایجاد گردید (Amirabedi et al., 2013). به منظور بررسی و مقایسه دقیق نتایج پژوهش حاضر، در فرآیند مدل

سازی SAW به سه روش هوشمند نروفازی (NF)، برنامه ریزی بیان ژن (GEP) و جنگل تصادفی (RF)، از همان ترکیب های متفاوت متغیر های ورودی در مدل های رگرسیونی ارائه شده توسط امیرعابدی و همکاران (Amirabedi et al., 2013) استفاده شد.

روش نروفازی (NF) ترکیبی از شبکه عصبی تطبیقی و سیستم استنتاج فازی می باشد که بر اساس قوانین اگر - آنگاه پایه ریزی شده است. پارامتر های سیستم فازی با به کارگیری الگوریتم های آموزشی شبکه عصبی، تشخیص داده می شوند (Shiri et al., 2017). در این پژوهش، مدل های نروفازی (NF) برای برآورد آب قابل استفاده خاک (SAW) با استفاده از نرم افزار MATLAB ایجاد شدند. بهترین ساختار مدل های نروفازی به روش سعی و خطا و با به کارگیری توابع عضویت و تکرار های مختلف برای هر کدام از مدل ها تعیین گردید (Ghorbani, 2019). به این ترتیب که توابع عضویت مثلثی (trimf)، دوزنقه ای (trapmf)، زنگوله ای (gbellmf) و پی شکل (pimf) به عنوان بهترین توابع عضویت لایه ورودی و تابع عضویت ثابت (Constant) به عنوان بهترین تابع عضویت لایه خروجی انتخاب شدند. همچنین، تعداد تابع عضویت ۳ و تعداد تکرار بهینه بین ۵۰ تا ۱۰۰ برای اجرای شش مدل NF به کار رفته در روش نروفازی تعیین گردید.

مدل برنامه ریزی بیان ژن (GEP) یک الگوریتم توسعه یافته بر اساس برنامه ریزی ژنتیک و الگوریتم ژنتیک می باشد و مبنای تمامی آنها بر اساس نظریه تکامل داروین استوار است که در آن کروموزوم های خطی و ساده با طول ثابت مشابه در الگوریتم ژنتیک استفاده می شود و ساختار های شاخه ای با اندازه ها و اشکال متفاوت مشابه با برنامه ریزی ژنتیک ترکیب می شوند. جزئیات بیشتر درباره فرآیند اجرای مدل GEP را می توان در مقاله زانگ و زانگ (Zhang & Zhang, 2024) پیدا نمود. در این پژوهش برای مدل سازی آب قابل استفاده خاک (SAW) بر اساس برنامه ریزی بیان ژن (GEP) از نرم افزار Gene Xpro Tools 4.0 استفاده گردید (Ferreira, 2006). در روش GEP، پدیده های مختلف با استفاده از مجموعه ای از توابع و ترمینال ها، مدل سازی می شوند. از مجموعه توابع اصلی حسابی، مثلثاتی و توابع تعریف شده توسط کاربر بر اساس سعی و خطا تا حصول حداکثر ضریب تبیین (R^2) و حداقل مجذور میانگین مربعات خطا (RMSE) در برازش مدل استفاده می شود. عملگر های ریاضی پر کاربرد در فرآیند اجرای مدل شامل چهار عمل اصلی (+، -، × و /)، جذر، لگاریتم نپری، لگاریتم طبیعی، توان ۲ و ۳ و ریشه سوم هستند. خلاصه پارامتر های لازم برای اجرای مدل GEP در جدول ۴ ارائه شده است.



شکل ۱- توزیع کلاس بافت در خاک‌های مورد مطالعه

Figure 1- Distribution of textural class in the studied soils

جدول ۱- آماره‌های توصیفی متغیرهای خاک در منطقه مورد مطالعه

Table 1- Descriptive statistics of soil variables in the studied area

متغیر Variable	آموزش Training (n= 82)				آزمون Testing (n= 20)			
	Min.	Max.	Mean	Sd.	Min.	Max.	Mean	Sd.
BD (g cm ⁻³)	1.04	1.70	1.35	0.15	1.15	1.60	1.39	0.15
Sand (%)	24.40	68.00	41.45	8.31	28.30	65.00	44.40	10.96
Silt (%)	24.40	57.60	37.93	7.75	23.30	49.90	36.40	8.43
Clay (%)	4.40	42.90	20.61	9.31	3.80	42.20	19.18	9.04
d _g (mm)	0.02	0.26	0.07	0.04	0.02	0.25	0.09	0.06
σ _g (-)	8.02	19.41	13.69	3.05	7.48	17.48	13.31	2.73
OC (%)	0.31	1.52	0.85	0.26	0.37	1.10	0.70	0.20
SAW (% g g ⁻¹)	5.10	20.40	13.72	3.21	5.10	25.1	14.04	4.78

BD: جرم مخصوص ظاهری (bulk density)، clay: رس، silt: سیلت، sand: شن، d_g: میانگین هندسی قطر ذرات (geometric mean diameter)، σ_g: انحراف استاندارد هندسی ذرات (geometric standard deviation)، OC: کربن آلی (organic carbon)، SAW: آب قابل استفاده خاک (soil available water)، Sd: انحراف استاندارد (Standard deviation)، Mean: میانگین، Max: حداکثر، Min: حداقل.

جدول ۲- ضریب همبستگی پیرسون (r) بین آب قابل استفاده خاک و متغیرهای ورودی مدل‌ها (n=102)

Table 2- Pearson correlation coefficient (r) between soil available water and input variables of models (n=102)

متغیر Variable	Sand	Silt	Clay	d _g	σ _g	OC	BD
SAW	-0.44*	-0.08 ^{ns}	0.56**	-0.42*	0.28*	0.45**	-0.59**

* و **: به ترتیب معنی‌دار در سطح احتمال ۱ و ۵ درصد. توصیف علائم متغیرها در زیر جدول ۱ آورده شده است.

Variable symbols description is available under Table 1. *, **: Significant at P < 0.05 and P < 0.01, respectively.

جدول ۳- مدل های رگرسیون خطی چندگانه (MLR) آب قابل استفاده خاک (Amirabedi et al., 2013)

Table 3- MLR models of soil available water (SAW) (Amirabedi et al., 2013)

شماره Number	مدل Model
1	SAW = 5.03 OC - 9.97 BD + 21.01
2	SAW = 5.35 OC - 6.97 BD + 21.01 + 0.09 Clay + 14.85
3	SAW = - 0.12 Sand + 5.55 OC + 11.99
4	SAW = - 0.08 Sand - 10.22 BD + 28.91
5	SAW = - 7.32 BD + 5.08 OC - 36.87 dg + 18.68
6	SAW = - 7.19 BD + 5.73 OC + 0.28 σ_g + 12.83

توصیف متغیرها در زیر جدول آورده شده است

The description of the variables is available in Table 1.

