

EXTENDED ABSTRACT

Investigating the Performance of Data-based Methods in Estimating Important Moisture Points in Shahrood Area

O. Norouzi Engenayi^{1*} M. J. Khalafi Khotbsara² and M. Karimi Sorand³

1* -Corresponding Author, MSc of Irrigation Drainage, Zabol University, Iran
(omidnoroozi67@gmail.com)

2 - MSc of Watershed Management, Zabol University, Iran.

3 - MSc of Soil Engineering, Shahrood University, Iran.

Received: 4 July 2017

Revised: 16 December 2017

Accepted: 20 December 2017

Keywords: Linear regression, Artificial neural networks, Support vector regression, Field capacity, Permanent wilting point. **DOI:** 10.22055/jise.2017.22685.1617.

Introduction

Awareness of the important moisture points is crucial for irrigation studies on the farm, but measuring this information in a direct way is very costly and time consuming. Therefore, several models and relationships have been developed as Pedotransfer functions which indirectly predict the hydrological properties of the soil using readily available soil data with the aid of a series of proper mathematical relationships (Nguyen et al., 2015). Since the measurement of important moisture points is a time consuming, costly and difficult work, many attempts have been made in order to use simpler soil properties such as texture, the amount of organic matter, and bulk density. Pedotransfer functions are indeed predictive functions which establish relationship between the soil's readily available and latency data (e.g., the percentage of sand, silt and clay, bulk density and organic matter) including the parameters of the moisture curve (field capacity and permanent wilting point) (Botulla et al., 2013). Moreover, the functions that can be successfully implemented in an area may not have suitable adaptations in another area with real values. There are several methods for obtaining Pedotransfer functions, among them are linear regression (LR), artificial neural networks, fuzzy adaptive-neural inference, and support regression vector.

Various researchers have studied the development of Pedotransfer functions and evaluated the predictive models in the water and soil sciences. As a sample, Shop and Lajj (1998) estimated the soil moisture curve using the neural network. They found that the artificial neural network was better than some of the regression Pedotransfer functions provided by other researchers, and if more readily available properties were used as inputs, the prediction accuracy increased. However, there was always a significant difference between the predicted and measured moisture values. Zhang et al. (2007) estimated the soil moisture curve for 110 non-calcareous soil samples with different tissue classes through the artificial neural networks and regression models. They showed that the neural network predicts the moisture curve better than the regression method with higher correlation coefficient in most tissue classes. Lin et al. (2009) argued that the SVM method was much faster trained than the artificial neural network. SVM was also found to have a more accurate prediction than the artificial neural network. Chen et al. (2010) used support vector machines to model daily rainfall and compared the results with that of the multivariate analysis method. It was found that the results of predictions from SVM were more accurate. In turn, Kaihua et al. (2014) used support

vector machines to predict cationic capacity on different horizons of the soil in Qingdao, China. They performed their studies at 208 points on two horizons of the soil, and concluded that the SVM model improved predictions.

Considering the significance of knowing the important points of soil moisture in Shahrood area for agricultural projects and irrigation schedules, developing appropriate Pedotransfer functions and evaluating models is necessary so as to obtain moisture of the field capacity and permanent wilting point. This research, thus, evaluates the performance of three models of support vector regression, artificial neural networks, and linear regression in the development of soil Pedotransfer functions and the effect of number and type of input variables on the performance of the models.

Methodology

In this research, early and late found data included field capacity and permanent wilting point related to soil studies in Shahrood area available at the Agricultural Research and Education Center of Semnan province. Data were collected from 100 profiles of a total area of 120 hectares and from different areas of Shahrood. To this end, disturbed and intact soil samples were collected from the surface depth (0-15 centimeters) and transferred to the laboratory for the measurement of physical, chemical, and hydraulic properties of soil. Field capacity and permanent wilting point were obtained by measuring the soil moisture in suctions 0.33 and 15 bar using a pressure plate. The first step for the development of Pedotransfer functions was to provide the empirical relationships between the basic properties of soil and the factors to be predicted. This can be achieved by various mathematical methods (such as support vector regression and artificial neural networks).

Evaluation of data-based methods

In the present study, the data were divided into two categories of training data and test data. The training of the model was carried out with 80% of the data and the performance of the methods was evaluated using 20% of the remaining data. In order to compare the measured values with the predicted values, the normalized root-mean-square deviation or error or NRMSE, the coefficient of determination and/or Root Mean Square Error (MSD) deviation were calculated. The more the values of the coefficient of determination and the efficiency of the Nash are closer to the number one, the higher will be the performance and accuracy of the model. The NRMSE value is between zero and one, which can be expressed as a percentage, so the lower the value, the better the estimate will be. Generally, the values less than 10 percent have an ideal estimation accuracy in modeling. Values between 10% to 20% and 20% to 30%, are good and average, respectively, and over 30% lack validation in modeling (Jafari Gilandeh et al., 2017).

Results and Discussion

To obtain more accurate models and to normalize the data, the zero and one method was used. Based on the correlation analysis of sand percentage, Clay percentage, Silt percentage, Bulk Density Special crime, and Organic Carbon percentage (OC), five different variables for input variables were selected to estimate the important moisture points. The results showed that the support regression method had a better performance than the other two. The values of coefficient of determination, the deviation of error root mean square, and the average root mean squared error normalized in the best model of support regression vector were 0.85, 3.21 and 12.89 for the field capacity, 0.83, 1.58, 14.84 for the permanent wilting point, respectively. For neural networks for which results were far better than the linear regression's, the values of 0.72, 3.48 and 14.36 for field capacity and 0.75, 1.90 and 17.91 for permanent wilting point were achieved, respectively. According to the results, it can be stated that the regression models of the support vector with linear kernel function of the radial base can predict the soil moisture points with low error and high coefficient of determination. Thus, they can also be a great alternative to traditional methods such as neural networks and linear regression. In

the data-based methods, the increase in inputs resulted in an increase in the coefficient of determination and reduction of errors, which is recommended using more early accessible variables.

