

## 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)

Sahar Shojaei<sup>1</sup>, Bahman Farhadi Bansouleh<sup>\*2</sup>, Shahrokh Fatehi<sup>3</sup>,  
Mahsa Rahmani<sup>4</sup>

1. M.Sc. Graduate, Dept. of Water Engineering, Faculty of Agriculture, Razi University, Kermanshah, Iran. E-mail: [sahar.shojaei1993@gmail.com](mailto:sahar.shojaei1993@gmail.com)
2. Corresponding Author, Associate Prof., Dept. of Water Engineering, Faculty of Agriculture, Razi University, Kermanshah, Iran. E-mail: [bfarhadi@razi.ac.ir](mailto:bfarhadi@razi.ac.ir)
3. Research Assistant Prof., Soil and Water Research Department, Kermanshah Agricultural and Natural Resources Research and Education Center, AREEO, Kermanshah, Iran. E-mail: [shahrokh.fatehi@gmail.com](mailto:shahrokh.fatehi@gmail.com)
4. Assistant Researcher, Dept. of Water Engineering, Faculty of Agriculture, Razi University, Kermanshah, Iran. E-mail: [mahsarahmani39@yahoo.com](mailto:mahsarahmani39@yahoo.com)

### Article Info

**Article type:**  
Research Full Paper

**Article history:**  
Received: 12.18.2022  
Revised: 04.07.2023  
Accepted: 04.11.2023

**Keywords:**  
Cubist model,  
Random forest model,  
Soil physical properties,  
Support vector machine  
model

### ABSTRACT

**Background and Objectives:** The physical properties of the soil, which cannot be easily measured, play an important role in the design of irrigation and drainage systems. Since the direct measurement of these characteristics is time-consuming and expensive, therefore, to estimate these parameters, most researchers use indirect methods such as transfer functions. This research aims to investigate and determine the best model for estimating soil moisture content in Field capacity (FC) and Permanent Wilting Point (PWP) using easily measured soil characteristics and pedotransfer functions in the R software environment and choosing the most suitable model for the soils of the Ravansar region in Kermanshah province.

**Materials and Methods:** In this research, the easily measurable properties of soil were used as input variables for five transfer functions of the multivariable linear, artificial neural network, Cubist, random forest, and support vector machine. At first, in the study area, the location of 120 profiles was determined using the Latin hypercube method. In these observation points, soil profile was dug and studied and samples were taken from its horizons. Then, laboratory analysis including measurement of electrical conductivity, pH, calcium carbonate equivalent, organic carbon, and percentage of sand, silt, and clay was performed on soil samples. Based on the range of changes of these characteristics, especially the soil texture, 75 surface soil samples and 33 soil samples from ten different soil profiles were selected. PWP measurement was performed on 33 samples and FC measurement was performed on surface and depth samples, i.e., 108, and then modeling operations were performed on them. Root mean square error (RMSE), mean absolute error (MAE), and  $R^2$  indices were used to evaluate the models.

**Results:** The results showed that the accuracy of pedotransfer functions in estimating PWP is higher than FC ( $R^2$  and RMSE values of the Cubist model for PWP are 0.81 and 0.054 and for FC are 0.53 and 0.085, respectively). Also, the results for FC showed that between the models, the Cubist and the artificial neural network have low MAE (0.066 and 0.068)

---

and RMSE (0.085) and high  $R^2$  (0.53 and 0.54) respectively, compared to other models.

**Conclusion:** The overall results showed that Cubist, Artificial Neural Network, and Random Forest models with lower error and higher  $R^2$  have higher efficiency for soil moisture estimation in FC than other models. The results showed that the Cubist and random forest models were the best models for estimating moisture at the PWP in terms of comparing the coefficient of determination. This research showed the importance of using new machine learning methods in studies related to soil transfer functions to estimate difficult-to-measure soil properties. Also, the results of this research are acceptable for a wide range of plains in Kermanshah province, which has similar soil formation conditions to the Ravansar region.

---

Cite this article: Shojaei, Sahar, Farhadi Bansouleh, Bahman, Fatehi, Shahrokh, Rahmani, Mahsa. 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.



© The Author(s).

DOI: 10.22069/jwsc.2023.20857.3606

Publisher: Gorgan University of Agricultural Sciences and Natural Resources

---

## مقایسه توابع انتقالی مبتنی بر روش‌های یادگیری ماشین برای تخمین رطوبت ظرفیت زراعی و پژمردگی (مطالعه موردی: منطقه روانسر کرمانشاه)

سحر شجاعی<sup>۱</sup>، بهمن فرهادی بانسوله<sup>۲\*</sup>، شاهرخ فاتحی<sup>۳</sup>، مهسا رحمانی<sup>۴</sup>

۱. دانش‌آموخته کارشناسی ارشد گروه مهندسی آب، دانشکده کشاورزی، دانشگاه رازی، کرمانشاه، ایران. رایانامه: [sahar.shojaei1993@gmail.com](mailto:sahar.shojaei1993@gmail.com)
۲. نویسنده مسئول، دانشیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه رازی، کرمانشاه، ایران. رایانامه: [bfarhadi@razi.ac.ir](mailto:bfarhadi@razi.ac.ir)
۳. استادیار پژوهش، بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کرمانشاه، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرمانشاه، ایران. رایانامه: [shahrokh.fatehi@gmail.com](mailto:shahrokh.fatehi@gmail.com)
۴. دستیار پژوهشی، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه رازی، کرمانشاه، ایران. رایانامه: [mahsarahmani39@yahoo.com](mailto:mahsarahmani39@yahoo.com)

اطلاعات مقاله	چکیده
<b>نوع مقاله:</b> مقاله کامل علمی- پژوهشی	<b>سابقه و هدف:</b> خصوصیات فیزیکی دیریافت خاک نقش مهمی در طراحی سامانه‌های آبیاری و زهکشی دارند. از آن‌جاکه اندازه‌گیری مستقیم این خصوصیات زمان‌بر و پرهزینه است، بیش‌تر پژوهش‌گران برای تخمین این پارامترها از روش‌های غیرمستقیم مانند توابع انتقالی استفاده می‌نمایند. هدف از این پژوهش بررسی و تعیین بهترین مدل برای برآورد رطوبت ظرفیت زراعی (FC) و نقطه پژمردگی دائم (PWP) با استفاده از خصوصیات زودیافت خاک و توابع انتقالی در محیط نرم‌افزار R و هم‌چنین انتخاب مناسب‌ترین تابع برای خاک‌های منطقه روانسر در استان کرمانشاه است.
<b>تاریخ دریافت:</b> ۰۱/۰۹/۲۷ <b>تاریخ ویرایش:</b> ۰۲/۰۱/۱۸ <b>تاریخ پذیرش:</b> ۰۲/۰۱/۲۲	
<b>واژه‌های کلیدی:</b> خصوصیات فیزیکی خاک، مدل جنگل تصادفی، مدل کیوبیست، مدل ماشین بردار پشتیبان	<b>مواد و روش‌ها:</b> در این پژوهش از خصوصیات زودیافت خاک به‌عنوان متغیرهای ورودی برای پنج تابع انتقالی خطی چندمتغیره <sup>۱</sup> ، شبکه عصبی مصنوعی <sup>۲</sup> ، کیوبیست <sup>۳</sup> ، جنگل تصادفی <sup>۴</sup> و ماشین بردار پشتیبان <sup>۵</sup> استفاده شد. در ابتدا در منطقه مورد مطالعه با روش ابر مکعب لاتین موقعیت مکانی ۱۲۰ خاکرخ تعیین شد. در این نقاط مشاهداتی خاکرخ‌ها حفر و از افق‌های آن نمونه‌برداری صورت گرفت. پس از تجزیه‌های آزمایشگاهی بر روی نمونه‌های خاک شامل اندازه‌گیری هدایت الکتریکی عصاره اشباع، واکنش خاک، کربنات کلسیم معادل، کربن آلی، درصد شن، سیلت و رس خاک، بر اساس دامنه تغییرات این ویژگی‌ها به‌ویژه اجزا بافتی خاک

- 1- Multivariable Linear
- 2- Artificial Neural Network
- 3- Cubist
- 4- Random Forest
- 5- Support Vector Machine

به ترتیب ۷۵ نمونه خاک سطحی و ۳۳ نمونه خاک از ده خاکریز مختلف انتخاب گردیدند. اندازه‌گیری PWP بر روی ۳۳ نمونه و اندازه‌گیری FC بر روی مجموع نمونه‌های سطحی و عمقی یعنی ۱۰۸ نمونه انجام شد و در مرحله بعد عملیات مدل‌سازی بر روی آن‌ها اجرا شد. برای ارزیابی مدل‌ها از شاخص‌های ریشه میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و ضریب تبیین ( $R^2$ ) استفاده شد.

**یافته‌ها:** نتایج نشان داد که دقت توابع انتقالی در برآورد رطوبت PWP بیش‌تر از FC است (مقدار  $R^2$  و RMSE مدل کیویست برای PWP به ترتیب برابر ۰/۸۱۳ و ۰/۰۵۴ و برای FC برابر ۰/۵۳ و ۰/۰۸۵ بود). هم‌چنین نتایج برآورد رطوبت FC نشان داد که مدل کیویست و شبکه عصبی مصنوعی به ترتیب دارای MAE (۰/۰۶۶ و ۰/۰۶۸) و RMSE (۰/۰۸۵) کم‌تر و  $R^2$  (۰/۵۳ و ۰/۵۴) بیش‌تری نسبت به سایر مدل‌ها هستند.