جدول ۴- مقادیر پارامترهای مدل برنامه ریزی بیان ژن (GEP)

Table 4- The values of GEP model parameters

مقادیر Values	عوامل ژنتیک Genetic factors	مقادیر Values	تنظیمات عمومی General adjustment
0.044	Mutation rate نرخ جهش	30	تعداد کروموزمها Number of chromosome
0.1	Inversion rate نرخ وارون سازی	8	اندازه سر Head size
0.3	One point recombination rate نرخ ترکیب تک نقطه ای	3	تعداد ژن ها Number of genes
0.3	Two point recombination rate نرخ ترکیب دو نقطه ای	جمع (+)	تابع پیوند Linking function
0.1	Gene recombination rate نرخ ترکیب ژن	RMSE	معیار خطای تابع Fitness function error type

برآورد می باشد. از لحاظ آماری کمترین مقدار RMSE، صفر می باشد که در آن وضعیت، مقادیر اندازه گیری و برآورد شده SAW باهم برابر هستند؛ هر چه مقدار RMSE کوچکتر باشد برآوردهای انجام شده توسط مدل برآورد کننده دقیق تر است. مقادیر مثبت ME بیش برآوردی و مقادیر منفی آن کم برآوردی مدل را نشان می دهد. اگر ضریب NS برابر یک باشد بیانگر آن است که مدل سازی با دقت بالایی انجام گرفته و مقادیر اندازه گیری و برآورد شده SAW در تطابق کامل هستند. مقدار برابر با صفر NS نشان دهنده آن است که برآوردهای مدل، دقتی برابر با دقت میانگین داده های اندازه گیری شده دارند.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{i(p)} - Y_{i(m)})^2}{n}} \quad (1)$$

$$ME = \frac{\sum_{i=1}^n Y_{i(p)} - Y_{i(m)}}{n} \quad (2)$$

$$NS = 1 - \frac{\sum_{i=1}^n (Y_{i(p)} - Y_{i(m)})^2}{\sum_{i=1}^n (Y_{i(m)} - \bar{Y}_m)^2} \quad (3)$$

در معادله های ۱، ۲ و ۳، $Y_{i(p)}$ ، $Y_{i(m)}$ و $\bar{Y}_m$ به ترتیب مقادیر اندازه گیری، برآورد شده و میانگین مقادیر اندازه گیری شده SAW توسط مدل و n تعداد داده ها برای SAW در هر مدل می باشد. از میان مدل های ایجاد شده برای برآورد SAW، مناسب ترین مدل NF، GEP و RF با ترکیب متغیرهای ورودی در داده های آموزشی براساس محک

مدل جنگل تصادفی (RF) اولین بار توسط بریمن (Breiman, 2001) ارائه شد. جنگل تصادفی یک الگوریتم یادگیری گروهی می باشد که مسائل رگرسیونی با ابعاد بزرگ را مدیریت می کند؛ این مدل یک روش گروهی بر پایه درختی است که در آن همه درختان به مجموعه ای از متغیرهای تصادفی وابسته هستند و جنگل از درختان رگرسیونی زیاد واقع در کنار یکدیگر رشد می یابد (Shiri et al., 2017). در مدل RF برای هر درخت یک زیر مجموعه ای از متغیرهای ورودی اعمال می شود. تعداد ورودی ها در هر درخت و نیز تعداد درختان بهینه، به روش سعی و خطا تعیین می گردد. در این پژوهش برای اجرای مدل RF از نرم افزار Weka استفاده شد؛ بهترین اجرای مدل RF برای برآورد آب قابل استفاده خاک (SAW)، با ۲ متغیر کمکی موجود در زیر مجموعه تصادفی هر گره (mtree= 2) و تعداد ۱۰۰ درخت به دست آمد.

در همه مدل های به کار رفته در تحقیق حاضر از مجموع ۱۰۲ داده برای هر متغیر، ۸۲ داده برای آموزش مدل و ۲۰ داده برای آزمون مدل به طور تصادفی انتخاب گردید. دامنه تغییرات داده های سری آموزش و سری آزمون در جدول ۱ بیانگر آن است که متغیرهای استفاده شده در پژوهش حاضر از پراکنش خوبی برخوردار هستند همچنین میانگین متغیرها در سری آموزش و سری آزمون تقریباً نزدیک هم می باشد. برای ارزیابی دقت و کارایی مدل های NF، GEP و RF در برآورد SAW، از آماره های R^2 ، RMSE، ME و NS (Ghorbani et al., 2019) استفاده شد. مقادیر بالای R^2 (نزدیک به ۱) بیانگر دقت بالایی

های آماری مذکور تعیین گردید. آزمون مدل‌ها در تشخیص دقت مدل به ازای داده‌هایی که در آموزش مدل استفاده نشده‌اند به کار می‌رود.

نتایج و بحث

نتایج مدل‌های نروفازی (NF) برای برآورد آب قابل استفاده خاک (SAW)

برای برآورد آب قابل استفاده خاک (SAW)، شش مدل نروفازی (NF) براساس متغیرهای ورودی استفاده شده در توابع رگرسیونی (جدول ۳) ایجاد گردید. در جدول ۵ مقادیر معیارهای ضریب تبیین (R^2)، مجذور میانگین مربعات خطا (RMSE)، میانگین خطا (ME) و ضریب نش-ساتکلیف (NS) برای ارزیابی مدل‌های NF از نظر دقت برآورد SAW ارائه شده است. ممکن است به کارگیری R^2 به تنهایی برای سنجش اعتبار مدل‌ها گول‌زننده باشد و الزاماً R^2 نزدیک به یک شاید بیانگر برازش دقیق مقادیر برآورد و اندازه‌گیری شده یک متغیر دریافت نباشد؛ به عبارت دیگر، بالا بودن R^2 همیشه نشان‌دهنده پایین بودن خطا نیست (Kozak, Pachepsky, Sokolowski, Sokolowska, & Stepniewski, 1996) لذا در این پژوهش، رتبه-بندی هر سه مدل هوشمند مورد استفاده (NF، GEP و RF) از نظر دقت برآورد SAW، بر اساس پایین بودن RMSE و بالا بودن NS در داده‌های آزمون انجام شد. واسیم و همکاران (Waseem, Mani, Andiego, & Usman, 2017) درجه‌بندی زیر را برای آماره نش-ساتکلیف (NS) از نظر ارزیابی عملکرد برآورد مدل ارائه نمودند: مقدار NS بین ۱ و ۰/۷۵ (بسیار خوب)، مقدار NS بین ۰/۷۵ و ۰/۶۵ (خوب)، مقدار NS بین ۰/۶۵ و ۰/۵۰ (قابل قبول) و مقدار NS کوچکتر ۰/۵۰ (نامطلوب). بر مبنای پایین بودن آماره RMSE و بالا بودن آماره NS در سری آزمون، مشاهده گردید مدل NF شماره ۲ (جدول ۵) با سه متغیر ورودی کربن آلی، جرم مخصوص ظاهری و رس از نظر برآورد SAW دارای دقت بالاتری در مقایسه با سایر مدل‌های NF بود. مقادیر R^2 ، RMSE، ME و NS این مدل برای داده‌های آموزشی به ترتیب برابر ۰/۶۶، ۱/۸۷ % $g\ g^{-1}$ و ۰/۰ % و برای داده‌های آزمون برابر ۰/۷۳، ۲/۵۱ % $g\ g^{-1}$ و ۰/۰۹ % محاسبه گردید (جدول ۵). مقدار جزئی مثبت ME برای داده‌های آزمون بیانگر برآورد اندکی بیش از مقدار واقعی SAW توسط مدل شماره ۲ می‌باشد و براساس آماره NS، عملکرد این مدل خوب ($NS = 0.71$) ارزیابی می‌شود. همچنین معادله رگرسیونی این مدل (جدول ۳) نشان می‌دهد که کربن آلی و رس با علامت مثبت و جرم مخصوص ظاهری با علامت منفی وارد مدل گردیدند. تأثیر مثبت متغیر کربن آلی به علت داشتن ضریب رگرسیون بالاتر در برآورد SAW به مراتب بیشتر از متغیر رس می‌باشد؛ همبستگی معنی‌دار و مثبت متغیرهای کربن آلی و رس و همبستگی معنی‌دار و منفی متغیر جرم مخصوص ظاهری با SAW