This research was intended to evaluate the performance of traditional linear regression and artificial neural network methods with the new method of support vector machine in estimating the critical points of moisture. Besides, using the Pedotransfer functions method, the readily available data as input and the field capacity and permanent wilting point as latency data or the output data were used. In summary, the results can be summarized as follows: Regression models of support vector with radial root linear kernel function can predict soil moisture points with low error and high coefficient of determination. In addition, these models at the training stage had better results with higher coefficients which could be a good alternative for traditional methods such as neural networks and linear regressions requiring data with high data numbers. In the data-based methods, the increase in input increased the coefficient of determination and reduced the errors. Accordingly, the use of readily available variable is much more recommended. However, there is a probability that in the regression method of the support vector or artificial neural network MLP the over-training phenomenon happens in relation to the training data. More studies and the use of other input parameters are, thus, required. Several researchers have suggested that other optimization methods be used to select the predictive variables and be referred to as SVM or ANN inputs. According to the efficiency of the support vector machine model, proven by numerous research in various applications, it is recommended to use this method to create soil Pedotransfer functions in different regions. Besides, it is suggested that a vaster database, and high initial correlation be used for better training of the models.

Acknowledgment

The authors of this article would like to thank Mr. Ahmad Akhiani, Research Assistant, in Semnan Province Agricultural Research and Education Center.

References

- 1- Botula, Y. D., Nemes, A., Mafuka, P., Van Ranst, E. and Cornelis, W., 2013. Prediction of water retention of soils from the humid tropics by the nonparametric neighbor approach. *Vadose Zone Journal*, 12(2), pp. 1-17.
- 2- Chen, S., Yu, P. and Tang, H., 2010. Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology*, 385, pp. 13-23.
- 3- Kaihua, L., Shaohui, X., Jichun, W., Qing, Z. and N. Lesheng., 2014. Using support vector machines to predict cation exchange capacity of different soil horizons in Qingdao City, China. *Journal Plant Nutr Soil Sci*, 177, pp. 775–782.
- 4- Lin, G., Chen, G., Huang, P. and Chou, Y., 2009. Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods. *Journal of Hydrology*, 372, pp. 17-29.
- 5- Jafari Gilandeh, S., Khodaverdillo, H. and Rasolzadeh, A., 2017. Application and comparison of parametric transfer functions of Van Genuchten model in simulating unsteady water flow in cultivated soil. *Soil Knowledge Journal*, 25 (2), pp. 82 - 92. (In Persian).
- 6- Nguyen, P.M., De Pue, J., Van, K. L. and Cornelis, W., 2015. Impact of regression methods on improved effects of soil structure on soil water retention estimates, *Journal of Hydrology*, 29, pp. 598-606.

-
- 7- Schaap, M. G. and F. Leij., 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil and Tillage Research*, 47, pp. 37-42.
- 8- Zhang, Y., 2007. Artificial neural networks based on principal component analysis input selection for clinical pattern recognition analysis. *Talanta*, 73 (1), pp. 68-75.



© 2019 by the authors. Licensee SCU, Ahvaz, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0 license) (<http://creativecommons.org/licenses/by/4.0/>).



بررسی عملکرد روش‌های داده مبنا در تخمین نقاط مهم رطوبتی در منطقه شاهرود

امید نوروزی انگنایی^{۱*}، محمدجواد خلفی خطبوسرا^۲ و محبوبه کریمی سورند^۳

*۱- نویسنده مسئول، دانش آموخته کارشناسی ارشد مهندسی آب، دانشگاه زابل،omidnoroozi67@gmail.com

۲- دانش آموخته کارشناسی ارشد آبخیزداری، دانشگاه زابل.

۳- دانش آموخته کارشناسی ارشد مهندسی خاک، دانشگاه صنعتی شاهرود.

پذیرش: ۱۳۹۶/۹/۲۹

بازنگری: ۱۳۹۶/۹/۲۵

دریافت: ۱۳۹۶/۴/۱۳

چکیده

آگاهی از نقاط مهم رطوبتی، برای مطالعه‌های آبیاری در مزرعه بسیار ضروری می‌باشد اما اندازه‌گیری این اطلاعات به روش مستقیم بسیار پرهزینه و وقت‌گیر است. روش‌های داده مبنا می‌توانند روش مناسبی برای تخمین این پارامترها باشد. تحقیق حاضر به برآورد نقاط مهم رطوبتی شامل ظرفیت زراعی و نقطه پژمردگی دایم به وسیله پارامترهای زودیافت با سه روش شبکه عصبی، رگرسیون خطی چندمتغیره و رگرسیون بردار پشتیبان در منطقه شاهرود پرداخته است. پس از نرمال‌سازی داده‌های مورد نظر جدول ضریب همبستگی متغیرهای ورودی احتمالی با خروجی‌های مورد نظر تشکیل شد و معنی‌داری همبستگی متغیرهای ورودی و خروجی از نظر آماری بررسی گردید. سپس، مدل‌سازی با روش‌های مذکور انجام و نتایج مورد ارزیابی قرار گرفت. نتایج نشان داد که روش رگرسیون بردار پشتیبان کارایی بهتری نسبت به دو روش دیگر دارد. مقادیر ضریب تعیین، انحراف جذر میانگین مربعات خطا و ریشه میانگین مربعات خطا نرمال‌شده در بهترین مدل رگرسیون بردار پشتیبان، به ترتیب برابر 0.85 ، $3/12$ و $12/89$ برای ظرفیت زراعی و 0.83 ، 0.58 و $14/84$ برای نقطه پژمردگی دایم و برای شبکه‌های عصبی مقادیر 0.72 ، $3/48$ و $14/36$ برای ظرفیت زراعی و 0.75 ، $1/90$ و $17/91$ برای نقطه پژمردگی به دست آمد. با توجه به بررسی‌های صورت گرفته در این تحقیق، می‌توان بیان نمود که مدل‌های رگرسیون بردار پشتیبان با تابع کرنل خطی پایه شعاعی قادر خواهند بود با خطای پایین و ضریب تعیین بالا نقاط مهم رطوبتی خاک را پیش‌بینی کنند و همچنین می‌توانند جایگزین بسیار خوبی برای روش‌های سنتی هم‌چون شبکه‌های عصبی و رگرسیون خطی باشند.

کلید واژه‌ها: رگرسیون خطی، شبکه‌های عصبی مصنوعی، رگرسیون بردار پشتیبان، ظرفیت زراعی، نقطه پژمردگی دایم.