**نتیجه‌گیری:** نتایج کلی نشان داد که مدل‌های کیویست و شبکه عصبی مصنوعی و پس‌از آن مدل جنگل تصادفی با خطای کم‌تر و ضریب تبیین بالاتر نسبت به سایر مدل‌ها از کارایی مناسبی برای برآورد رطوبت ظرفیت زراعی برخوردارند. نتایج مربوط به رطوبت PWP نشان داد که مدل کیویست و پس‌از آن مدل جنگل تصادفی از نظر مقایسه ضریب تبیین بهترین مدل‌ها برای برآورد رطوبت PWP هستند. این پژوهش اهمیت استفاده از روش‌های نوین یادگیری ماشین در مطالعات مربوط به توابع انتقالی خاک برای برآورد خصوصیات دیریافت خاک را نشان داد. هم‌چنین نتایج این پژوهش برای دامنه وسیعی از دشت‌های استان کرمانشاه که شرایط تشکیل خاک مشابه با منطقه روانسر را دارند، قابل قبول است.

**استناد:** شجاعی، سحر، فرهادی بانسوله، بهمن، فاتحی، شاهرخ، رحمانی، مهسا (۱۴۰۲). مقایسه توابع انتقالی مبتنی بر روش‌های یادگیری ماشین برای تخمین رطوبت ظرفیت زراعی و پژمردگی (مطالعه موردی: منطقه روانسر کرمانشاه). پژوهش‌های حفاظت آب و خاک، ۳۰ (۱)، ۲۷-۴۷.

DOI: 10.22069/jwsc.2023.20857.3606



© نویسندگان.

ناشر: دانشگاه علوم کشاورزی و منابع طبیعی گرگان

## مقدمه

خاک را با دقت مناسبی برآورد می‌کنند؛ ولی در برآورد رطوبت قابل‌استفاده در خاک، مدل‌های رگرسیونی کارایی لازم را ندارند. هم‌چنین دریافتند که دقت تخمین ضرایب رطوبتی توسط مدل شبکه عصبی مصنوعی بیش‌تر از مدل‌های رگرسیونی بوده است (۸). نوروزی و همکاران (۱۳۹۸) عملکرد روش‌های داده‌مبنا در تخمین نقاط مهم رطوبتی در شاهرود را بررسی کردند. نتایج مطالعات آن‌ها نشان داد که روش رگرسیون بردار پشتیبان کارایی بهتری نسبت به روش شبکه عصبی مصنوعی و رگرسیون خطی چندمتغیره دارد. هم‌چنین آن‌ها دریافتند که روش رگرسیون بردار پشتیبان با خطای پایین و ضریب تبیین بالا نقاط مهم رطوبتی خاک را پیش‌بینی نموده و می‌تواند جایگزین خوبی برای سایر روش‌های سنتی باشد (۹). لایب و همکاران (۲۰۱۲)، از روش جنگل تصادفی و درخت رگرسیون برای پیش‌بینی توزیع بافت خاک استفاده کرده و نشان دادند که روش جنگل تصادفی در مقایسه با روش درخت رگرسیون از دقت بیش‌تری برخوردار است (۱۰). مک‌برانتی و همکاران (۲۰۱۱) با مطالعه روی ۸۰۶ خاکرخ در پنج عمق برای پیش‌بینی آب موجود در خاک با استفاده از توابع انتقالی به این نتیجه رسیدند که مدل کیوبیست بهترین مدل با مقدار خطای کم‌تر است (۱۱). راستگو و همکاران (۲۰۲۰) روش رگرسیون غیرخطی چندمتغیره و روش داده‌کاوی جنگل تصادفی را در برآورد منحنی رطوبتی خاک مقایسه نمودند. نتایج نشان داد که دقت توابع انتقالی مبتنی بر جنگل تصادفی به‌طور قابل‌توجهی بهتر از توابع انتقالی مبتنی بر روش رگرسیون غیرخطی چندمتغیره بود و قابلیت اطمینان توابع انتقالی مبتنی بر روش رگرسیون غیرخطی چندمتغیره به‌طور قابل‌توجهی

رطوبت خاک ظرفیت زراعی (FC) و نقطه پژمردگی دائم (PWP) از مهم‌ترین داده‌های موردنیاز در مطالعات آبیاری هستند که اندازه‌گیری مستقیم آن‌ها دشوار، وقت‌گیر و پرهزینه است؛ بنابراین متخصصین آب‌و‌خاک روش‌های به‌صرفه‌ای از نظر زمان و هزینه مانند توابع انتقالی را برای برآورد این دسته از خصوصیات فیزیکی خاک ارائه داده‌اند. ازجمله مدل‌های توابع انتقالی می‌توان به روش شبکه عصبی مصنوعی (۱)، روش غیرپارامتریک (۲)، مدیریت گروهی داده‌ها (۳)، مدیریت چندهدفه گروهی داده‌ها (۴) و روش ماشین بردار پشتیبان (۵) اشاره نمود. توابع انتقالی هم‌چون پلی برای برقراری ارتباط بین ویژگی‌های هیدرولیکی خاک (متغیرهای دیریاخت) و ویژگی‌های مبنایی خاک (متغیرهای زودیاخت هم‌چون توزیع اندازه ذرات خاک، جرم مخصوص ظاهری و کربن آلی) عمل می‌کنند (۶). مهربانی گوهری و همکاران (۱۳۹۱) با برآورد رطوبت خاک در نقاط FC و PWP با استفاده از چند روش یادگیری ماشین به این نتیجه رسیدند که شبکه عصبی بهترین عملکرد و بعدازآن مدل رگرسیون چندمتغیره عملکرد بهتری نسبت به بقیه توابع انتقالی دارند. در بین توابع انتقالی نیز توابع انتقالی کلاسی عملکرد بهتری نسبت به توابع انتقالی نقطه‌ای و پارامتری داشته‌اند. به‌طورکلی، آن‌ها دریافتند که فرایند آموزش در تعیین پارامترهای مدل با استفاده از داده‌های یک منطقه اهمیت فراوانی دارد (۷). امیر عابدی و همکاران (۱۳۹۲) در پژوهشی نقاط مهم رطوبتی خاک‌های دشت اردبیل را با استفاده از مدل‌های رگرسیونی و شبکه عصبی مصنوعی برآورد نمودند. نتایج آن‌ها نشان داد که هر دو روش، ضرایب رطوبتی

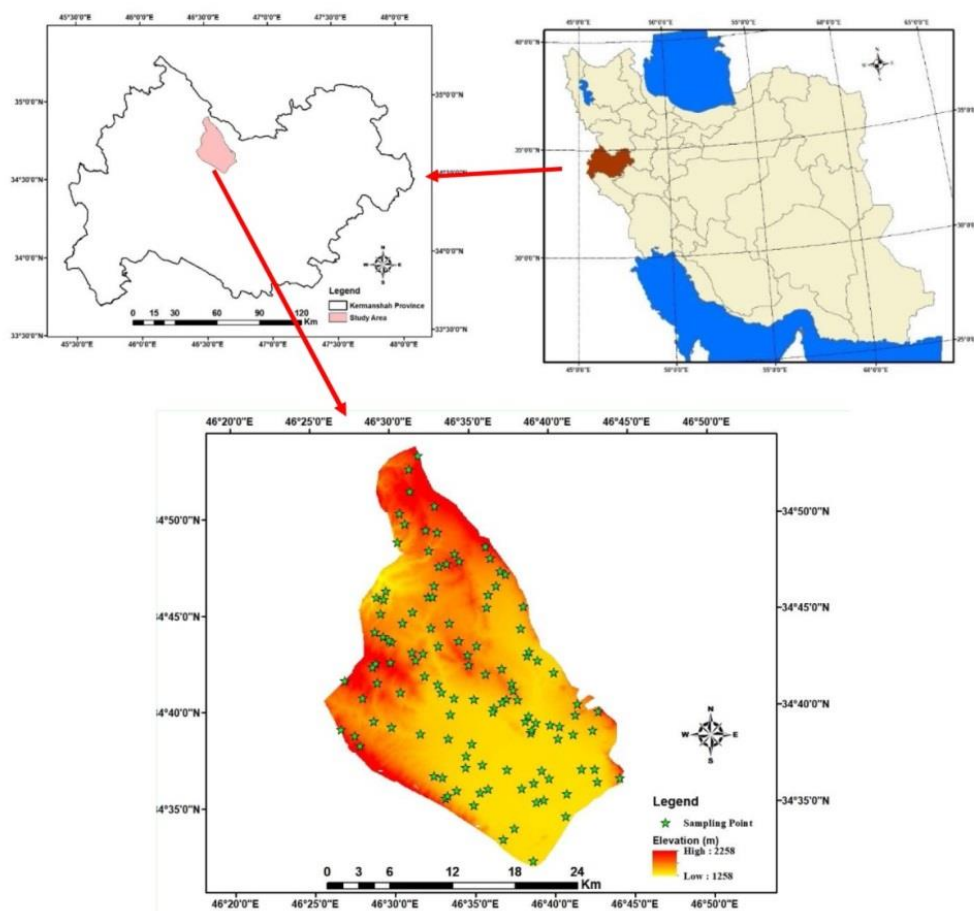
هم‌چنین مشخص شد که مدل MLP در مرحله پیش‌بینی مدل تا حدودی بهتر از SVM عمل می‌کند (۱۴).