(جدول ۲)، این موضوع را تأیید می‌کند. در رتبه‌های بعدی تخمین، مدل شماره ۱ با دو متغیر ورودی کربن آلی و جرم مخصوص ظاهری ($RMSE = 2.77\ %\ g\ g^{-1}$ و $NS = 0.65$) و مدل شماره ۴ با دو متغیر ورودی جرم مخصوص ظاهری و شن ($RMSE = 2.87\ %\ g\ g^{-1}$ و $NS = 0.62$) از نظر برآورد SAW قرار دارد (جدول ۵)؛ پس در مجموع براساس آماره‌های ارزیابی شش مدل نروفازی فوق، مدل شماره ۲ برای برآورد SAW ارجح‌تر است. در پژوهش قبلی امیرعابدی و همکاران (Amirabedi et al., 2013) نیز بر روی داده‌های تحقیق حاضر، بهترین مدل MLR ($RMSE = 3.07\ %\ g\ g^{-1}$) و بهترین مدل ANN ($RMSE = 3.05\ %\ g\ g^{-1}$) با سه متغیر ورودی کربن آلی، رس و جرم مخصوص ظاهری برای برآورد SAW در خاک‌های زراعی دشت اردبیل تشخیص داده شد؛ بنابراین، استفاده از مدل نروفازی در تحقیق حاضر در مقایسه با مدل‌های رگرسیونی و شبکه عصبی مصنوعی (Amirabedi et al., 2013) به علت کاهش RMSE باعث بهبود تخمین آب قابل استفاده خاک گردید. این نتایج با یافته‌های شیری و همکاران (Shiri et al., 2017) در خصوص کارایی بهتر روش نروفازی (ضرایب تبیین برابر ۰/۸۵ و ۰/۹۶) در مقایسه با روش رگرسیون خطی چندگانه (ضرایب تبیین برابر ۰/۴۱ و ۰/۴۱) در برآورد رطوبت‌های FC و PWP در خاک‌های استان فارس همخوانی دارد.

نتایج مدل‌های برنامه‌ریزی بیان ژن (GEP) برای برآورد آب قابل استفاده خاک (SAW)

شش مدل برنامه‌ریزی بیان ژن (GEP) با همان ترکیب متغیرهای ورودی مدل‌های رگرسیون خطی چندگانه (جدول ۳) برای برآورد آب قابل استفاده خاک (SAW) طراحی گردید. در جدول ۶ مقادیر آماره‌های ضریب تبیین (R^2)، مجذور میانگین مربعات خطا (RMSE)، میانگین خطا (ME) و ضریب نش-ساتکلیف (NS) برای ارزیابی عملکرد مدل‌های GEP در تخمین SAW آورده شده است. همانند مدل‌های NF، برای رتبه‌بندی عملکرد مدل‌های GEP در برآورد SAW نیز از ملاک پایین بودن RMSE و بالا بودن NS در مقایسه‌ها استفاده گردید. بر این اساس، مشاهده می‌شود مدل GEP شماره ۵ در سری آزمون (جدول ۶) با سه متغیر ورودی کربن آلی، جرم مخصوص ظاهری و میانگین هندسی قطر ذرات (dg) از نظر برآورد SAW به علت داشتن RMSE کم و NS زیاد، دارای دقت تخمین بالاتری در مقایسه با سایر مدل‌های GEP بود.

جدول ۵- معیارهای ارزیابی مدل های نروفازی (NF) آب قابل استفاده خاک (SAW)

Table 5- Evaluation criteria for NF models of SAW

شماره Number	متغیرهای ورودی Input variables	آزمون Testing set				آموزش Training			
		R ²	RMSE (% g g ⁻¹)	ME (% g g ⁻¹)	NS	R ²	RMSE (% g g ⁻¹)	ME (% g g ⁻¹)	NS
1	OC, BD	0.66	2.77	-0.08	0.650.56	2.11	0.00	0.56	
2	OC, BD, clay	0.73	2.51	0.09	0.710.66	1.87	0.00	0.66	
3	OC, sand	0.51	3.29	0.07	0.500.48	2.31	0.00	0.48	
4	BD, sand	0.63	2.87	-0.07	0.620.47	2.32	0.00	0.47	
5	OC, BD, dg	0.62	2.99	0.88	0.590.72	1.70	0.00	0.72	
6	OC, BD, σg	0.64	2.96	-0.07	0.590.65	1.90	0.00	0.65	

NS: ضریب نش-ساتکلِف، RMSE: مجذور میانگین مربعات خطا، R²: ضریب تبیین. توصیف متغیرها در زیر جدول ۱ آورده شده است.

R²: coefficient of determination; RMSE: root mean square error; NS: Nash-Sutcliff coefficient. The description of the variables is available in Table 1.

SAW می باشد؛ همبستگی معنی دار و مثبت متغیر کربن آلی و همبستگی معنی دار و منفی متغیرهای جرم مخصوص ظاهری و dg با SAW (جدول ۲) نیز مؤید این قضیه است. به نظر می رسد با افزایش ماده آلی خاک به علت تأثیر مثبت آن در خاکدانه سازی، رطوبت FC و در نتیجه SAW افزایش می یابد در حالی که افزایش سهم ذرات درشت خاک (شن) از طریق افزایش dg و نیز افزایش جرم مخصوص ظاهری خاک احتمالاً به علت تخریب خاکدانه سازی و تضعیف ساختمان خاک موجب کاهش رطوبت FC و در نتیجه کاهش SAW می گردد (Jury & Horton, 2004).