مقدمه

ظاهری به دست آورد. توابع انتقالی خاک در واقع توابع پیش‌بینی‌کننده می‌باشند که ارتباط بین داده‌های زودیافت خاک (مانند درصد شن، سیلت و رس، جرم مخصوص ظاهری و مواد آلی) را با داده‌های دیریاخت خاک مانند پارامترهای منحنی رطوبتی (ظرفیت زراعی و نقطه پژمردگی دایم) برقرار می‌سازد (Botula et al., 2013). البته توابعی که می‌توانند در یک منطقه به طور موفقیت آمیزی به کار روند، ممکن است در منطقه جدید سازگاری مناسبی با مقادیر واقعی نداشته باشند. روش‌های مختلفی برای به دست آوردن توابع انتقالی وجود دارد که از جمله می‌توان به مدل‌های رگرسیون خطی (Linear Regression (LR)، شبکه‌های عصبی مصنوعی (Artificial Neural Networks (ANN)، استنتاج تطبیقی-عصبی فازی (Adaptive Neuro Fuzzy Inference System (ANFIS) و رگرسیون بردار پشتیبان (Support Vector Regression (SVM) اشاره نمود. محققین مختلفی به توسعه توابع انتقالی و بررسی مدل‌های پیش-بینی کننده در علوم آب و خاک پرداخته‌اند. به عنوان مثال: Leji و Schaap (1998) با استفاده از شبکه عصبی منحنی رطوبتی خاک را تخمین زدند. آن‌ها دریافتند که شبکه عصبی

منحنی مشخصه آب خاک که بیانگر تغییرات رطوبت در برابر مکش آب خاک است، در مسائل مربوط به حرکت آب در خاک غیراشباع کاربرد فراوان دارد و شناسایی رفتار فیزیکی خاک در رطوبت‌های مختلف را آسان تر می‌کند. نقاط مهم رطوبتی شامل ظرفیت زراعی و نقطه پژمردگی دایم از عوامل مهم در رسم منحنی مشخصه رطوبتی هستند. اندازه‌گیری نقاط مهم رطوبتی در آزمایشگاه وقت‌گیر و پرهزینه است، بنابراین، مدل‌ها و روابطی تحت عنوان توابع انتقالی (Pedotransfer functions) توسعه یافته‌اند که خصوصیات هیدرولیکی خاک را به طور غیر مستقیم با استفاده از داده‌های زودیافت خاک و با کمک یک سری روابط ریاضی مناسب پیش‌بینی می‌کنند (Nguyen et al., 2015). توابع انتقالی اولین بار توسط Boma پیشنهاد شد و در مسائل مربوط به فیزیک خاک به کار گرفته شد (Minasny and Mcbratney., 2002). از آن‌جا که اندازه‌گیری نقاط مهم رطوبتی، کاری وقت‌گیر، پرهزینه و دشوار می‌باشد، کوشش‌های زیادی به عمل آمده است که بتوان این نقاط را با استفاده از ویژگی‌های ساده‌تر خاک مانند بافت، مقدار ماده آلی و چگالی

Kakaeilafdani et al. (2013) با استفاده از روش رگرسیون بردار پشتیبان و شبکه عصبی به منظور تخمین بار رسوبی رودخانه دوراج پرداختند. آنها کارایی بالای هر دو روش را گزارش نمودند.

Kaihua et al. (2014) با استفاده از ماشین های بردار پشتیبان، به پیش بینی ظرفیت کاتیونی در افق های مختلف خاک در شهر چینگدائو چین پرداختند. آن ها بررسی های خود را در ۲۰۸ نقطه و در دو افق خاک انجام دادند و در پایان بیان نمودند که مدل SVM پیش بینی ها را بهبود می بخشد.

با توجه به اهمیت آگاهی از نقاط مهم رطوبتی خاک در منطقه شاهرود برای طرح های کشاورزی و برنامه ریزی های آبیاری و دشواری اندازه گیری مستقیم آن ها، توسعه توابع انتقالی مناسب و یا ارزیابی مدل ها در به دست آوردن رطوبت ظرفیت زراعی و نقطه پژمردگی داریم در این منطقه لازم می باشد. بنابراین در این تحقیق، عملکرد دو مدل شبکه های عصبی مصنوعی و رگرسیون خطی با روش نوین رگرسیون بردار پشتیبان، در توسعه توابع انتقالی خاک و اثر تعداد و نوع متغیرهای ورودی بر عملکرد مدل های مورد استفاده، مورد ارزیابی قرار می گرفت. لازم به ذکر است استفاده از روش رگرسیون بردار پشتیبان در توسعه توابع انتقالی خاک در منابع داخلی گزارش نشده است.

مواد و روش ها

در این تحقیق از داده های زودیافت و دیریافت شامل ظرفیت زراعی و نقطه پژمردگی داریم، مربوط به مطالعات خاکشناسی شهرستان شاهرود موجود در مرکز تحقیقات و آموزش کشاورزی استان سمنان استفاده شده است. داده ها از ۱۰۰ نمونه پروفیل از مساحتی بالغ بر ۱۲۰ هکتار و از مناطق مختلف شهرستان شاهرود جمع آوری شده است. بدین منظور از عمق سطحی (۱۵-۰ سانتی متر) نمونه های دست خورده و دست نخورده خاک جمع آوری شده و برای اندازه گیری خصوصیات فیزیکی و شیمیایی و هیدرولیکی خاک به آزمایشگاه منتقل گردید. ظرفیت زراعی و نقطه پژمردگی داریم با اندازه گیری رطوبت خاک در مکش های به ترتیب ۰/۳۳ و ۱۵ بار با استفاده از دستگاه صفحات فشاری به دست آمد. خصوصیات زودیافت خاک شامل بافت خاک (درصد سیلت، شن و رس)، جرم مخصوص ظاهری و موادالی خاک به ترتیب به روش های هیدرومتری (Gee and Bauder, 1986)، استوانه با حجم مشخص (Blake and Hartge, 1986) اندازه گیری شدند.

اولین گام برای توسعه توابع انتقالی، ارایه روابط تجربی میان خصوصیات پایه خاک و عواملی است که باید پیش بینی شوند. این امر می تواند به وسیله روش های مختلف ریاضی (مانند رگرسیون بردار پشتیبان و شبکه های عصبی مصنوعی) محقق شود. در ادامه توضیح مختصری در مورد روش مورد استفاده در این پژوهش ارایه می گردد:

مصنوعی به کار رفته بهتر از تعدادی از توابع انتقالی رگرسیونی ارائه شده توسط سایر محققان بوده و معمولاً اگر خصوصیات زودیافت بیشتری به عنوان ورودی به کار رود، صحت پیش بینی ها افزایش می یابد ولی همیشه اختلاف قابل توجهی بین مقادیر رطوبت پیش بینی شده و اندازه گیری شده وجود دارد.