باتوجه به بررسی منابع صورت‌گرفته، می‌توان بیان کرد که توابع انتقالی خاک روابطی تجربی هستند و به‌منظور بررسی قابلیت استفاده از آن‌ها در سایر خاک‌ها، لازم است کارآیی آن‌ها در مقابل داده‌ها و اطلاعات خاک‌های مختلف ارزیابی شود. لذا هدف از این پژوهش بررسی و تعیین بهترین تابع انتقالی برای برآورد رطوبت در نقاط FC و PWP با استفاده از خصوصیات زودیافت خاک در محیط نرم‌افزار R برای خاک‌های منطقه روانسر در استان کرمانشاه است.

#### مواد و روش‌ها

این پژوهش در شهرستان روانسر در ناحیه‌ای به وسعت حدود ۵۰,۰۰۰ هکتار در ۶۰ کیلومتری شمال‌غربی استان کرمانشاه، در بین مختصات طول جغرافیایی  $۲۶/۸^{\circ}$  تا  $۴۶^{\circ}$   $۴۲/۹'$  شرقی و عرض جغرافیایی  $۲/۳۳^{\circ}$  تا  $۳۴^{\circ}$   $۵۱/۶'$  شمالی انجام شد (شکل ۱). محدوده مورد مطالعه از شمال به منطقه اورامانات، از شمال غرب به شهرستان پاوه، از غرب به شهرستان جوانرود، از جنوب به شهرستان دالاهو و از شرق به شهرستان کرمانشاه محصور می‌شود.

بهتر از توابع انتقالی مبتنی بر جنگل تصادفی بود. به‌طور کلی، روش جنگل تصادفی به دلیل بیش‌برازش، عدم قطعیت در تعیین مقیاس و ناپایداری در مرحله آزمایش، قابلیت اطمینان کم‌تری برای پیش‌بینی نقاط منحنی رطوبتی خاک در مقایسه با روش رگرسیون غیرخطی چندمتغیره داشت (۱۲). کواف و همکاران (۲۰۲۱) در پژوهشی به ارزیابی عملکرد توابع انتقالی موجود برای پیش‌بینی نگهداری آب در خاک در اراضی تحت کشاورزی حفاظتی در فرانسه پرداختند و سپس توابع انتقالی جدیدی را با استفاده از سه روش رگرسیون خطی چندگانه، درخت رگرسیون و جنگل تصادفی برای پیش‌بینی محتوای آب خاک در نقاط FC و PWP توسعه دادند. نتایج نشان داد که به‌طور کلی، خطای مطلق پیش‌بینی PWP کم‌تر از FC بوده است و روش جنگل تصادفی مقدار آب خاک را با دقت بهتری برآورد کرده است (۱۳). عبدالحفید و همکاران (۲۰۲۱) در صحرای الجزایر بر روی ۱۲۰ نمونه خاک روش‌های رگرسیونی را به‌منظور پیش‌بینی میزان رطوبت خاک در FC و PWP در سه رویکرد رگرسیون خطی چندگانه (MLR)، شبکه عصبی مصنوعی چندلایه (MLP) و ماشین بردار پشتیبان (SVM) مقایسه کردند. نتایج نشان داد که هر دو الگوریتم هوش مصنوعی MLP و SVM قادر به ارائه عملکرد بهتری نسبت به MLR معمولی هستند.



شکل ۱- موقعیت نقاط نمونه‌برداری در ایران و استان کرمانشاه.

Figure 1. Location of sampling points in Iran and Kermanshah Province.

۳۳ نمونه از افق‌های مختلف ۱۰ خاکرخ مطالعه شده در بخش‌های مختلف منطقه انتخاب شدند. اندازه‌گیری PWP به دلیل هزینه زیاد و زمان‌بر بودن تنها بر روی ۳۳ نمونه خاک مربوط به خاکرخ‌های انتخابی انجام شد؛ اما مقدار FC خاک علاوه بر این نمونه‌ها، بر روی ۷۵ نمونه خاک سطحی یعنی در مجموع بر روی ۱۰۸ نمونه خاک اندازه‌گیری شد. به‌منظور تعیین رطوبت خاک در نقاط FC و PWP، نمونه‌های خاک در دستگاه صفحات فشاری به ترتیب تحت مکش‌های ۳۰۰ و ۱۵۰۰۰ پاسکال قرار گرفتند. در این پژوهش دامنه‌ای از مدل‌های مرسوم مورد استفاده به‌عنوان توابع انتقالی خاک شامل مدل‌های رگرسیون خطی چندمتغیره، شبکه عصبی

ابتدا با روش ابرمکعب لاتین موقعیت مکانی ۱۲۰ نقطه مشاهداتی در منطقه مورد مطالعه تعیین شد. پس از حفر خاکرخ‌ها و مطالعه آن‌ها، نمونه‌های مختلف سطحی و عمقی به آزمایشگاه منتقل شد. سپس در آزمایشگاه خصوصیات زودپافت خاک از جمله هدایت الکتریکی عصاره گل اشباع (EC)، واکنش خاک (pH)، کربنات کلسیم معادل (CCE)، کربن آلی (OC)، درصد شن (Sand)، سیلت (Silt) و رس (Clay) که در این مطالعه به‌عنوان متغیرهای ورودی توابع انتقالی مورد استفاده قرار گرفتند با روش‌های استاندارد اندازه‌گیری شدند. بر اساس دامنه تغییرات ویژگی‌های زودپافت خاک به‌ویژه درصد شن، سیلت و رس خاک، ۷۵ نمونه از افق‌های سطحی خاکرخ‌ها و

پرسپترون چندلایه دسته‌ای از شبکه‌های عصبی مصنوعی پیش‌خور است که حداقل سه لایه (شامل یک لایه ورودی، یک لایه پنهان و یک لایه خروجی) دارد (شکل ۲). به جز نرون‌های لایه ورودی، هر نرون از یک تابع فعال‌سازی غیرخطی استفاده می‌کند. خروجی هر نرون طبق رابطه ۲ مشخص می‌شود:

$$y_j = \sum w_{ij}x_i + b_j \quad (2)$$

سپس  $y_j$  در تابع فعال‌سازی  $f$  پردازش شده و خروجی نهایی طبق رابطه ۳ به دست می‌آید. از توابع فعال‌سازی می‌توان به تابع خطی و سیگموئیدی اشاره کرد.

$$\hat{y} = f(y_j) = \frac{1}{1+e^{-y_j}} \quad (3)$$

در این روابط،  $w$  وزن‌های شبکه،  $x$  متغیرهای ورودی و  $b$  بایاس است. باید توجه داشت که پارامترهای  $b$  و  $w$  قابل تنظیم هستند و تابع محرک  $f$  نیز توسط طراح انتخاب می‌شود. در این مطالعه از بسته نرم‌افزاری Neuralnet برای اجرای این مدل در محیط نرم‌افزار R استفاده شد (۱۵).

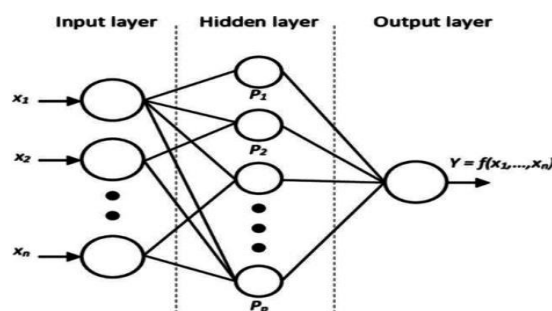
مصنوعی و روش‌های یادگیری ماشین که کارایی آن‌ها توسط پژوهش‌گران مختلف تأیید شده از جمله کیوبیست و جنگل تصادفی برای توسعه این توابع انتقالی خاک استفاده شد. در ذیل به تشریح مختصر مدل‌ها پرداخته می‌شود.

**مدل خطی چندمتغیره:** در این روش رابطه بین یک متغیر وابسته و تعدادی متغیر مستقل در قالب یک معادله خطی چندمتغیره تعیین می‌شود. در این مدل‌های آماری فرض بر این است که رابطه بین متغیرهای مستقل و متغیرهای وابسته به صورت زیر است (رابطه ۱):

$$Y = a + b_1(x_1) + b_2(x_2) + \dots + b_p(x_p) \quad (1)$$

که در آن،  $Y$  متغیر وابسته مانند FC و PWP بوده و  $x_1, x_2, \dots, x_p$  متغیرهای مستقل مانند درصد رس، سیلت و شن هستند. پارامترهای  $b_1, b_2, \dots, b_p$  ضرایب رگرسیون جزئی و مقدار  $a$  عرض از مبدأ است که با رگرسیون‌گیری تعیین می‌شود.

**مدل شبکه عصبی مصنوعی:** در این پژوهش از شبکه عصبی مصنوعی پرسپترون چندلایه<sup>۱</sup> به عنوان تابع انتقالی برای پیش‌بینی FC و PWP استفاده شد.



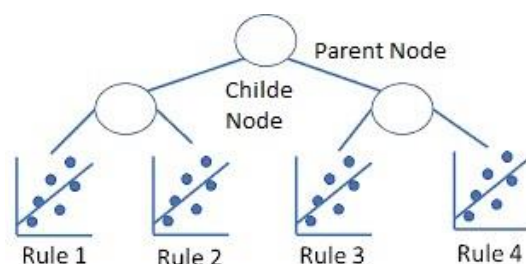
شکل ۲- ساختار ساده‌ای از شبکه عصبی مصنوعی.

Figure 2. A simple structure of an artificial neural network.