مقادیر R²، RMSE، ME و NS این مدل برای سری آموزش به ترتیب برابر ۰/۵۵، ۲/۱۵ % g g⁻¹ و ۰/۰۴ و ۰/۵۵ و برای سری آزمون به ترتیب برابر ۰/۷۶، ۳/۱۰ % g g⁻¹ و ۱/۴۱ - و ۰/۵۶ به دست آمد (جدول ۶). مقدار منفی ME برای داده های آزمونی نشان دهنده تخمین کمتر از مقدار واقعی SAW توسط مدل شماره ۵ می باشد ولی با این حال براساس آماره NS، عملکرد این مدل قابل قبول (NS = 0.56) است (Waseem et al., 2017). معادله رگرسیونی این مدل (جدول ۳) بیانگر آن است که کربن آلی با علامت مثبت و جرم مخصوص ظاهری و dg با علامت منفی وارد مدل گردیدند و تأثیر منفی متغیر dg به مراتب بالاتر از جرم مخصوص ظاهری در برآورد

جدول ۶- معیارهای ارزیابی مدل های برنامه ریزی بیان ژن آب قابل استفاده خاک (SAW)

Table 6- Evaluation criteria for GEP models of SAW

شماره Number	متغیرهای ورودی Input variables	آزمون Testing set				آموزش Training			
		R ²	RMSE (% g g ⁻¹)	ME (% g g ⁻¹)	NS	R ²	RMSE (% g g ⁻¹)	ME (% g g ⁻¹)	NS
1	OC, BD	0.74	3.23	-1.58	0.52	0.53	2.18	-0.02	0.53
2	OC, BD, clay	0.69	3.42	-1.61	0.46	0.56	2.12	-0.07	0.56
3	OC, sand	0.51	3.91	-1.81	0.30	0.46	2.35	0.01	0.46
4	BD, sand	0.68	3.32	-1.01	0.49	0.46	2.35	-0.12	0.46
5	OC, BD, dg	0.76	3.10	-1.41	0.56	0.55	2.15	0.04	0.55
6	OC, BD, σg	0.74	3.32	-1.69	0.49	0.55	2.13	-0.04	0.55

NS: ضریب نش-ساتکلِف، RMSE: مجذور میانگین مربعات خطا، ME: میانگین خطا، R²: ضریب تبیین. توصیف متغیرها در زیر جدول ۱ آورده شده است.

R²: coefficient of determination; RMSE: root mean square error; ME: mean error; NS: Nash-Sutcliff coefficient. The description of the variables is available in Table 1.

در رتبه‌های بعدی تخمین SAW، مدل شماره ۱ با دو متغیر ورودی کربن آلی و جرم مخصوص ظاهری ($g\ g^{-1}$) $RMSE = 3.23\%$ و $NS = 0.52$ و مدل‌های شماره ۴ و ۶ (هر دو) با متغیرهای ورودی کربن آلی، جرم مخصوص ظاهری، شن و σ_g ($g\ g^{-1}$) $RMSE = 3.32\%$ و $NS = 0.49$ قرار دارد (جدول ۶). بنابراین براساس مقادیر آماره‌های ارزیابی شش مدل GEP اشاره شده در (جدول ۶)، مدل شماره ۵ نسبت به سایر مدل‌های GEP در برآورد SAW اراضی دشت اردبیل بهترین عملکرد را نشان داد که تفاوت چندانی با عملکرد بهترین مدل MLR ($RMSE = 3.07\%$) و بهترین مدل ANN ($RMSE = 3.05\%$) در برآورد SAW در پژوهش امیرعابدی و همکاران (Amirabedi et al., 2013) ندارد؛ به عبارت دیگر، استفاده از روش GEP در مقایسه با روش‌های رگرسیونی و شبکه عصبی مصنوعی نتوانست باعث بهبود تخمین آب قابل استفاده خاک گردد. جعفری و همکاران (Jafari et al., 2021) گزارش نمودند مدل GEP توانست رطوبت حجمی خاک را از روی متغیرهای زودیافت خاک شامل رس، سیلت، جرم مخصوص ظاهری، تخلخل کل و مکش خاک در اراضی جنگلی فنلاند با ضریب تبیین ۰/۹۳ و مجذور میانگین مربعات خطای ۴/۱۰ درصد حجمی پیش‌بینی نماید. نتایج پژوهش چیم و همکاران (Cheema et al., 2024) نیز نشان داد که مدل GEP کارایی لازم را برای برآورد منحنی رطوبتی خاک در دامنه خشک تا اشباع از روی ویژگی‌های سهل‌الوصول خاک با دقت بالا و خطای پایین دارد.

نتایج مدل‌های جنگل تصادفی (RF) برای برآورد آب قابل استفاده خاک (SAW)

شش مدل جنگل تصادفی (RF) با ترکیب متغیرهای ورودی استفاده شده در توابع رگرسیونی (جدول ۳) برای برآورد آب قابل استفاده خاک (SAW) توسعه داده شد. جدول ۷ مقادیر آماره‌های ضریب تبیین (R^2)، مجذور میانگین مربعات خطا ($RMSE$)، میانگین خطا (ME) و ضریب نش-ساتکلیف (NS) برای مقایسه عملکرد مدل‌های RF از نظر صحت تخمین SAW نشان می‌دهد. همانند دو مدل هوشمند قبلی، رتبه‌بندی مدل‌های RF از نظر دقت برآورد SAW نیز بر اساس پایین بودن آماره $RMSE$ و بالا بودن آماره NS انجام گرفت. همانند روش GEP (جدول ۶)، مدل RF شماره ۵ در سری آزمون (جدول ۷) با سه متغیر ورودی کربن آلی، جرم مخصوص ظاهری و میانگین هندسی قطر ذرات (dg) توانست SAW را به علت داشتن $RMSE$ کم

و NS زیاد در مقایسه با سایر مدل‌های RF با دقت بالایی برآورد نماید. مقادیر R^2 ، $RMSE$ ، ME و NS این مدل برای سری آموزش به‌ترتیب برابر ۰/۹۴، $g\ g^{-1}$ ۰/۹۲، $g\ g^{-1}$ ۰/۰۰ و ۰/۹۲ و برای سری آزمون به‌ترتیب برابر ۰/۶۸، $g\ g^{-1}$ ۳/۳۰، $g\ g^{-1}$ ۱/۴۵ و ۰/۵۰ تعیین گردید (جدول ۷). مقدار منفی ME برای سری آزمون‌یابی‌نگر برآورد کمتر از مقدار واقعی SAW توسط مدل شماره ۵ می‌باشد و مطابق آماره NS ، عملکرد این مدل رضایت‌بخش ($NS = 0.50$) است (Waseem et al., 2017). معادله رگرسیونی مدل ۵ در جدول ۳ نیز بیانگر تأثیر مثبت کربن آلی و تأثیر منفی dg و جرم مخصوص ظاهری در تخمین SAW می‌باشد که با نتایج همبستگی این سه متغیر ورودی با SAW (جدول ۲) در یک راستا می‌باشد. در رتبه‌های بعدی تخمین SAW توسط روش جنگل تصادفی، مدل شماره ۲ با سه متغیر ورودی کربن آلی، جرم مخصوص ظاهری و رس ($g\ g^{-1}$) $RMSE = 3.33\%$ و $NS = 0.49$ و مدل شماره ۴ با دو متغیر ورودی جرم مخصوص ظاهری و شن ($g\ g^{-1}$) $RMSE = 3.42\%$ و $NS = 0.46$ قرار دارد (جدول ۷). به‌طور کلی نتایج ضد و نقیضی در خصوص قابلیت روش جنگل تصادفی (RF) در برآورد منحنی رطوبتی خاک گزارش شده است. در مجموع براساس مقادیر آماره‌های ارزیابی شش مدل RF مورد استفاده در تحقیق حاضر (جدول ۷)، مدل شماره ۵ نسبت به سایر مدل‌های RF در برآورد SAW از خطای کمی برخوردار بود که در مقایسه با بهترین مدل MLR ($RMSE = 3.07\%$) و بهترین مدل ANN ($RMSE = 3.05\%$) در برآورد SAW (Amirabedi et al., 2013) عملکرد ضعیفی از خود نشان داد. این نتایج با یافته‌های کیانی و بیات (Kiani & Bayat, 2024) همخوانی ندارد که نشان دادند مدل جنگل تصادفی در مقایسه با مدل رگرسیونی به‌طور معنی‌دار از صحت تخمین بالاتری در تخمین منحنی رطوبتی خاک از روی ویژگی‌های زودیافت خاک برخوردار بود. راستگو و همکاران (Rastgou, Bayat, Mansoorizadeh, & Gregory, 2020) گزارش نمودند مدل RF به علت بیش‌برازشی، عدم قطعیت در تعیین مقیاس و ناپایداری در مرحله آزمون از قابلیت اطمینان کمتری در مقایسه با مدل رگرسیون غیرخطی چند متغیره در برآورد نقاط منحنی رطوبتی خاک برخوردار بود. نتایج پژوهش شجاعی و همکاران (Shojaei et al., 2023) نیز نشان داد که مدل جنگل تصادفی در مقایسه با مدل شبکه عصبی مصنوعی دارای دقت کمی در پیش‌بینی رطوبت‌های ظرفیت زراعی و نقطه پژمردگی از روی ویژگی‌های زودیافت خاک بود.