Vali et al. (2009) بیان نمودند که ماشین های بردار پشتیبان برای داده هایی با ابعاد بالا به خوبی پاسخ گو می باشد، همچنین مدلی را انتخاب می کند که بین میزان خطای تجربی و تقریب یک حالت بهینه را انتخاب کند.

Booltink و Koekkoek (1999) نشان دادند که استفاده از شبکه های عصبی مصنوعی برای تخمین منحنی رطوبتی خاک های با مواد آلی بالا و غیرآهکی نتایج بهتری از توابع انتقالی ارایه می دهد. همچنین Nemes et al. (2003) عنوان نمودند که اشتقاق توابع انتقالی با استفاده از داده های کم ولی مناسب بهتر از استفاده از پایگاه داده های بزرگتر ولی عمومی و نامناسب می باشد.

Ungaro et al. (2005) بیان نمودند که با توجه به این مطلب که خصوصیات جغرافیایی و منطقه ای تنوع زیادی را در خاک ها پدید می آورد، اعتمادپذیر بودن توابع انتقالی رابطه ای بسیار قوی با مجموعه داده هایی دارد که برای واسنجی آن ها به کار می رود. لذا خصوصاً در هنگامی که از شبکه های عصبی استفاده می شود، بررسی کارایی شبکه آموزش داده شده با پایگاه های داده عظیم موجود، برای نمونه های خاک محلی بسیار مهم می باشد. به طوری که حتی ممکن است نتایج ضعیفی از اجرای شبکه های عصبی به دست آید. از طرفی به علت اینکه شبکه های عصبی نیازمند تعداد داده های زیادی برای آموزش می باشند، کاربرد آن ها در مقیاس های کوچک و سطح مزرعه صرفاً با استفاده از داده های موجود مورد سوال می باشد.

Zhang (2007) منحنی رطوبتی خاک را برای تعداد ۱۱۰ نمونه خاک غیرآهکی با کلاس های بافتی مختلف به روش شبکه عصبی مصنوعی و رگرسیونی تخمین زدند و نشان دادند شبکه عصبی منحنی رطوبتی را در بیشتر کلاس های بافتی بهتر از روش رگرسیونی و با ضریب همبستگی بالاتری تخمین می زند.

Noori et al. (2013) برای کاهش متغیرهای ورودی به مدل ماشین بردار پشتیبان در پیش بینی میزان زباله های تولیدی در شهر مشهد از تحلیل مولفه های اصلی (Component Principal Analysis) استفاده نمودند و با مقایسه ی مدل ترکیبی PCA - SVM با مدل SVM برتری مدل PCA - SVM را نتیجه گرفتند.

Lin et al. (2009) بیان نمودند که نتایج گویای این واقعیت است که روش SVM خیلی سریع تر از شبکه عصبی مصنوعی آموزش می بیند. همچنین SVM پیش بینی دقیق تری را نسبت به شبکه عصبی مصنوعی نشان داده است.

Chen et al. (2010) از ماشین های بردار پشتیبان برای مدل کردن بارش روزانه استفاده کردند و نتایج را با نتایج حاصل از روش آنالیز چند متغیره مقایسه نمودند و نشان دادند نتایج پیش بینی های حاصله از SVM دقیق تر است.

مورد استفاده جهت آموزش از نوع الگوریتم لونیبرگ-مارکواد بوده است. الگوریتم لونیبرگ-مارکواد سریع‌ترین نرخ همگرایی را در مسائل فنی و مهندسی دارد. الگوریتم لونیبرگ-مارکواد این حسن را دارد که در شبکه تغییر میزان نرخ یادگیری به صورت تطبیقی است. روش لونیبرگ-مارکواد روشی است که در آن برای اصلاح فراسنج‌ها علاوه بر مشتق اول (ماتریس ژاکوبین) از مشتق‌های دوم نیز (موسوم به ماتریس هسین) استفاده می‌شود (MahdaviMeymand and Ahadian, 2015).

مدل‌سازی بر پایه ماشین‌های بردار پشتیبان

ساختار مدل ماشین‌های بردار پشتیبان (SVM) توسط وپنیک که یک ریاضیدان روسی در دهه 1960 بود، ارایه شد. مدل‌های SVM که بر پایه‌ی تئوری آموزش احتمال قرار دارند، یک کلاس جدید از مدل‌هایی هستند که به منظور طبقه‌بندی و پیش‌بینی در علوم مختلف استفاده می‌شوند (Vapnik and Cortes, 1995). اخیراً این مدل‌ها در یک گستره وسیع از مسائل هیدرولوژیکی مورد استفاده قرار گرفته‌اند. مشابه با هر مدل ریاضی و آماری، مدل‌های SVM نیز دارای معایبی هستند. وجود تعداد زیاد متغیر ورودی اصلی‌ترین مشکل در توسعه این مدل‌هاست. تعداد زیاد متغیرهای ورودی ممکن است مانع یافتن مدل بهینه توسط SVM شود. راه‌های متعددی برای کاهش تعداد متغیرهای ورودی پیشنهاد شده است که از آن جمله می‌توان به روش تحلیل مؤلفه‌های اصلی اشاره نمود (Zhang, 2007). ویژگی مهم ماشین بردار پشتیبان این است که برخلاف الگوریتم‌های کلاسیک و رگرسیون‌های خطی که به وسیله حداقل کردن قدرمطلق خطا یا توان دوم خطا عمل می‌کنند، ریسک عملیاتی را حداقل می‌کنند. ماشین بردار پشتیبان برخلاف شبکه‌های عصبی با مشکل گیر افتادن در حداقل‌های محل تابع خطا مواجه نمی‌باشد (Yin and Log, 2011 و Hong, 2011). شکل (۲) ساختار ماشین بردار پشتیبان را نشان می‌دهد.