استفاده شده‌اند. در هر مرحله از تشکیل درخت مدل‌های خطی حد واسطی هم تشکیل می‌شود. پیش‌بینی نهایی با استفاده از مدل رگرسیون خطی در گره انتهایی درخت انجام می‌شود، اما برای پیرایش نتیجه نهایی از مدل‌های خطی گره‌های پیشین درخت تشکیل شده استفاده می‌شود (۱۶).

مدل کیوبیست: مدل کیوبیست یک مدل مبتنی بر قاعده بوده و توسعه یافته مدل درختی M5 کویلاین است (۱۶). یک درخت تا جایی رشد می‌کند که گره‌های انتهایی تشکیل یک سری مدل‌های رگرسیونی بدهند (شکل ۳). این مدل‌های رگرسیونی مبتنی بر پیش‌بینی‌کننده‌هایی (متغیرهای کمکی) است که در تفکیک‌های درخت (توسعه شاخه و گره‌های درخت)



شکل ۳- ساختار ساده‌ای از مدل کیوبیست.

Figure 3. A simple structure of Cubist model.

یادگیرنده‌های ضعیف به صورت تکرار شونده آموزش داده می‌شوند و به مجموعه قبلی اضافه می‌شوند تا در نهایت یک درخت طبقه‌بندی یا رگرسیون قوی حاصل شود. یادگیرنده‌های ضعیف در حین اضافه شدن به مجموعه، وزن‌دهی می‌شوند که این وزن‌دهی معمولاً بر اساس میزان دقت در نتایج درختان طبقه‌بندی یا رگرسیون هست (۱۷). اما در طرح کمیته در مدل کیوبیست درخت‌های تکراری به ترتیب ایجاد می‌شوند. درخت اول که بر اساس قواعد ذکر شده در بالا تولید شده و درخت‌های بعدی با استفاده از نسخه‌های تنظیم شده بر اساس نتایج مجموعه آموزشی ایجاد می‌شوند. اگر مدل یک مقدار را بیش از حد پیش‌بینی کرد، پاسخ در مدل بعدی به سمت مقدار کم‌تر تنظیم می‌شود. برخلاف روش بوستینگ، وزن هر مرحله برای هر کمیته برای میانگین پیش‌بینی‌های هر درخت مدل استفاده نمی‌شود. پیش‌بینی نهایی میانگین ساده پیش‌بینی‌های هر درخت در مدل است (۱۸).

مدل کیوبیست از نزدیک‌ترین همسایه‌ها<sup>۱</sup> برای اصلاح پیش‌بینی مبتنی بر قاعده استفاده می‌کند. در مدل کیوبیست، مدل‌های تشکیل شده توسط دو مدل خطی به صورت زیر بیان می‌شوند (رابطه ۴):

$$\hat{Y}_{par} = a * \hat{Y}_{kid} + (1 - a)\hat{Y}_{par} \quad (4)$$

در این رابطه،  $\hat{Y}_{par}$  پیش‌بینی از گره والد و  $\hat{Y}_{kid}$  پیش‌بینی از گره فرزند است. ضریب  $a$  هم از رابطه‌ای شامل واریانس پیش‌بینی‌های والد و فرزند و اختلاف آن‌ها به دست می‌آید. مدل کیوبیست همچنین می‌تواند از طرحی شبیه بوستینگ<sup>۲</sup> که اصطلاحاً کمیته<sup>۳</sup> نامیده می‌شود، استفاده نماید. بوستینگ، روشی برای تبدیل سیستم‌های یادگیری ضعیف به قوی بر اساس ترکیب نتایج درختان طبقه‌بندی یا رگرسیون مختلف است.

- 1- Nearest Neighbors
- 2- Boosting
- 3- Committee

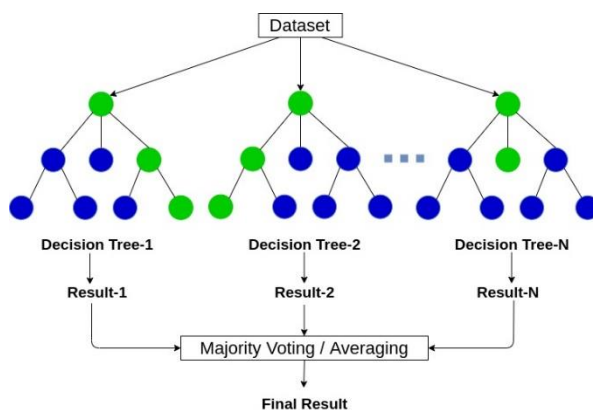
بهترین تفکیک‌کننده از میان این متغیرها و  $\epsilon$ ) پیش‌بینی نهایی از ترکیب پیش‌بینی‌های حاصل از  $n$  تعداد درخت. در مدل جنگل‌های تصادفی، تعداد  $k$  درخت رگرسیون ساخته می‌شود و مقادیر پیش‌بینی شده نهایی با تجمیع نتایج همه درختان منفرد به دست می‌آید و پس از رشد  $k$  درخت  $\{Tk(x)\}$ ، تخمین‌گر جنگل‌های تصادفی رگرسیونی با رابطه ۵ توصیف می‌شود (شکل ۴):

$$f(x) = \frac{\sum_{k=1}^K T_k(x)}{K} \quad (5)$$

در این رابطه،  $T$  درخت،  $X$  متغیرهای کمکی،  $k$  تعداد درختان می‌باشد

در این پژوهش برای اجرای مدل جنگل تصادفی از بسته نرم‌افزاری `randoamForest` در محیط نرم‌افزار `R` استفاده شد (۱۹، ۲۰).

مدل جنگل تصادفی: برای اولین بار بریمن (۲۰۰۱) الگوریتم جنگل تصادفی را پیشنهاد نمود. در مدل درختان رگرسیون و طبقه‌بندی استاندارد، هر گره با استفاده از بهترین تفکیک‌کننده از میان همه متغیرهای کمکی ایجاد می‌شود. اما در جنگل تصادفی هر گره با استفاده از بهترین زیرمجموعه از بین زیرمجموعه‌های متغیرهای کمکی موجود در گره، منشعب می‌شود (۱۹). مدل جنگل تصادفی دارای دو پارامتر اصلی شامل تعداد متغیرهای کمکی موجود در زیرمجموعه تصادفی هر گره و دیگری تعداد درختان در جنگل است. الگوریتم جنگل تصادفی طی چندمرحله اجرا می‌شود: ۱) استخراج نمونه‌ها به روش بازنمونه‌گیری از داده‌های اولیه (مشاهدات نقطه‌ای خاک) به تعدادی مشخص ( $ntree$ )، ۲) ایجاد رگرسیون برای هر نمونه بازنمونه‌گیری شده، ۳) انتخاب تعدادی از متغیرهای کمکی ( $mtry$ ) در هر گره به طور تصادفی و انتخاب



شکل ۴- ساختار ساده مدل جنگل تصادفی.

Figure 4. A simple structure of random forest model.

روابط غیرخطی در یک فضای مشخص القایی با ابعاد بالا استفاده می‌کند (۲۱، ۲۲). فرایند تخمین‌گر `SVM` در رگرسیون را می‌توان به صورت رابطه ۶ بیان کرد:

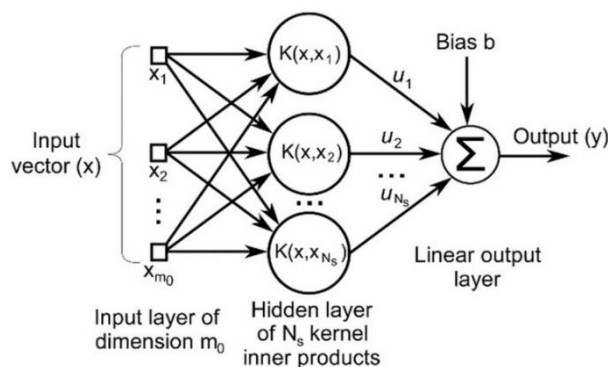
$$f(x) = w \cdot \theta(x) + b \quad (6)$$

مدل ماشین بردار پشتیبان (`SVM`): ماشین‌های بردار پشتیبان از الگوریتم‌های نظارت شده یادگیری ماشین مبتنی بر کرنل است. این مدل از سطوح تصمیم‌گیری (تعریف شده توسط یک تابع کرنل) برای ترسیم

#### 1- Bootstrap

در نظر گرفتن روابط غیرخطی فراهم می‌شود. تابع کرنل این امکان را می‌دهد. توابع کرنل رایج شامل تابع خطی، چندجمله‌ای، رادیال و سیگموئیدی هستند که در این پژوهش از تابع رادیال استفاده شد (۲۲). شکل ۵ ساختار ساده‌ای از ماشین بردار پشتیبان را نشان می‌دهد. در این پژوهش برای پیاده‌سازی SVM از بسته e1071 در محیط نرم‌افزار R استفاده شد (۲۱).

که در آن،  $w$  بردار وزن،  $b$  عرض از مبدا و  $\theta$  تابع غیرخطی است؛ بنابراین، در مدل‌های SVM همانند مدل‌های شبکه عصبی، بردار وزن باید تعیین شود. معمولاً متغیرهای پیش‌بینی‌کننده در فضای ویژگی ارتباط غیرخطی با متغیر هدف دارند. این محدودیت در مدل‌های SVM با ارتقای فضای برداری به فضای چندبعدی ویژگی با استفاده از تابع کرنل، امکان



شکل ۵- ساختار ساده‌ای از ماشین بردار پشتیبان.