جدول ۷- معیارهای ارزیابی مدل جنگل تصادفی آب قابل استفاده خاک (SAW)

Table 7- Evaluation criteria for RF models of SAW

شماره	متغیرهای ورودی	آزمون				آموزش			
		Testing set				Training			
Number	Input variables	R ²	RMSE (% g g ⁻¹)	ME (% g g ⁻¹)	NS	R ²	RMSE (% g g ⁻¹)	ME (% g g ⁻¹)	NS
1	OC, BD	0.60	3.54	-1.68	0.420.91	1.01	0.00	0.90	
2	OC, BD, clay	0.72	3.33	-1.66	0.490.93	0.96	-0.03	0.91	
3	OC, sand	0.40	4.08	-1.86	0.230.92	1.03	0.00	0.89	
4	BD, sand	0.56	3.42	-1.29	0.460.92	1.08	0.01	0.89	
5	OC, BD, dg	0.68	3.30	-1.45	0.500.94	0.92	0.00	0.92	
6	OC, BD, σg	0.72	3.37	-1.78	0.480.93	0.96	0.05	0.92	

NS: ضریب نش-ساتکلیف، RMSE: مجذور میانگین مربعات خطا، ME: میانگین خطا، R²: ضریب تبیین. توصیف متغیرها در زیر جدول ۱ آورده شده است.

R²: coefficient of determination; RMSE: root mean square error; ME: mean error; NS: Nash-Sutcliffe coefficient. The description of the variables is available in Table 1.

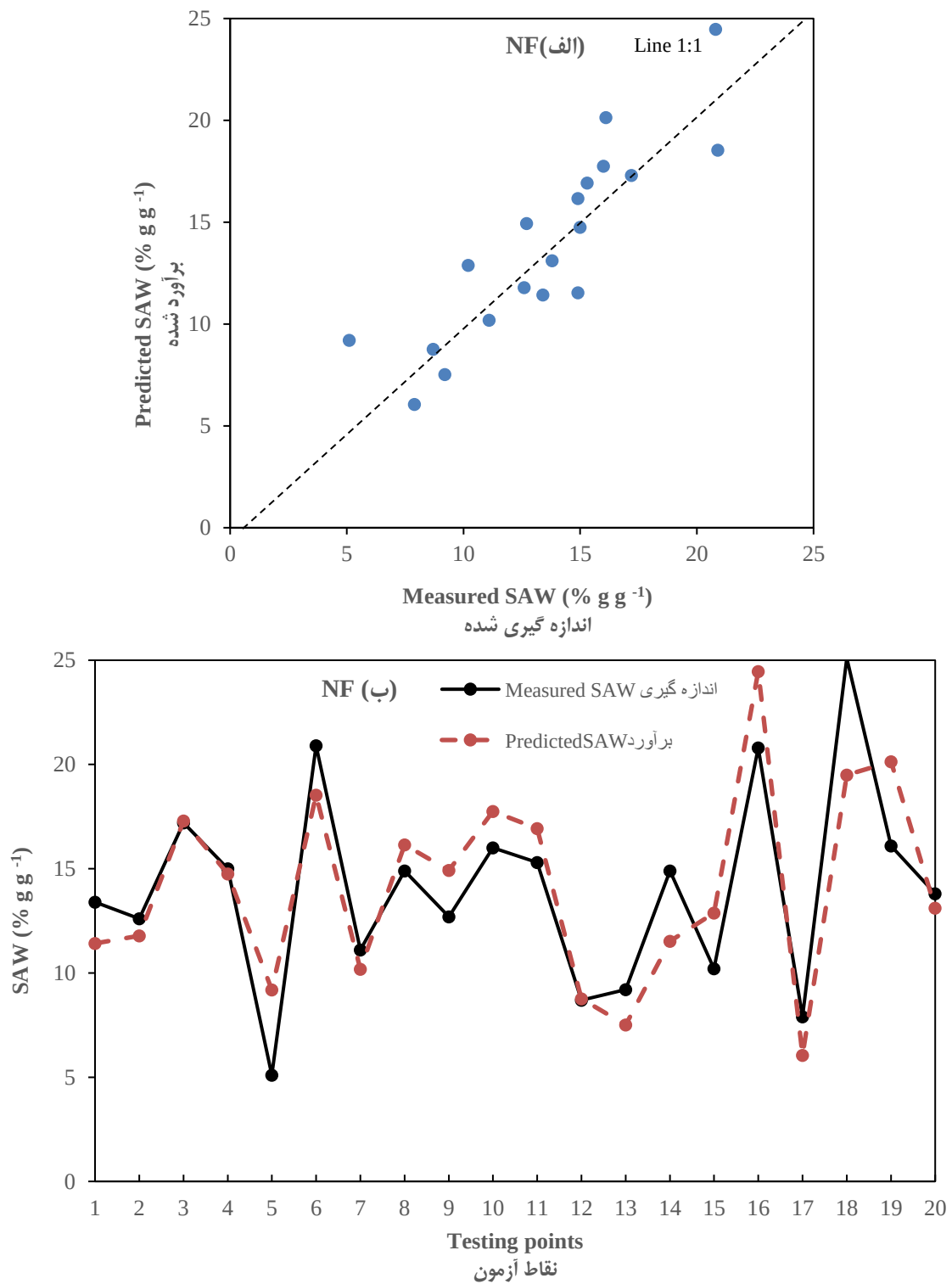
GEP و RF کمتر است. بنابراین می توان چنین استنباط نمود که در این پژوهش، مدل های نروفازی در برآورد SAW از روی متغیرهای زود یافت خاک به علت استفاده از روابط غیر خطی پیچیده به مراتب موفق تر عمل نموده اند در حالی که مدل های جنگل تصادفی از عملکرد ضعیفی در این زمینه برخوردار بوده اند. شیر و همکاران (Shiri et al., 2017) نیز بیان نمودند که مدل نروفازی (R² = ۰/۸۴۷ و ۰/۹۶۰) در مقایسه با مدل های برنامه ریزی بیان ژن (R² = ۰/۸۳۸ و ۰/۸۷۳)، جنگل تصادفی (R² = ۰/۷۸۴ و ۰/۷۱۴) و رگرسیون خطی چندگانه (R² = ۰/۴۱۱ و ۰/۴۱۴) از دقت بالاتری در برآورد به ترتیب رطوبت های FC و PWP از روی متغیرهای زود یافت شامل رس، سیلت، جرم مخصوص ظاهری و d_g در خاک های استان فارس برخوردار بود. همچنین این نتایج با یافته های جعفری و همکاران (Jafari et al., 2021) در یک راستا می باشد که گزارش نمودند مدل های مبتنی بر شبکه های عصبی مصنوعی در مقایسه با مدل برنامه ریزی بیان ژن توانستند با دقت بالاتری آب قابل استفاده خاک را از روی متغیرهای زود یافت خاک تخمین بزنند.