مدل‌های ماشین‌های بردار پشتیبان به دو گروه عمده، مدل طبقه‌بندی ماشین بردار پشتیبان و مدل رگرسیون ماشین بردار پشتیبان تقسیم‌بندی می‌شود. از مدل طبقه‌بندی ماشین بردار پشتیبان جهت حل مسائل طبقه‌بندی داده‌هایی که در کلاس‌های مختلف قرار می‌گیرند استفاده می‌گردد و مدل رگرسیون ماشین بردار پشتیبان در حل مسائل پیش‌بینی کاربرد دارد. همان‌گونه که بیان شد، ماشین بردار پشتیبان مبتنی بر حداقل کردن ساختاری ریسک می‌باشد که از تئوری آموزش آماری گرفته شده است. Vapnik و Cortes (1995) برای کاربرد ماشین‌های بردار پشتیبان در مسائل رگرسیون از تابع خطایی استفاده کرد که خطاهایی که در یک ϵ -Insensitive به نام فاصله معین از مقادیر

رگرسیون خطی چند متغیره (Multivariate linear regression)

در روش رگرسیون خطی چند متغیره (MLR)، مقادیر متغیر خروجی بر اساس داده‌های ورودی و خطوط تصمیم‌گیری پیش‌بینی می‌شوند. در واقع روش رگرسیون، مجموعه‌ای از داده‌های آزمایشی را دریافت نموده و فرمولی را به آن داده‌ها برآزش می‌نماید. رگرسیون می‌تواند برای حل مسائل کلاس‌بندی و نیز برای پیش‌گویی مورد استفاده قرار بگیرد. روش رگرسیون خطی، یک تکنیک یادگیری نظارتی است که به وسیله آن تغییرات یک متغیر وابسته به وسیله ترکیب خطی از یک یا چند متغیر مستقل مدل می‌گردد. حالت کلی معادله مدل رگرسیون خطی به صورت رابطه (۱) بیان می‌گردد:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i \quad (1)$$

که در این رابطه، X_i ها متغیرهای مستقل و B_i ها ضرایب ثابت و Y متغیر وابسته می‌باشند (Asadi et al., 2014).

مدل‌سازی بر پایه شبکه‌های عصبی مصنوعی

شبکه‌های عصبی مصنوعی (ANN) ابزار ریاضی قدرتمندی هستند که با تقلید ساده از سیستم عصبی بیولوژیک ساخته شده‌اند و قدرت انعطاف و تصحیح‌پذیری بالایی در انطباق با داده‌های موجود دارند. مهم‌ترین موضوع در این مدل‌ها، انتخاب ورودی‌های مناسب به مدل برای رسیدن به خروجی‌های مد نظر می‌باشد. علاوه بر آن ساختار شبکه‌های عصبی مصنوعی و نحوه انتخاب ارتباط بین نرون‌ها و وزنی که هر نرون به خود اختصاص خواهد داد از اهمیت بالایی برخوردار است. ساختار یک شبکه عصبی متشکل از تعداد لایه‌ها، تعداد نرون‌ها در هر لایه، نحوه ارتباط لایه‌ها با هم، روش آموزش شبکه و نحوه توزیع پارامترها می‌باشد. شکل (۱) ساختار ساده یک شبکه عصبی پرسپترون سه لایه را نشان می‌دهد. هر نرون توسط لایه‌ی خروجی‌اش به نرون‌های لایه بعد متصل است ولی با نرون‌های لایه‌ی خودش ارتباط ندارد (Schaap and Leij, 1998).

در این تحقیق از شبکه عصبی پیش‌رو با تعداد لایه ورودی مختلف (بسته به سناریوی تعریف‌شده) و یک خروجی (ظرفیت زراعی یا نقطه پژمردگی دایم) استفاده شده است. برای بررسی ارتباط خطا با ورودی‌ها، وزن‌ها و خروجی‌ها، روش‌های متفاوتی از جمله روش گرادیان شیب، روش نیوتن، روش اندازه حرکت، روش آنترویی متقابل و روش لونیبرگ-مارکواد وجود دارد که در این تحقیق تابع

واقعی می‌باشند، را نادیده می‌گیرد. این تابع به صورت رابطه (۲) تعریف می‌شود.

$$L(y, f(X, \alpha)) = |y - f(X, \alpha)|_\varepsilon = \begin{cases} 0 & \text{if } |y - f(X, \alpha)| \leq \varepsilon \\ |y - f(X, \alpha)| - \varepsilon & \text{otherwise} \end{cases} \quad (2)$$

این تابع خطا، مقادیر خطای کمتر از ε را در نظر نمی‌گیرد. مسأله تقریب یک مجموعه از داده‌های زیر را در نظر می‌گیرد:

$$D = \{(x^1, y^1), \dots, (x^l, y^l)\}, \quad (x \in R^n, y \in R) \quad (3)$$

تابع رگرسیون به‌وسیله تابع زیر تخمین زده می‌شود.

$$D = \{(x^1, y^1), \dots, (x^l, y^l)\}, \quad (x \in R^n, y \in R) \quad (4)$$

که در روابط (۲) و (۵)، $\langle \cdot \rangle$ ضرب داخلی می‌باشد. تابع بهینه رگرسیون به‌وسیله حداقل تابع زیر بیان می‌گردد.

$$\Phi(\omega, C) = \frac{1}{2} \|\omega\|^2 + C \sum_i 1(C_i^- + C_i^+) \quad (5)$$

$$\text{Subject to } \begin{cases} y_i - ((\omega, x_i) + b) \leq \varepsilon + C_i^- \\ ((\omega, x_i) + b) - b \leq \varepsilon + C_i^+ \\ C_i^-, C_i^+ \geq 0 \end{cases}$$

که C یک مقدار از پیش تعیین شده بود، C_i^- و C_i^+ متغیرهای به‌دست آمده هستند که قیود بالا و پایین خروجی سیستم را تعیین می‌کنند. اگر داده‌ها به صورت خطی مجزا از هم باشند، یک سطح بهینه که داده‌ها را بدون خطا و با حداکثر فاصله میان صفحه و نزدیک‌ترین نقاط آموزشی (بردارهای پشتیبان) تفکیک می‌نماید، آموزش می‌دهد. اگر نقاط آموزشی به صورت (x_i, y_i) و بردار ورودی $x_i \in R^n$ تعریف شود، در حالتی که داده‌ها به صورت خطی قابل تفکیک باشند، معادله به صورت زیر است:

$$y = f(x) = \text{sign} \left[\sum_{i=1}^N y_i \alpha_i \langle x_i, x \rangle + b \right] \quad (6)$$