Figure 5. Simple structure of support vector machine.

$$R^2 = \frac{\sum_{i=1}^n (M_i - \bar{M})^2 - \sum_{i=1}^n (E_i - \bar{E})^2}{\sum_{i=1}^n (M_i - \bar{M})^2} \quad (9)$$

که در این روابط،  $M_i$  و  $E_i$  به ترتیب متغیر اندازه‌گیری شده و پیش‌بینی شده و  $\bar{M}$  و  $\bar{E}$  به ترتیب میانگین مشاهدات و پیش‌بینی متغیر هدف و  $n$  تعداد نمونه‌ها می‌باشد. هرچه مقدار RMSE و MAE به صفر نزدیک باشد، اختلاف بین مقادیر تخمین زده شده با اندازه‌گیری شده کم‌تر بوده و مدل مناسب‌تر است (۲۴).

برای اجرای مدل‌های فوق‌الذکر باتوجه به نوع اعتبارسنجی در نظر گرفته شده (حذف یک داده و پیش‌بینی آن) در این پژوهش به دلیل داده‌های اندک، از بسته نرم‌افزاری CARET استفاده شد که مدل‌های دیگر را از تعیین داده‌های آموزشی و آزمون، روش

برای ارزیابی صحت و دقت مدل‌ها از روش حذف یک داده و پیش‌بینی آن<sup>۱</sup> استفاده شد (۲۳). در شرایطی که زمان و بودجه کافی برای برداشت نمونه وجود نداشته باشد، بهترین گزینه اعتبارسنجی روش مذکور است. برای ارزیابی نتایج مدل‌ها از شاخص‌های آماری ریشه میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و ضریب تبیین ( $R^2$ ) استفاده شد (روابط ۷ تا ۹):

$$RMSE = \sqrt{\frac{\sum (M_i - E_i)^2}{N}} \quad (7)$$

$$MAE = \frac{\sum_{i=1}^n (M_i - E_i)}{N} \quad (8)$$

1 - Leave-one-out cross-validation

### نتایج و بحث

داده‌های بافت، شوری، کربن آلی، درصد آهک و pH خاک برای پیش‌بینی FC مورد استفاده قرار گرفت. ابتدا نتایج تجزیه‌های آزمایشگاهی نمونه‌ها در محیط نرم‌افزار اکسل وارد شده و سپس داده‌های حاصل به محیط نرم‌افزار R فراخوانی گردید. خلاصه آماری داده‌های موردنظر برای پیش‌بینی FC در جدول ۱ ارائه شده است. از آنجا که مقدار p-value حاصل از آزمون نرمال بودن شاپیرو-ویلک برای پارامترهای هدایت الکتریکی عصاره گل اشباع، واکنش خاک، درصد کربنات کلسیم، کربن آلی، شن خاک، رس خاک و رطوبت ظرفیت زراعی کم‌تر از ۰/۰۵ بود، این متغیرها از توزیع نرمال پیروی نمی‌کنند. با توجه به این که مقدار p-value حاصل از آزمون نرمال بودن شاپیرو-ویلک برای درصد سیلت خاک بیش‌تر از ۰/۰۵ بود می‌توان بیان کرد که داده‌های درصد سیلت از توزیع نرمال پیروی می‌کند. گومز و گارسیا (۲۰۰۲) ضریب تغییرات را در سه دسته تغییرپذیری کم (کم‌تر از ۱۰٪)، زیاد (۲۰٪-۳۰٪) و خیلی زیاد (بیش از ۳۰٪) طبقه‌بندی نمودند (۲۵). براین اساس شن خاک دارای تغییرپذیری زیاد و مقدار pH خاک دارای تغییرپذیری کم است.

اعتبارسنجی و معیارهای آن، تنظیم پارامترهای هر مدل و انتخاب متغیرهای مؤثر در اجرای مدل را تحت پوشش قرار می‌دهد (۲۰). تخمین سهم هر متغیر در مدل‌های مورد استفاده با نرم‌افزار CARET به صورت زیر انجام می‌شود.

در مدل خطی، قدرمطلق آماره t برای هر پارامتر مدل استفاده می‌شود. در مدل جنگل تصادفی برای هر درخت، دقت پیش‌بینی در بخش خارج از سبد داده‌ها ثبت می‌شود. به دنبال آن همین کار پس از جابه‌جایی هر متغیر انجام می‌شود. سپس تفاوت بین این دو دقت در تمام درختان به طور میانگین محاسبه و با خطای استاندارد نرمال می‌شود. در به‌کارگیری مدل جنگل تصادفی برای داده‌های پیوسته خاک مثل FC و PWP، مقدار RMSE بر روی داده‌های خارج از سبد برای هر درخت محاسبه می‌شود، و پس از تغییر یک متغیر، همان محاسبه می‌شود. تفاوت‌ها با خطای استاندارد نرمال می‌شوند. در مدل کیوبیست اهمیت نسبی متغیرها بر اساس درصد دفعاتی که هر متغیر در قواعد شرطی مدل و مدل خطی استفاده شده است بیان می‌شود (۲۰).

جدول ۱- آماره‌های توصیفی خصوصیات زودیافت خاک بر اساس ۱۰۸ نمونه خاک مورد استفاده برای تعیین FC

**Table 1. Descriptive statistics of soil easily available properties based on 108 soil samples used to determine FC.**

آزمون شاپیرو-ویلک	کشیدگی	چولگی	انحراف استاندارد	ضریب تغییرات	حداکثر	میانگین	حداقل	واحد	خصوصیات خاک
Shapiro-Wilk test p-value	Kurtosis	Skewness	Standard deviation	Coefficient of variation	Max	Mean	Min	Unit	Soil properties
1E-13	6.60	2.49	0.42	59.22	2.56	0.71	0.15	dS.m <sup>-1</sup>	EC
6E-06	0.46	-0.96	0.37	4.80	8.26	7.73	6.61	(H <sup>+</sup> )	pH
2E-07	2.06	1.22	11.98	81.27	63.50	14.74	1.00	%	CCE
1E-05	2.53	1.20	0.68	51.07	4.39	1.33	0.12	%	OC
3E-11	2.75	1.75	15.97	90.07	81.00	17.73	1.20	%	Sand
2E-01	-0.24	-0.34	11.15	27.49	65.40	40.56	11.00	%	Silt
3E-05	1.40	-1.02	9.75	23.36	62.40	47.71	8.00	%	Clay

ضرایب همبستگی خصوصیات زودبافت خاک با رطوبت FC تعیین شد که نتایج آن در جدول ۲ گزارش شده است. درصد شن با میزان همبستگی ۰/۷۲۹ بالاترین و EC با میزان ۰/۰۶۱ کم‌ترین همبستگی را با ظرفیت زراعی دارند. همبستگی درصد شن، سیلت و رس با FC در سطح احتمال ۰/۰۱ معنی‌دار و pH با FC در سطح ۰/۰۵ معنی‌دار است. لازم به ذکر است که درصد رس و سیلت دارای همبستگی مثبت و شن دارای همبستگی منفی با FC است.

جدول ۲- ضرایب همبستگی ویژگی‌های مورد مطالعه با FC.

Table 2. Correlation coefficients of studied properties with FC.

	EC	pH	CCE	OC	Clay	Silt	Sand	FC
EC	1							
pH	-0.035	1						
CCE	-0.067	0.366**	1					
OC	0.294**	-0.044	-0.395**	1				
Clay	0.091	0.085	-0.115	-0.063	1			
Silt	0.011	0.366**	0.087	-0.199*	0.165	1		
Sand	-0.063	-0.287**	0.009	-0.1	-0.725**	-0.799**	1	
FC	0.061	0.240*	-0.139	0.137	0.546**	0.567**	-0.729**	1

\* و \*\* به ترتیب معنی‌دار در سطح احتمال پنج (P≤۰/۰۵) و یک درصد (P≤۰/۰۱)

\* and \*\* Significant at P ≤ 0.05 and P ≤ 0.01 respectively

شامل هدایت الکتریکی عصاره گل اشباع، واکنش خاک، کربنات کلسیم معادل، کربن آلی، درصد شن، سیلت و رس، و ظرفیت زراعی رابطه خطی چندمتغیره برقرار شد (رابطه ۱۰):

نتایج این جدول با نتایج پژوهش بوتولا و همکاران (۲۰۱۲) و مین انگوین و همکاران (۲۰۱۵) از نظر بیش‌ترین میزان همبستگی شن با FC همخوانی دارد (۲۶، ۲۷). برای مدل رگرسیون خطی چندمتغیره بین FC و خصوصیات زودبافت خاک

(۱۰)

$$FC = 0.45 + 0.0009 * EC + 0.0127 * pH - 0.0212 * CCE + 0.0005 * OC + 0.0552 * Clay + 0.0594 * silt \quad R^2 = 0.504$$

درصد شن، کربنات کلسیم معادل، واکنش خاک، هدایت الکتریکی عصاره گل اشباع، درصد رس و سیلت بودند (شکل ۶).