نتیجه گیری

عملکرد سه مدل هوشمند نروفازی (NF)، برنامه ریزی بیان ژن (GEP) و جنگل تصادفی (RF) در برآورد آب قابل استفاده خاک (SAW) با شش ترکیب متفاوت از متغیرهای زود یافت و سهل الوصول خاک شامل شن، رس، میانگین و انحراف معیار هندسی قطر ذرات، کربن آلی و جرم مخصوص ظاهری خاک در اراضی زراعی دشت اردبیل مورد ارزیابی و مقایسه قرار گرفت.

مقایسه عملکرد مدل های NF، GEP و RF در برآورد SAW

مقایسه مقادیر اندازه گیری و برآورد شده آب قابل استفاده خاک (SAW) بر اساس نقاط آزمونی (۲۰ داده تصادفی) برای بهترین مدل در روش های نروفازی (NF)، برنامه ریزی بیان ژن (GEP) و جنگل تصادفی (RF) به ترتیب در شکل ۲، ۳ و ۴ (ب) نشان داده شده است. مقادیر آماره های مجذور میانگین مربعات خطا (RMSE) و ضریب نش-ساتکلیف (NS) برای مدل های شش گانه NF (جدول ۵)، GEP (جدول ۶) و RF (جدول ۷) ارائه شده برای برآورد SAW در داده های آزمونی نشان می دهد که مدل های NF به دلیل داشتن NS زیاد (از ۰/۵۰ تا ۰/۷۱) و RMSE کم (از ۲/۵۱ تا ۳/۲۹ g g⁻¹) در مقایسه با مدل های GEP (NS از ۰/۳۰ تا ۰/۵۶ و RMSE از ۳/۱۰ تا ۳/۹۱ g g⁻¹) و RF (NS از ۰/۲۳ تا ۰/۵۰ و RMSE از ۳/۳۰ تا ۴/۰۸ g g⁻¹) دارای دقت بالا و خطای پایین در برآورد SAW بودند. همچنین مطابق تقسیم بندی واسیم و همکاران (Waseem et al., 2017) برای ضریب نش-ساتکلیف (NS)، می توان استنباط نمود که در تحقیق حاضر مدل های شش گانه NF (جدول ۵) به علت برخورداری از مقادیر بالای ۰/۵۰ برای آماره NS، دارای عملکرد قابل قبول تا بسیار خوب از نظر برآورد آب قابل استفاده خاک هستند در حالی که برخی مدل های GEP (جدول ۶) و RF (جدول ۷) به علت داشتن مقادیر کوچکتر از ۰/۵۰ برای آماره NS، دارای عملکرد نامطلوب برای برآورد آب قابل استفاده خاک در اراضی دشت اردبیل هستند. شکل ۲، ۳ و ۴ (الف) نیز نمودار ۱:۱ مقادیر اندازه گیری و برآورد شده SAW را با استفاده از بهترین مدل NF (مدل شماره ۲ در جدول ۵)، بهترین مدل GEP (مدل شماره ۵ در جدول ۶) و بهترین مدل RF (مدل شماره ۵ در جدول ۷) برای داده های آزمونی نشان می دهد؛ با توجه به این شکل مشاهده می شود که پراکنش داده ها حول محور ۱:۱ در بهترین مدل NF در مقایسه با بهترین مدل های