که در آن y خروجی معادله و y_i ارزش طبقه نمونه آزمایشی x_i می‌باشد. بردار $x = (x_1, x_2, \dots, x_n)$ نشان‌دهنده یک داده ورودی و بردارهای $i=1, 2, \dots, N$ بردارهای پشتیبان هستند. اگر داده‌ها به صورت خطی قابل تفکیک نباشند، می‌توان با اعمال پیش‌پردازش، نمونه‌ها را به فضای بالاتر برد. در این حالت رابطه (۶) به رابطه (۷) تغییر می‌یابد:

$$y = f(x) = \text{sign} \left[\sum_{i=1}^N y_i \alpha_i K(x, x_i) + b \right] \quad (7)$$

تابع $K(x, x_i)$ تابع کرنلی (Kernel Function) است که برای ایجاد ماشین‌هایی با انواع مختلفی از سطوح غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند. برای مدل رگرسیون بردار پشتیبان از کرنل‌های مختلفی استفاده می‌شود که عبارتند از خطی، درجه دوم، گوسی و چند جمله‌ای. معمولاً تابع کرنل گوسی شعاعی برای پیش‌بینی عملکرد بهتری دارد که در این تحقیق از این تابع استفاده شده است. معادله (۸)، تابع کرنل شعاعی را نشان می‌دهد (Shukri et al., 2013).

$$k(x, y) = \exp\left(-\frac{\|x_i - y_i\|^2}{2\sigma^2}\right) \quad (8)$$

ارزیابی روش‌های داده‌مبنا

در تحقیق حاضر، داده‌های موجود به دو دسته داده‌های آموزش و داده‌های آزمون تقسیم شدند. آموزش مدل به‌وسیله ۸۰ درصد داده‌ها، انجام شد و عملکرد روش‌ها با استفاده از ۲۰ درصد داده‌های باقی‌مانده ارزیابی گردید. برای مقایسه مقادیر اندازه‌گیری شده با مقادیر پیش‌بینی شده از شاخص‌های کمی ریشه میانگین مربعات خطا نرمال شده (Normalized Root Mean Square Error (NRMSE)) (رابطه ۹)، ضریب تعیین یا R^2 (رابطه ۱۰) و انحراف جذر میانگین مربعات خطا یا MSD (رابطه ۱۱) استفاده گردید. مقادیر ضریب تعیین هرچه به عدد یک نزدیکتر باشد، عملکرد و صحت مدل بالاتر خواهد بود. مقدار NRMSE بین صفر و یک است که می‌توان آن را به صورت درصد نیز بیان نمود و هر چه مقدار آن کمتر باشد صحت تخمین بهتر خواهد بود. به‌طور کلی مقادیر کمتر از ده درصد، صحت تخمین ایده‌آل در مدل‌سازی دارند. مقادیر بین ده تا ۲۰ و ۲۰ تا ۳۰ درصد به‌ترتیب صحت تخمین خوب و متوسط و مقادیر بالای ۳۰ درصد فاقد اعتبار لازم در مدل‌سازی را داراست (Shirani, 2011 و Jafari Gilandeh et al. 2017).

$$NRMSE = \left(\sqrt{\frac{1}{N} \sum_{i=1}^N (E_i - M_i)^2} \right) \times \left(\frac{100}{M} \right) \quad (9)$$

$$R^2 = \frac{\left(\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P}) \right)^2}{\sum_{i=1}^n (O_i - \bar{O})^2 \cdot \sum_{i=1}^n (P_i - \bar{P})^2} \quad (10)$$

$$MSD = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (11)$$

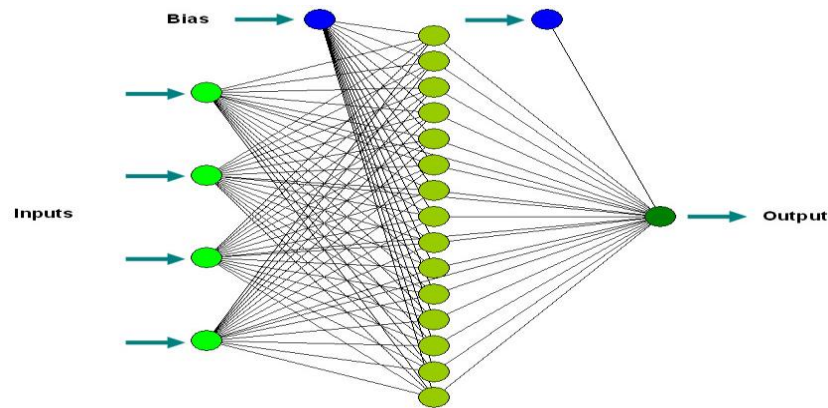


Fig. 1- Schematic of a multi-layer neural network model (input layer, middle layer and output layer)
 شکل ۱- شماتیک یک مدل شبکه عصبی چند لایه (لایه ورودی، لایه میانی و لایه خروجی)

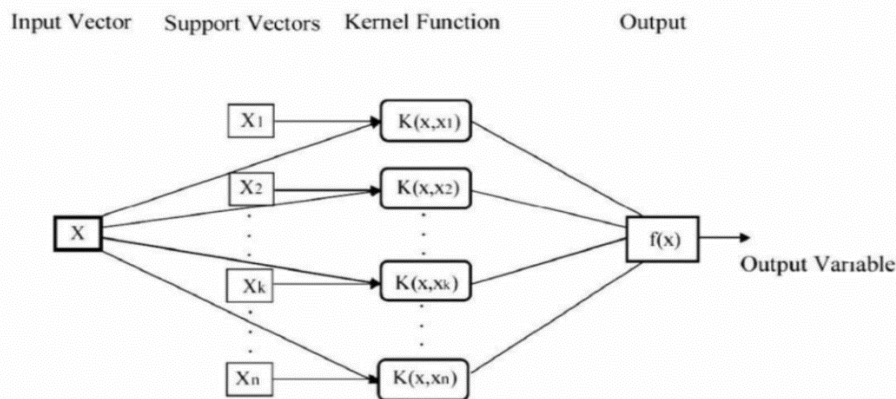


Fig. 2- Backup vector machine structure (Yoon et al., 2011)
 شکل ۲- ساختار ماشین بردار پشتیبان

در ادامه براساس تحلیل همبستگی از بین متغیرهای درصد شن (Sand)، درصد رس (Clay)، درصد سیلت (Silt)، چگالی ظاهری خاک (bd) و درصد کربن آلی خاک (OC)، پنج ترکیب متفاوت برای متغیرهای ورودی جهت تخمین نقاط مهم رطوبتی انتخاب شد که در جدول (۲) ارایه شده است.