در شکل ۶ اهمیت نسبی متغیرهای مستقل در مدل خطی چندمتغیره برای برآورد نقطه ظرفیت زراعی نشان داده شده است. ضریب تبیین (R<sup>2</sup>) نشان می‌دهد که این مدل می‌تواند ۵۰٪ تغییرات FC را توضیح دهد و متغیرهای مهم در این رابطه به ترتیب اهمیت شامل

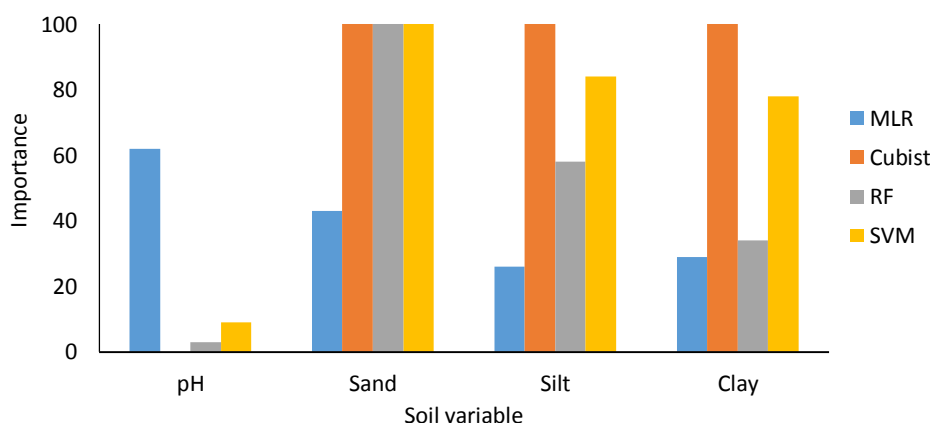
در این پژوهش مدل شبکه عصبی مصنوعی با مشخصاتی شامل لایه اول با ۲ نرون، لایه دوم با ۴

بهترین مدل جنگل تصادفی (RF) برای پیش‌بینی FC، مدلی با ۲ متغیر کمکی موجود در زیرمجموعه تصادفی هر گره (mtree=۲) و تعداد درختان ۱۰۰۰ بود. طبق شکل ۶ نتایج نشان داد که درصد شن مؤثرترین و واکنش خاک کم‌اثرترین پیش‌بینی‌کننده در برآورد رطوبت ظرفیت زراعی با مدل جنگل تصادفی بوده‌اند.

نتایج مدل ماشین بردار پشتیبان (SVM) نشان داد که درصد شن مؤثرترین و کربن آلی خاک کم‌اثرترین پیش‌بینی‌کننده در برآورد رطوبت ظرفیت زراعی با ماشین بردار پشتیبان بودند (شکل ۶).

نرون و لایه سوم با ۲ نرون به‌عنوان بهترین مدل شناخته شد. در مدل شبکه عصبی مصنوعی چون مدل دارای ماهیت جعبه سیاه است، بنابراین متغیرهای کمی مهم قابل استخراج نبوده است.

در مدل کیوبیست طبق شکل ۶ متغیرهای کمی به‌ترتیب اهمیت شامل درصد شن، کربنات کلسیم معادل، سیلت، رس، کربن آلی، واکنش خاک و هدایت الکتریکی عصاره گل اشباع هستند. بهترین مدل با تعداد کمیتی ۳ و تعداد نزدیک‌ترین همسایه معادل صفر به‌عنوان بهترین مدل مشخص شد. با توجه به نتایج فوق در مدل کیوبیست درصد شن و کربنات کلسیم معادل، بیش‌ترین ارتباط را با رطوبت خاک در ظرفیت زراعی داشتند (شکل ۶).



شکل ۶- اهمیت نسبی متغیرهای مستقل در مدل‌های خطی چندمتغیره، کیوبیست، جنگل تصادفی و ماشین بردار پشتیبان برای برآورد FC.

Figure 6. Relative importance of independent variables in multivariable linear, Cubist, random forest and support vector machine models for estimating FC.

و شن خاک به دلیل دارا بودن p-value کم‌تر از ۰/۰۵، از توزیع نرمال پیروی نمی‌کنند. درحالی‌که مقدار p-value حاصل از آزمون نرمال بودن شاپیرو-ویلک مقادیر کربن آلی، درصد سیلت و رس خاک و مقدار رطوبت در نقطه پژمردگی دایم بیش‌تر از ۰/۰۵ بوده و بنابراین، این ویژگی‌ها از توزیع نرمال پیروی می‌کنند. طبق طبقه‌بندی پیشنهادی گومز و گارسیا (۲۰۰۲)

به‌منظور برآورد رطوبت PWP خصوصیات زودیافت مورداستفاده همان خصوصیات مورداستفاده برای برآورد FC بود. آماره‌های توصیفی این خصوصیات در جدول ۳ ارائه شده است. طبق نتایج جدول ۳ و مقادیر p-value حاصل از آزمون نرمال بودن شاپیرو-ویلک، پارامترهای هدایت الکتریکی عصاره گل اشباع، واکنش خاک، درصد کربنات کلسیم

برای ضریب تغییرات، شن خاک و کربنات کلسیم  
معادل دارای تغییرپذیری زیاد و مقدار جرم مخصوص  
ظاهری و حقیقی خاک دارای تغییرپذیری کم هستند  
(۲۵).

جدول ۳- آماره‌های توصیفی خصوصیات زودیافت خاک براساس ۳۳ نمونه خاک مورد استفاده برای تعیین PWP.

Table 3. Descriptive statistics of easily available soil properties based on 33 soil samples used to determine PWP.

آزمون شاپیرو-ویلک	کشیدگی	چولگی	انحراف استاندارد	ضریب تغییرات	حداکثر	میانگین	حداقل	واحد	خصوصیات خاک
Shapiro-Wilk test p-value	Kurtosis	Skewness	Standard deviation	Coefficient of variation	Max	Mean	Min	Unit	Soil properties
1.3E-06	8.31	2.63	0.36	58.61	2.17	0.61	0.28	dS.m <sup>-1</sup>	EC
9.9E-03	-0.40	-0.80	0.44	5.70	8.26	7.70	6.61	(H <sup>+</sup> )	PH
3.4E-03	0.83	1.08	15.47	86.11	63.50	17.96	1.00		CCE
2.0E-01	-0.13	0.40	0.52	56.84	2.34	0.91	0.12	%	OC
7.0E-04	-1.11	0.61	24.01	89.09	81.00	26.95	1.20	%	Sand
3.3E-01	-1.20	-0.03	12.80	37.55	56.60	34.08	11.00	%	Silt
1.9E-01	-0.86	-0.45	14.05	36.04	62.40	38.97	8.00	%	Clay

رطوبت PWP دارند. همبستگی درصد شن، سیلت و رس با میزان PWP در سطح احتمال ۰/۰۱ معنی دار است. لازم به ذکر است که درصد رس و سیلت دارای همبستگی مثبت و درصد شن دارای همبستگی منفی با رطوبت در نقطه PWP است.

ضرایب همبستگی خصوصیات زودیافت خاک برای تخمین PWP تعیین شد که نتایج آن در جدول ۴ گزارش شده است. براساس این نتایج درصد شن با میزان همبستگی ۰/۸۸۵ و هدایت الکتریکی با میزان ۰/۳۷ به ترتیب بیش‌ترین و کم‌ترین همبستگی را با

جدول ۴- ضرایب همبستگی ویژگی‌های مورد مطالعه با رطوبت PWP.

Table 4. Correlation coefficients of studied properties with PWP.

	EC	pH	CCE	OC	Clay	Silt	Sand	PWP
EC	1							
pH	-0.011	1						
CCE	-0.086	0.402*	1					
OC	0.123	-0.047	-0.521**	1				
Clay	0.216	0.263	0.057	0.153	1			
Silt	-0.032	0.286	-0.004	0.126	0.599**	1		
Sand	-0.011	-0.307	-0.011	0.157	-0.904**	-0.883**	1	
PWP	-0.037	0.233	-0.049	0.059	0.756**	0.803**	-0.885**	1

\* و \*\* به ترتیب معنی دار در سطح احتمال پنج (P ≤ ۰/۰۵) و یک درصد (P ≤ ۰/۰۱)

\* and \*\* Significant at P ≤ 0.05 and P ≤ 0.01 respectively

بین PWP و خصوصیات زودبافت خاک شامل هدایت الکتریکی عصاره گل اشباع، واکنش خاک، کربنات کلسیم معادل، کربن آلی خاک، درصد شن، سیلت و رس و رطوبت PWP رابطه خطی چندمتغیره برقرار شد (رابطه ۱۱):

نتایج این جدول با نتایج پژوهش بوتولا و همکاران (۲۰۱۲) و مین انگوین و همکاران (۲۰۱۵) از نظر بیش‌ترین میزان همبستگی شن با PWP همخوانی دارد (۲۶، ۲۷).

(۱۱)

$$PWP = 0.2672 - 0.0139 * EC + 0.0017 * pH - 0.0205 * CCE - 0.0194 * OC + 0.0575 * Clay + 0.0676 * Silt \quad R^2 = 0.6995$$

برای خاک‌های مورد مطالعه تعیین گردید. طبق شکل ۷ در مدل کیوبیست درصد شن بیش‌ترین ارتباط را با نقطه پژمردگی دارا است.

هم‌چنین، بهترین مدل جنگل تصادفی (RF) برای پیش‌بینی PWP، مدلی با ۲ متغیر کمکی موجود در زیرمجموعه تصادفی هر گره (۲=mtree) و تعداد درختان ۱۰۰۰ بوده است. نتایج طبق شکل ۷ نشان داد که درصد سیلت مؤثرترین و هدایت الکتریکی عصاره اشباع خاک کم اثرترین پیش‌بینی‌کننده در برآورد رطوبت نقطه PWP با مدل جنگل تصادفی بوده‌اند.