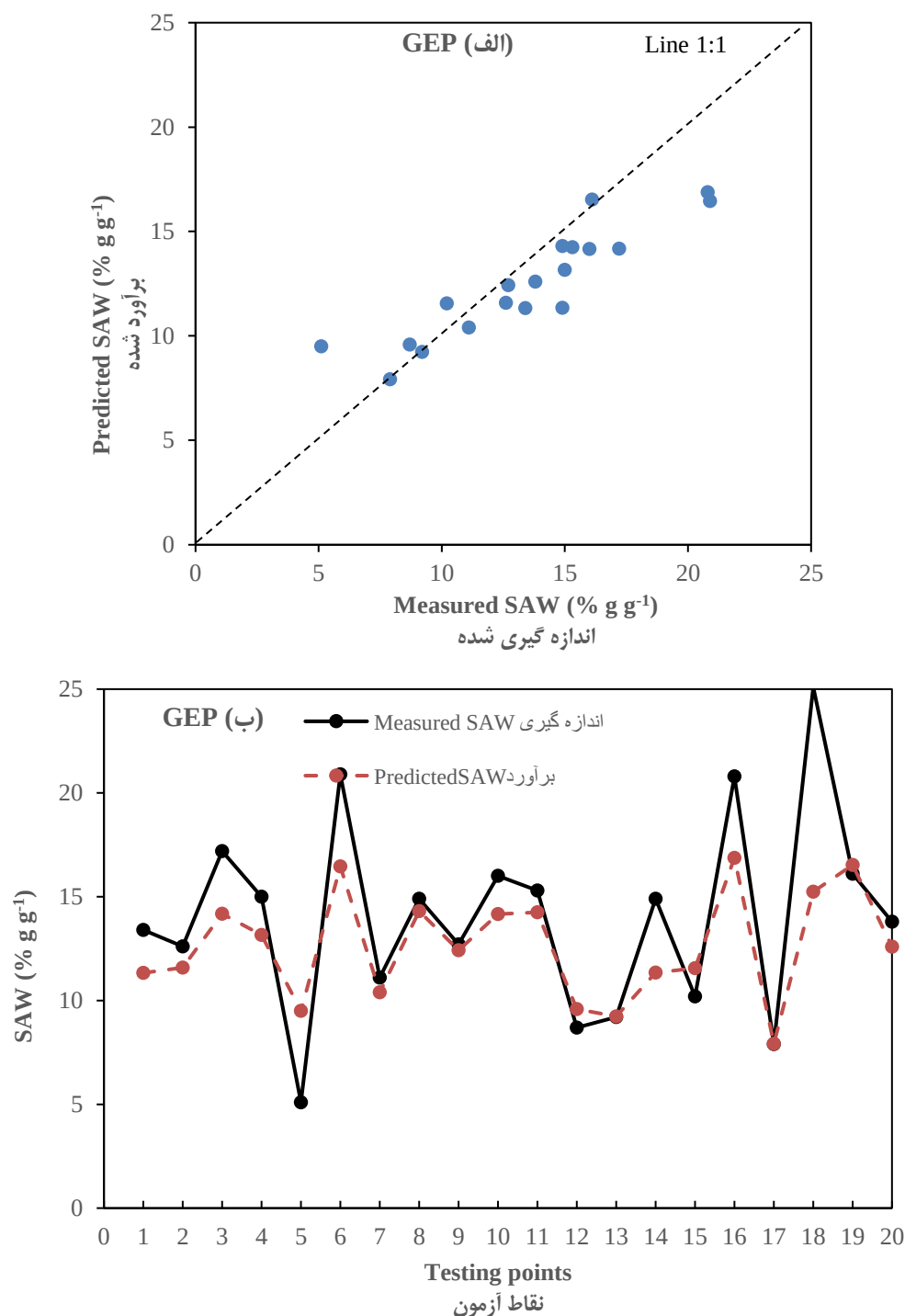


شکل ۲- مقایسه مقادیر اندازه‌گیری و برآورد شده آب قابل استفاده خاک (SAW) با استفاده از بهترین مدل نروفازی (NF) بر اساس داده‌های آزمونی (n=20)

اطلاعات مربوط به مدل و متغیرهای ورودی در جدول ۵ آورده شده است.

Figure 2- Comparison of predicted and measured values of soil available water (SAW) using the best NF model based on testing data (n=20)

Table 5 provides detailed information about model and input variables.

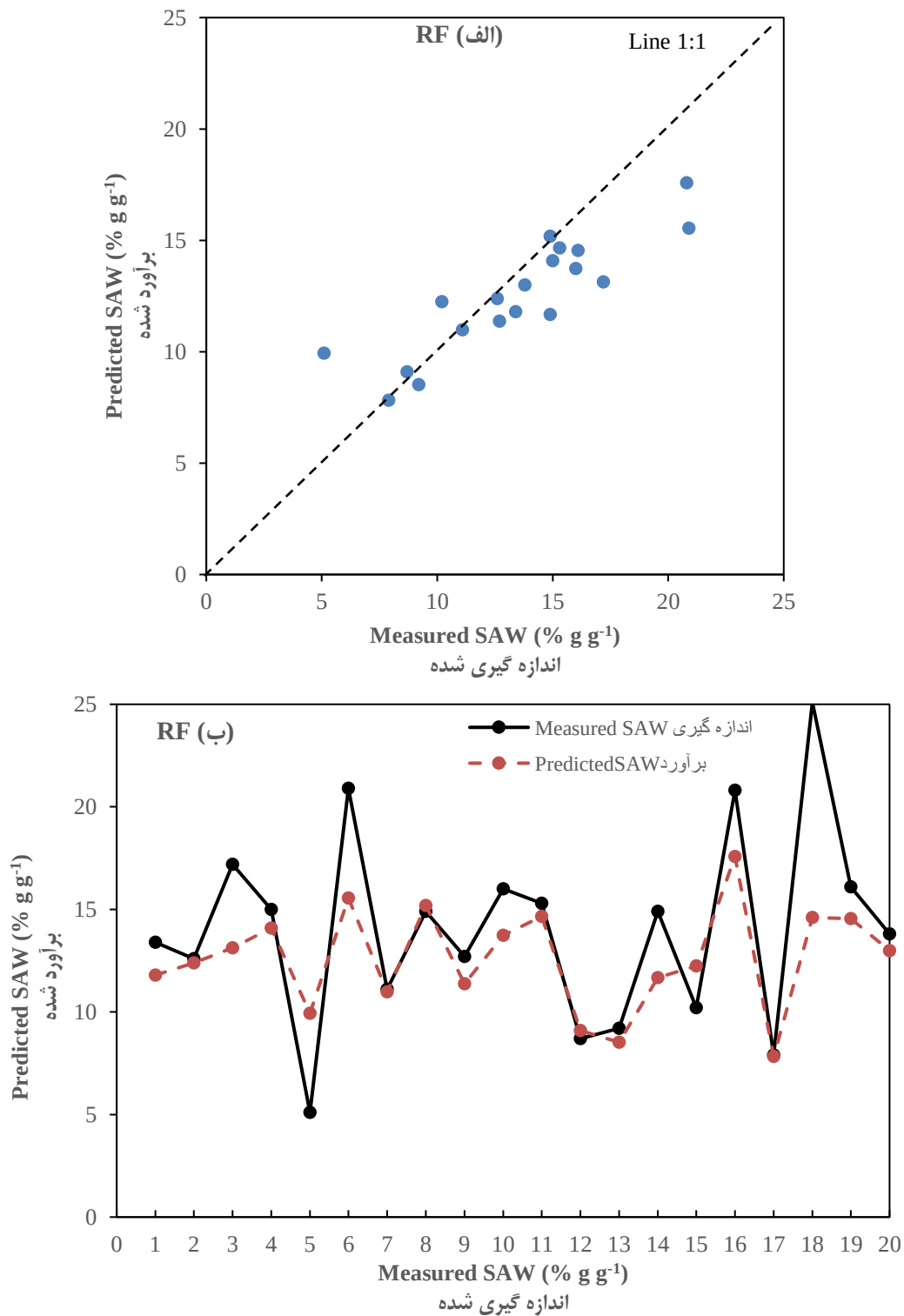


شکل ۳- مقایسه مقادیر اندازه گیری و برآورد شده آب قابل استفاده خاک (SAW) با استفاده از بهترین مدل برنامه ریزی بیان ژن (GEP) بر اساس داده های آزمونی (n=20)

اطلاعات مربوط به مدل و متغیرهای ورودی در جدول ۶ آورده شده است.

Figure 3- Comparison of predicted and measured values of soil available water (SAW) using the best GEP model based on testing data (n=20)

Table 6 provides detailed information about model and input variables.



شکل ۴- مقایسه مقادیر اندازه‌گیری و برآورد شده آب قابل استفاده خاک (SAW) با استفاده از بهترین مدل جنگل تصادفی (RF) بر اساس داده‌های آزمونی (n=20)

اطلاعات مربوط به مدل و متغیرهای ورودی در جدول ۷ آورده شده است.

Figure 4- Comparison of predicted and measured values of soil available water (SAW) using the best RF model based on testing data (n=20)

Table 7 provides detailed information about model and input variables.

کربن آلی، جرم مخصوص ظاهری و رس به دلیل داشتن مقادیر بالای آماره NS و مقادیر پایین آماره RMSE، عملکرد مناسب تری در مقایسه با بهترین مدل های GEP و RF در برآورد SAW داشت؛ به نظر می رسد می توان از این مدل برای برآورد آب قابل استفاده خاک در سایر مناطق نیمه خشک کشور و جهان با شرایط اقلیمی مشابه دشت اردبیل نیز با دقت مناسب بهره برد که البته این موضوع نیاز به راستی آزمایی دارد. پیشنهاد می شود در مطالعات بعدی از رگرسیون غیر خطی چندگانه و سایر مدل های هوشمند مانند ماشین بردار پشتیبان نیز در برآورد SAW در منطقه مورد مطالعه استفاده گردد.