نتایج مدل سازی رگرسیون چندمتغیره

مدل رگرسیون خطی چندمتغیره با استفاده از نرم افزار SPSS به هر کدام از مدل های ذکر شده در دو نقطه رطوبتی مورد نظر برازش داده شد. رابطه رگرسیونی به دست آمده برای هر مدل با استفاده از داده های آزمون مورد ارزیابی قرار گرفت و معیارهای ارزیابی برای آن محاسبه گردید. نتایج به دست آمده از ارزیابی روش رگرسیون خطی چندمتغیره در جدول شماره (۳) ارایه شده است.

در روابط فوق P_i و O_i به ترتیب مقادیر مشاهداتی و پیش بینی شده در نمونه i ام، n تعداد نمونه ها و \bar{P} و \bar{O} میانگین مقادیر مشاهداتی و پیش بینی شده هستند.

نتایج و بحث

تحلیل آماری داده های مورد استفاده در جدول شماره (۱) ارایه شده است. به منظور دستیابی به مدل های با صحت بیشتر و همچنین برای نرمال سازی داده ها از روش صفر و یک استفاده شد که در رابطه (۱۲) به آن اشاره شده است.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (12)$$

که در این فرمول X_{norm} مقادیر نرمال شده، X مقدار مشاهده شده، X_{min} و X_{max} به ترتیب مقدار حداکثر و حداقل مشاهده شده است.

جدول ۱- داده‌های مورد استفاده برای ظرفیت زراعی و نقطه پژمردگی دائم

N=100							Parameter
Sand	Clay	Silt	bd	OC	FC	PWP	MAX
89	34	58	1.79	2.06	33.02	21.17	Mean
18	4	7	1.39	0.13	6.54	5.89	Average
43	18	37	1.54	0.72	23.07	10.11	Standard deviation
16.93	5.83	14.34	0.08	0.44	5.54	2.64	kurtosis
-0.20	0.30	-0.85	-0.84	0.99	0.36	4.49	Skewness
0.59	0.18	-0.49	0.17	1.04	-0.47	1.78	Mode
36	16	54	1.51	0.36	20.12	8.77	

جدول ۲- سناریوهای تعریف شده

Table 2- Defined Scenarios

Scenario 1	Sand- Silt- Clay- bd- OC
Scenario 2	Sand- Silt- Clay- bd
Scenario 3	Sand- Silt- Clay
Scenario 4	Sand- Silt
Scenario 5	Silt

جدول ۳- نتایج آزمون مدل‌سازی به وسیله رگرسیون خطی چندمتغیره

Table 3- Modeling test results by multivariate linear regression

FC			PWP			Scenario
R^2	NRMSE	MSD	R^2	NRMSE	MSD	
0.50	18.67	4.53	0.45	25.63	2.73	1
0.49	19.72	4.78	0.27	28.25	3.01	2
0.43	20.57	4.99	0.18	29.55	3.15	3
0.28	22.67	5.50	0.16	32.00	3.41	4
0.11	24.89	6.03	0.20	29.52	3.14	5

$$FC=(2.68)+(9.63 \times \text{sand})+(13.85 \times \text{clay})+(16.09 \times \text{silt})+(-0.83 \times \text{bd})+(3.41 \times \text{oc}) \quad (13)$$

$$PWP=(5.48)+((0.83 \times \text{sand})+(6.05 \times \text{clay})+(1.94 \times \text{silt})+(1.49 \times \text{bd})+(-0.95 \times \text{oc}) \quad (14)$$

الگوریتم‌های آموزش مختلف در تعداد نوروهای متفاوت (بین ۲ تا ۱۰) مورد بررسی قرار گرفتند. در این تحقیق، نهایتاً از دو لایه پنهان و الگوریتم لونیبرگ-مارکواد جهت آموزش شبکه استفاده شد، همچنین بهترین معماری مدل شبکه عصبی پرسپترون چند لایه طی فرآیند سعی و خطا برای هر کدام از ترکیب‌های ورودی در هر دو نقطه‌ی رطوبتی به دست آمد. نتایج به دست آمده برای مدل شبکه عصبی در جداول (۵ و ۴) ارائه شده است.

با توجه به جداول (۵ و ۴) مشاهده می‌شود که همانند روش رگرسیونی هرچه از ورودی‌های بیشتری استفاده شود نتایج بهتری حاصل شده است، همچنین در بهترین حالت مقادیر ضریب تعیین، انحراف جذر میانگین مربعات خطا و ریشه میانگین مربعات خطا نرمال شده به ترتیب برابر با ۰/۷۲، ۳/۴۸ و ۱۴/۳۶ برای ظرفیت زراعی و ۰/۷۵، ۱/۹۰ و ۱۷/۹۱ برای نقطه پژمردگی، بیانگر ارتقا مدل‌ها و توانایی بالای روش شبکه‌های عصبی مصنوعی در

با توجه به نتایج جدول (۳) مشاهده می‌شود مدل رگرسیونی در سناریو اول بهترین عملکرد را از نظر بالا بودن ضریب تعیین و پایین بودن میزان خطا و NRMSE به ترتیب با مقادیر ۰/۵۰، ۴/۵۳ و ۱۸/۶۷ برای ظرفیت زراعی و ۰/۴۵، ۲/۷۳ و ۲۵/۶۳ برای نقطه پژمردگی داریم داراست. به طور کلی از نظر معیار NRMSE تمام مدل‌ها برای ظرفیت زراعی خوب یا متوسط و برای نقطه پژمردگی داریم در سطح متوسط و ضعیف قرار گرفته‌اند. همچنین رابطه (۱۳) و (۱۴)، برای محاسبه ظرفیت زراعی و نقطه پژمردگی دائم در حالت استفاده از تمام ورودی‌ها به دست آمده است.