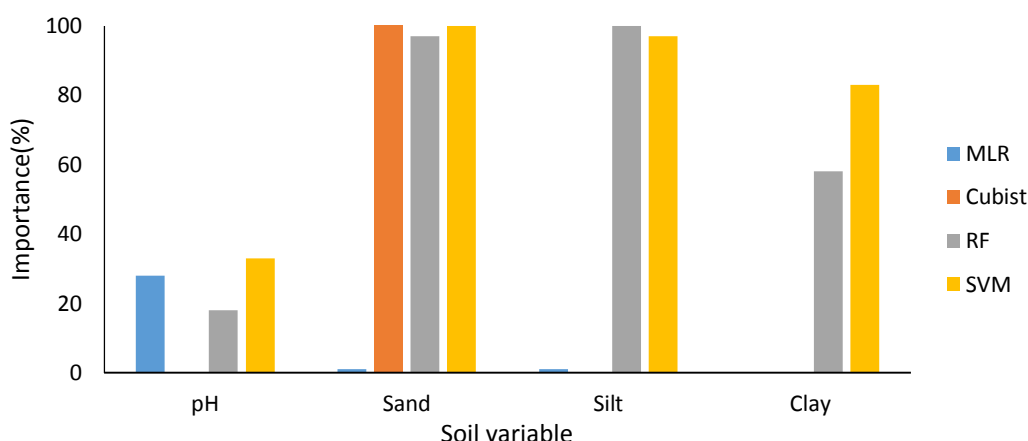
نتایج اهمیت نسبی متغیرهای کمکی در پیش‌بینی PWP با مدل ماشین بردار پشتیبان (SVM) طبق شکل ۷ نشان داد که درصد شن مؤثرترین و هدایت الکتریکی کم اثرترین پیش‌بینی‌کننده در برآورد رطوبت نقطه پژمردگی دائم با ماشین بردار پشتیبان بوده‌اند.

در شکل ۷ اهمیت نسبی متغیرهای مستقل در مدل خطی چندمتغیره برای نقطه پژمردگی دائم نشان داده شده است. ضریب تبیین ( $R^2$ ) محاسبه شده نشان‌دهنده توصیف ۷۰٪ تغییرات با استفاده از مدل است. متغیرهای مهم در این رابطه به ترتیب اهمیت شامل کربنات کلسیم معادل، کربن آلی، هدایت الکتریکی عصاره اشباع، واکنش خاک، درصد شن و سیلت می‌باشند (شکل ۷).

در این پژوهش شبکه عصبی مصنوعی با مشخصاتی شامل لایه اول با ۶ نرون، لایه دوم با ۳ نرون و لایه سوم با ۶ نرون به‌عنوان بهترین مدل شناخته شد. در مدل شبکه عصبی مصنوعی چون مدل دارای ماهیت جعبه سیاه است متغیرهای کمکی مهم قابل استخراج نبود.

علاوه بر این، در این پژوهش بهترین مدل کیوبیست با تعداد کمیتی ۱ و تعداد نزدیک‌ترین همسایه معادل ۳ به‌عنوان بهترین مدل پیش‌بینی PWP





شکل ۷- اهمیت نسبی متغیرهای مستقل در مدل‌های خطی چندمتغیره، کیوبیست، جنگل تصادفی و ماشین بردار پشتیبان برای برآورد PWP.

Figure 7. Relative importance of independent variables in multivariate linear, cubist, random forest and support vector machine models for estimating PWP.

نتایج مندرج در جدول ۵ نشان داد که دقت توابع انتقالی در برآورد PWP بیش‌تر از FC است (ضریب تبیین مدل کیوبیست برای PWP برابر ۰/۸۱ و برای FC برابر ۰/۵۳ می‌باشد). نتایج این پژوهش با نتایج هاتسون و کاس (۱۹۸۷)، مردان و همکاران (۲۰۰۶) و مک براتنی و همکاران (۲۰۱۱) همخوانی داشت (۱۱، ۲۸، ۲۹)، ولی با نتایج پژوهش‌های مصدقی و محبوبی (۲۰۱۱) و سالشو و همکاران (۱۹۹۶) همخوانی نداشت (۳۰، ۳۱). بدین معنی که نتایج آن‌ها نشان داد که توابع انتقالی برای برآورد FC دقت کم‌تری نسبت به PWP داشت. علت این امر را ساختمان بسیار متغیر خاک در پژوهش خود ذکر کردند که این پارامتر تأثیر زیادی روی رطوبت ظرفیت زراعی دارد. میناسنی و مک براتنی (۲۰۰۲) اثر تعداد نمونه بر دقت توابع انتقالی را بررسی کردند. آن‌ها بیان کردند توابع انتقالی حاصل از پایگاه داده‌های بزرگ پیش‌بینی‌های مناسب‌تری را انجام می‌دهند (۳۲). اما بوتولا و همکاران (۲۰۱۲) و مدیروس و همکاران (۲۰۱۴) بیان نمودند که کارایی توابع انتقالی خاک می‌تواند علاوه بر اندازه نمونه تحت تأثیر تغییرپذیری مجموعه داده‌های آموزشی قرار گیرد

بعد از محاسبه رطوبت با توابع انتقالی موردنظر برای تعیین مناسب‌ترین تابع از پارامترهای ریشه میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و ضریب تبیین ( $R^2$ ) استفاده شد. بیش‌ترین مقدار ضریب تبیین در تخمین FC مربوط به مدل شبکه عصبی، کیوبیست و جنگل تصادفی (۰/۵) و کم‌ترین مقدار آن مربوط به مدل خطی و ماشین بردار پشتیبان (۰/۴) است. مدل کیوبیست بالاترین مقدار ضریب تبیین (۰/۸۲) و مدل خطی، شبکه عصبی و ماشین بردار پشتیبان پایین‌ترین مقدار ضریب تبیین (۰/۷) در برآورد PWP داشتند (جدول ۵).

نتایج اعتبارسنجی برای FC نشان داد که مدل‌های کیوبیست و شبکه عصبی مصنوعی دارای میانگین خطا و ریشه میانگین مربعات خطای پایین و ضریب تبیین تصدیق شده بالاتری نسبت به مدل‌های دیگر هستند. هم‌چنین نتایج اعتبارسنجی برای PWP نشان داد که باتوجه به پارامترهای آماری، مدل کیوبیست و پس‌از آن مدل جنگل تصادفی از نظر مقایسه میانگین خطا (۰/۰۴) و (۰/۰۴۴) و ریشه میانگین مربعات خطا (۰/۰۵۴) و (۰/۰۵۷) بهترین مدل برای برآورد رطوبت PWP هستند (جدول ۵).

(۲۶، ۳۳). همان‌طور که ذکر شد در پژوهش حاضر دامنه تغییرات ویژگی‌های خاک برای تولید توابع انتقالی برای FC و PWP در نظر گرفته شد. بالا بودن کارایی توابع انتقالی با وجود اندازه نمونه کم بر اهمیت در نقش تغییرپذیری خصوصیات خاک صحه گذاشت.

جدول ۵- نتایج اعتبارسنجی مدل‌های موردبررسی برای برآورد FC و PWP

Table 5. Validation results of the studied models for estimating FC and PWP.

R <sup>2</sup>	MAE	RMSE	Model Type	واحد	Parameter
0.40	0.070	0.100	خطی Linear		
0.54	0.068	0.085	شبکه عصبی Neural Network		
0.53	0.066	0.085	کیوبیست Cubist	cm <sup>3</sup> . cm <sup>-3</sup>	ظرفیت زراعی Field Capacity (FC)
0.51	0.072	0.087	جنگل تصادفی Random Forest		
0.42	0.075	0.095	ماشین بردار پشتیبان Support Vector Machine		
0.71	0.048	0.067	خطی Linear		
0.70	0.046	0.066	شبکه عصبی Neural Network		
0.81	0.040	0.054	کیوبیست Cubist	cm <sup>3</sup> . cm <sup>-3</sup>	نقطه پژمردگی دائمی Permanent Wilting Point (PWP)
0.78	0.044	0.057	جنگل تصادفی Random Forest		
0.73	0.049	0.064	ماشین بردار پشتیبان Support Vector Machine		

### نتیجه‌گیری کلی

منطقه روانسر در استان کرمانشاه موردبررسی قرار گرفت. به این منظور با استفاده از ویژگی‌های زودیافت خاک شامل کربن آلی خاک، کربنات کلسیم معادل، واکنش خاک و هدایت الکتریکی عصاره گل اشباع و پنج مدل تابع انتقالی شامل مدل خطی چندمتغیره، شبکه عصبی مصنوعی، کیوبیست، جنگل

در این پژوهش باتوجه به اهمیت رطوبت خاک در برنامه‌ریزی‌های آبیاری، امکان استفاده از خصوصیات فیزیکی زودیافت خاک و توابع انتقالی در برآورد خصوصیات دیریافت خاک مانند نقاط مهم رطوبتی ظرفیت زراعی و نقطه پژمردگی دائم برای خاک‌های

### داده‌ها، اطلاعات و دسترسی

پژوهش حاضر حاصل پایان‌نامه کارشناسی ارشد است که در آزمایشگاه‌های گروه مهندسی آب دانشگاه رازی و بخش تحقیقات خاک و آب مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کرمانشاه به اجرا درآمده است.

### تعارض منافع

در این مقاله تضاد منافی وجود ندارد و این مسأله مورد تأیید همه نویسندگان است.

### مشارکت نویسندگان

نویسنده اول: دانشجوی کارشناسی ارشد است که در همه مراحل پژوهش مشارکت داشته است.

نویسندگان دوم و سوم: اساتید راهنمای این پایان‌نامه می‌باشند و در همه موارد انجام پژوهش نظارت و مشارکت داشته‌اند.