رابطه همبستگی مثبت و معنی دار بین SAW با کربن آلی و رس و نیز رابطه همبستگی منفی و معنی دار بین SAW با شن و جرم مخصوص ظاهری خاک یافت شد. نتایج مدل سازی نشان داد که کربن آلی، جرم مخصوص ظاهری و رس از مهم ترین و اثرگذارترین متغیرهای ورودی در برآورد SAW در منطقه مورد مطالعه بودند. همچنین ارزیابی و مقایسه نتایج مدل ها بیانگر آن بود که هر شش مدل نروفازی (NF) مورد استفاده در پژوهش به علت برخورداری از مقادیر نش-سانکلیف (NS) بالاتر از ۰/۵۰، دارای عملکرد قابل قبول در برآورد SAW هستند. در مجموع، بهترین مدل نروفازی با متغیرهای ورودی

References

- 1- Amirabedi, H., Asghari, Sh., Mesri, T., & Keivan behjo, F. (2013). Estimating of field capacity, permanent wilting point and available water content in Ardabil plain using regression and artificial neural network models. *Applied Soil Research*, 1(1), 60-72. (In Persian with English abstract)
- 2- Bauer, A., & Black, A.L. (1992). Organic carbon effects on available water capacity of three soil textural groups. *Soil Science Society of America Journal*, 56, 248-254. <https://doi.org/10.2136/sssaj1992.03615995005600010038x>
- 3- Blake, G.R., & Hartge, K.H. (1986). *Bulk density*. p. 363-375. In: Klute A. (ed). *Methods of Soil Analysis Part 1, Physical and Mineralogical Methods*. 2nd ed. American Society of Agronomy, Madison, WI. <https://doi.org/10.2136/sssabookser5.1.2ed.c13>
- 4- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.
- 5- Cheema, S.J., Shah, S.H.H., Farooque, A.A., Jamei, M., Abbas, F., Esau, T.J., & Grewal, K.S. (2024). A comprehensive analytical and computational assessment of soil water characteristics curves in Atlantic Canada: Application of a novel SelectKbestbased GEP model. *Agricultural Water Management*, 298, 108868. <https://doi.org/10.1016/j.agwat.2024.108868>
- 6- Ferreira, C. (2006). *Gene Expression Programming: Mathematical modeling by an artificial intelligence*. Springer, Berlin, Heidelberg, New York, p. 478.
- 7- Gardner, W.H. (1986). *Water content*. p. 493-544. In: Klute A. (ed). *Methods of Soil Analysis. Part 1*. 2nd ed. Agronomy. Monograph. 9. ASA, Madison, WI.
- 8- Gee, G.W., & Or, D. (2002). *Particle-size analysis*. p. 255-293. In: Dane J. H., & Topp G. C. (eds.). *Methods of Soil Analysis. Part 4*. SSSA Book Series No. 5. Soil Science Society of America, Madison, WI. <https://doi.org/10.2136/sssabookser5.4.c12>
- 9- Ghorbani, M.A., Deo, R.C., Kashani, M.H., Shahabi M., & Ghorbani, S. (2019). Artificial intelligence-based fast and efficient hybrid approach for spatial modeling of soil electrical conductivity. *Soil and Tillage Research*, 186, 152-164. <https://doi.org/10.1016/j.still.2018.09.012>
- 10- Jafari, M.M., Ojaghlo, H., & Karbasi, M. (2021). Comparison of the performance of artificial neural networks and gene expression programming in estimating the forest soil water characteristic curve. *Iranian Journal of Soil and Water Research*, 52(8), 2093-2109.
- 11- Jury, W., & Horton, R. (2004). *Soil Physics*. John Wiley and Sons, Inc.
- 12- Kiani, R., & Bayat, H. (2024). Capability of regression and random forest methods to estimate soil water retention curve by developing pseudo-continuous pedotransfer functions. *Journal of Soil and Plant Science (Water and Soil Science)*, 34(4), 15-36. <https://doi.org/10.3390/w12123425>
- 13- Kozak, E., Pachepsky, Y.A., Sokolowski, S., Sokolowska, Z., & Stepniewski, W. (1996). A modified number-based method for estimating fragmentation fractal dimensions of soils. *Soil Science Society of America Journal*, 60, 1291-1297. <https://doi.org/10.2136/sssaj1996.03615995006000050002x>
- 14- Merdun, H., Cinar, O., Meral, R., & Apan, M. (2006). Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. *Soil and Tillage Research*, 90, 108-116. <https://doi.org/10.1016/j.still.2005.08.011>
- 15- Nelson, D.W., & Sommers, L.E. (1982). *Total carbon, organic carbon, and organic matter*. p. 539-579. In: A.L. Page et al. (ed.) *Methods of Soil Analysis. Part 2*. 2nd ed. Agron. Monogr. 9. ASA and SSSA, Madison, WI. <https://doi.org/10.2134/agronmonogr9.2.2ed.c29>
- 16- Ostovari, Y., Asgari, K., & Cornelis, W. (2015). Performance evaluation of pedotransfer functions to predict field

- capacity and permanent wilting point using UNSODA and HYPERS datasets. *Arid Land Research Management*, 29, 383-398. <https://doi.org/10.1080/15324982.2015.1029649>
- 17- Rastgou, M., Bayat, H., Mansoorizadeh, M., & Gregory, A.S. (2020). Estimating the soil water retention curve: comparison of multiple nonlinear regression approach and random forest data mining technique. *Computer and Electronics in Agriculture*, 174, 1-13. <https://doi.org/10.1016/j.compag.2020.105502>
- 18- Shirazi, M.A., & Boersma, L. (1984). A unifying quantitative analysis of soil texture. *Soil Science Society American Journal*, 48(1), 142-147. <https://doi.org/10.2136/sssaj1984.03615995004800010026x>
- 19- Shiri, J., Keshavarzi, A., Kisi, O., & Karimi, S. (2017). Using soil easily measured parameters for estimating soil water capacity: Soft computing approaches. *Computers and Electronics in Agriculture*, 141, 327-339. <https://doi.org/10.1016/j.compag.2017.08.012>
- 20- Shojaei, S., Farhadi Bansouleh, B., Fatehi, Sh., & Rahmani, M. (2023). Comparison of pedotransfer functions based on machine learning methods to estimate soil moisture at field capacity and permanent wilting point (Case study: Ravansar District, Kermanshah Province). *Journal of Water and Soil Conservation*, 30(1), 27-47. <https://doi.org/10.3390/w13192639>
- 21- Waseem, M., Mani, N., Andiego, G., & Usman, M. (2017). A review of criteria of fit for hydrological models. *International Research Journal of Engineering and Technology*, 4(11), 1765-1772.
- 22- Zhang, R., & Zhang, S. (2024). Coefficient of permeability prediction of soils using gene expression programming. *Engineering Applications of Artificial Intelligence*, 128 (107504). <https://doi.org/10.1016/j.engappai.2023.107504>