نویسنده چهارم: دستیار پژوهشی نویسنده مسئول در دانشگاه رازی است که بعد از دفاع دانشجو نسبت به انجام برخی آنالیزهای تکمیلی، به‌روزرسانی بررسی منابع و تدوین مقاله زیر نظر نویسندگان دوم و سوم پرداخته است.

### اصول اخلاقی

نویسندگان اصول اخلاقی را در انجام و انتشار این اثر عملی رعایت نموده‌اند و این موضوع مورد تأیید همه آن‌ها است.

### حمایت مالی

این پژوهش با حمایت مالی دانشگاه رازی در قالب کمک‌هزینه انجام پروپوزال‌های دانشجویان تحصیلات تکمیلی انجام شده است.

تصادفی و ماشین بردار پشتیبان، اقدام به برآورد درصد رطوبت خاک در ظرفیت زراعی و نقطه پژمردگی دائم شد. تحلیل نتایج نشان داد مدل کیویست و مدل شبکه عصبی مصنوعی و پس‌از آن مدل جنگل تصادفی با خطای کم‌تر و ضریب تبیین بالاتر نسبت به سایر مدل‌ها دارای کارایی بالاتری برای برآورد رطوبت ظرفیت زراعی می‌باشند. در مورد رطوبت نقطه پژمردگی دائم با توجه به پارامترهای آماری مدل کیویست و پس‌از آن مدل جنگل تصادفی از نظر ضریب تبیین بهترین مدل برای برآورد رطوبت PWP بودند. در پژوهش‌های گذشته کم‌تر از مدل‌های درختی به‌عنوان تابع انتقالی استفاده شده است و این در حالی است که مدل درختی معایب جعبه سیاه بودن شبکه عصبی مصنوعی و شرط نرمال بودن داده‌ها در مدل خطی را ندارد. این پژوهش بر این امر صحه گذاشت و نشان داد که استفاده از روش‌های نوین یادگیری ماشین می‌تواند در توسعه توابع انتقالی خاک برای برآورد خصوصیات دیریافت خاک مهم باشد. نتایج این پژوهش برای دامنه وسیعی از دشت‌های استان کرمانشاه که شرایط تشکیل خاک مشابه با منطقه روانسر را دارند، قابل قبول است.

### تقدیر و تشکر

از آقای دکتر حسین میرزایی کارشناس گروه مهندسی آب دانشگاه رازی که در انجام آزمایش‌های انجام شده با دستگاه صفحات فشاری، نویسنده اول را یاری نمودند تشکر می‌گردد. از مدیریت محترم مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کرمانشاه به خاطر هماهنگی در خصوص استفاده از امکانات آزمایشگاهی آن مرکز تشکر می‌شود.

منابع

1. Schaap, M.G., and Leij, F.J. 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil and Tillage Research*. 47: 1-2. 37-42.
2. Nemes, A., and Rawls, W.J. 2006. Evaluation of different representations of the particle-size distribution to predict soil water retention. *Geoderma*. 132: 1-2. 47-58.
3. Tomasella, J., Pachepsky, Y., Crestana, S., and Rawls, W.J. 2003. Comparison of two techniques to develop pedotransfer functions for water retention. *Soil Science Society of American Journal*, 67: 4. 1085-1092.
4. Bayat, H., Neyshabouri, M., Mohammadi, K., Nariman-Zadeh, N., Irannejad, M., and Gregory, A.S. 2013. Combination of artificial neural networks and fractal theory to predict soil water retention curve. *Computers and Electronics in Agriculture*. 92: 92-103.
5. Twarakavia, N.K.C., Simunek, J., and Schaap, M.G. 2009. Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. *Soil Science Society of American Journal*. 73: 5. 1443-1452.
6. Babaeian, E., Homae, M., and Norouzi, A.A. 2014. Evaluating point and parametric spectral transfer functions for a prediction of soil water characteristics. *Iranian Journal of Soil and Water Research*. 45: 4. 475-490. (In Persian)
7. Mehrabi Gohri, A., Sarmidian, F., and Taghizadeh Mehrjardi, R. 2013. Prediction of the amount of water at Field Capacity and Permanent Wilting Point using Artificial Neural Network and Multivariate Regression. *J. of Irrigation and water engineering*. 3: 10. 42-52. (In Persian)
8. Amir Abedi, H., Asghari, S., Mesri Gandoshmin, T., and Keivan Behjo, F. 2013. Estimating of field capacity, permanent wilting and available water content in Ardabil plain soils using regression and artificial neural network models. *Applied Soil Research*. 1: 1. 60-72. (In Persian)
9. Norouzi Engnaee, O., Khalafi, M., and Karimi Soorvand, M. 2019. Investigating the performance of data-based methods in estimating important moisture points in Shahrood area. *Journal of Irrigation Sciences and Engineering*. 42: 4. 29-44. (In Persian)
10. Lieb, M., Glaser, B., and Huwe, B. 2012. Uncertainty in the spatial prediction of soil texture comparison of regression tree and random forest models. *Geoderma*. 170: 70-79.
11. McBratney, A.B., Minasny, B., and Tranter, G. 2011. Necessary meta-data for pedotransfer functions. *Geoderma*. 160: 3-4. 627-629.
12. Rastgou, M., Bayat, H., Mansoorizadeh, M., and Gregory, Andrew S. 2020. Estimating the soil water retention curve: Comparison of multiple nonlinear regression approach and random forest data mining technique. *Computers and Electronics in Agriculture*. 174: 1-13.
13. Cueff, S., Coquet, Y.B., Aubertot, J.N., Bel, L., Pot, V., and Alletto, L. 2021. Estimation of soil water retention in conservation agriculture using published and new pedotransfer functions. *Soil and Tillage Research*. 209: 104967.
14. Abdelhafid, Y., Chebbah, M., and Rechachi, M.Z. 2021. Comparison of regression methods for predicting soil water contents at field capacity and wilting point in Bas Sahara of Algeria. *International Journal of Forest, Soil and Erosion*. 11: 2. 45-62.
15. Fritsch, S., Guenther, F., and Wright, M.N. 2019. neuralnet: Training of Neural Networks. URL <https://CRAN.R-project.org/package=neuralnet>. R package version 1.44.2.
16. Quinlan, J.R. 1992. Learning with continuous classes. P 343-348, In: *Proceedings of 5th Australian conference on artificial intelligence*. World Scientific. Singapore.
17. Zhou, Zhi-Hua. 2012. *Ensemble Methods: Foundations and Algorithms*. Chapman and Hall/CRC. 23p.

18. Kuhn, M., Weston, S., Keefer, C., Coulter, N., and Quinlan, R. 2023. Cubist: Rule-and Instance-based Regression Modeling, R package version 0.4.2.1. (<https://cran.r-project.org/web/packages/Cubist/Cubist.pdf>. Last access date: 3 May 2023).
19. Breiman, L. 2001. Random forests. *Machine Learning*. 45: 1. 5-32.
20. Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel B., Team, R.C., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., and Hunt, T. 2023. caret: Classification and Regression Training. R package version 6.0-94. (<https://cran.r-project.org/web/packages/caret/caret.pdf>. Last access date: 3 May 2023).
21. Chang, C.C., and Lin, C.J. 2001. Libsvm: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2: 3. 1-27.
22. Cortes, C., and Vapnik, V. 1995. Support-vector networks. *Machine Learning*. 20: 3. 273-297.
23. Brus, D., Kempen, B., and Heuvelink, G. 2011. Sampling for validation of digital soil maps. *European Journal of Soil Science*. 62: 3. 394-407.
24. Manyam, C., Morgan, C.L., Heilman, J.L., Fatondji, D., Gerard, B., and Payne, W.A. 2007. Modeling hydraulic properties of sandy soils of Niger using pedotransfer functions. *Geoderma*. 141: 3-4. 407-415.
25. Gomes, F.P., and Garcia, C.H. 2002. *Estatística Aplicada an Experimentos Agronomicos e Florestais*. FEALQ: Piracicaba, 309p. (In Portuguese)
26. Botula, Y.D., Cornelis, W.M., Baert, G., and Van Ranst, E. 2012. Evaluation of pedotransfer functions for predicting water retention of soils in Lower Congo (D.R. Congo). *Agriculture Water Management*, 111: 1-10.
27. Minh Nguyen, P., Van Le, Kh., Dady Botula, Y., and M. Cornelis, W. 2015. Evaluation of soil water retention pedotransfer functions for Vietnamese Mekong Delta soils. *Agricultural Water Management*. 158: 126-138.
28. Hutson, J.L., and Cass, A. 1987. A retentivity function for use in soil-water simulation models. *Journal of Soil Science*. 38: 1. 105-113.
29. Merdun, H., Cinar, O., Meral, R., and 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: 1-2. 108-116.
30. Mosaddeghi, M.R., and Mahboubi, A.A. 2011. Point pedotransfer functions for prediction of water retention of selected soil series in a semi-arid region of western Iran. *Archives of Agronomy and Soil Science*. 57: 4. 327-342.
31. Salchow, E., Lal, R., Fausey, N.R., and Ward, A. 1996. Pedotransfer functions for variable alluvial soils in Southern Ohio. *Geoderma*. 73: 3-4. 165-181.
32. Minasny, B., and McBratney, A. B. 2002. The Neuro-m Method for Fitting Neural Network Parametric Pedotransfer Functions. *Soil Science Society American Journal*. 66: 2. 352-361.
33. Medeiros, J.C., Cooper, M., Dalla Rosa, J., Grimaldi, M., and Coquet, Y.M. 2014. Assessment of pedotransfer functions for estimating soil water retention curves for the amazon region. *Rev. Bras. Cien. Solo*. 38: 3. 730-743.