نتایج شبکه عصبی

به منظور دسترسی به بهترین معماری مدل شبکه عصبی مصنوعی در پیش‌بینی نقاط مهم رطوبتی، توابع محرک و

معیارهای ارزیابی ذکر شده و نتایج به دست آمده انتخاب شد. جهت انجام عملیات یادگیری و استفاده از رگرسیون بردار پشتیبان در برآورد نقاط مهم رطوبتی از نرم افزار LibSVM استفاده شد (Nikbakht Shahbazi et al. 2013). این نرم افزار دو حالت کلی برای حل مسئله رگرسیونی دارد (epsilon- nu-SVR و SVR)، که تفاوت آنها در چگونگی بیان مسأله به صورت یک مسئله بهینه سازی و هم چنین پارامتر مورد تنظیم است. در این پژوهش یک روش برای توابع کرنل خطی، یعنی تابع شعاعی (RBF) مورد بررسی قرار گرفت.

نتایج به دست آمده برای مدل سازی نقاط مهم رطوبتی با روش رگرسیون بردار پشتیبان در جداول (۶) و (۷) ارائه شده است.

پیش بینی نقاط مهم رطوبتی است. علاوه بر آن، مدل ها برای ظرفیت زراعی از سطح متوسط به سطح خوب و برای نقطه پژمردگی دایم از سطح ضعیف یا متوسط به سطح خوب یا متوسط ارتقا پیدا کرده اند. در این تحقیق عملکرد مدل شبکه عصبی در مقایسه با روش سنتی رگرسیون آماری با توجه به معیارهای ارزیابی ذکر شده، به مراتب بهتر بوده است. این نتیجه با نتایج پژوهش های (Schaap and Leji, 1998)، (Zhang, 2007) و (Ellison and Baker, 2008) که از برتری شبکه عصبی در مقایسه با رگرسیون خطی گزارش داده بودند، مطابقت دارد.

نتایج مدل سازی رگرسیون بردار پشتیبان (SVR)

برای دست یافتن به بهترین عملکرد روش های به کار رفته، برای هر کدام از مدل ها، بهترین مدل از هر نوع با توجه به

جدول ۴- نتایج شبکه عصبی برای ظرفیت زراعی در مدل های مختلف

Table 4- Neural network results for crop capacity in different models

Neuron	Test			Train		Validation		Scenario
	R^2	NRMSE	MSD	R^2	NRMSE	R^2	NRMSE	
8	0.72	14.36	3.48	0.83	5.11	0.82	7.17	1
3	0.71	15.64	3.79	0.79	5.80	0.80	7.74	2
5	0.67	16.32	3.96	0.71	7.00	0.68	9.05	3
2	0.63	17.61	4.27	0.66	7.86	0.66	9.16	4
9	0.60	18.62	4.51	0.60	8.63	0.63	9.43	5

جدول ۵- نتایج شبکه عصبی برای نقطه پژمردگی دائم در مدل های مختلف

Table 5- Neural network results for permanent wilting point in different models

Neuron	Test			Train		Validation		Scenario
	R^2	NRMSE	MSD	R^2	NRMSE	R^2	NRMSE	
8	0.75	17.91	1.90	0.78	11.84	0.86	9.12	1
5	0.74	18.69	1.99	0.75	12.54	0.81	13.36	2
6	0.57	22.26	2.37	0.71	14.75	0.72	16.52	3
8	0.47	26.99	2.86	0.69	15.08	0.70	18.22	4
2	0.26	28.05	3.00	0.59	17.70	0.46	23.21	5

جدول ۶- نتایج SVR در برآورد ظرفیت زراعی

Table 6- SVR results in crop capacity estimation

Neuron	Train		Test			Model parameters			Scenario
	R^2	MSD	R^2	NRMSE	MSD	C	ϵ	kernel	
8	0.89	2.12	0.85	12.89	3.12	3.26	0.00	RBF	1
3	0.89	2.18	0.83	13.49	3.27	12.26	0.01	RBF	2
5	0.85	2.51	0.83	13.72	3.33	7.82	0.02	RBF	3
2	0.84	2.97	0.79	15.30	3.71	6.51	0.02	RBF	4
9	0.81	3.08	0.72	16.21	3.94	6.62	0.02	RBF	5

جدول ۷- نتایج SVR در برآورد نقطه پژمردگی دائم
Table 7- SVR results in estimation of permanent wilting point

Train		Test			Model parameters			
R^2	MSD	R^2	NRMSE	MSD	C	ϵ	kernel	Scenario
0.92	0.97	0.83	14.84	1.58	0.42	0.00	RBF	1
0.90	1.12	0.78	16.01	1.70	4	0.01	RBF	2
0.86	1.42	0.74	16.74	1.78	5.65	0.02	RBF	3
0.86	1.63	0.74	16.95	1.80	1.18	0.01	RBF	4
0.80	1.97	0.72	20.63	2.20	13.45	0.01	RBF	5

مقادیر ۰/۸۹ و ۰/۹۲ برای SVM در مقایسه با ۰/۸۳ و ۰/۷۸ برای ANN می‌تواند گویای این مطلب باشد. این نتیجه با گزارش‌های حاصله از محققانی همچون Kisi and Noori et al. (2011)، Cimen et al. (2011)، Kakaeilafdani et al. (2013) و Kaihua et al. (2014) برابری می‌کند.

مقایسه نمودارهای برازش روش‌های داده مینا

شکل (۳) تا (۵)، نمودارهای برازش داده‌های اندازه‌گیری شده با پیش‌بینی شده در مرحله آزمون، برای سه روش رگرسیونی، شبکه‌های عصبی مصنوعی و رگرسیون بردار پشتیبان در بهترین سناریوها را نشان می‌دهد.

با توجه به نمودارهای به‌دست آمده در شکل (۳) تا (۵)، مشاهده می‌شود که پراکنش نقاط اطراف خط یک به یک در مدل‌های رگرسیون بردار پشتیبان نسبت به مدل‌های شبکه‌های عصبی و رگرسیون خطی چندمتغیره فشرده‌تر می‌باشد. این امر بیانگر همبستگی بالاتر مقادیر پیش‌بینی با مشاهده‌ای است. مهمترین عامل پایین بودن همبستگی و در نتیجه ضریب تعیین برای مدل‌های حاصله از شبکه‌های عصبی مصنوعی و رگرسیون چند متغیره این است که این گونه مدل‌ها نیازمند داده‌های با تعداد بالا می‌باشند، در صورتی که مدل رگرسیون بردار پشتیبان با تعداد داده‌های کم و با کیفیت، می‌تواند آموزش خوبی داشته باشد (Ungaro et al. (2005)، Lin et al. (2009)).

با توجه به جدول (۶) مشاهده می‌شود که در بهترین حالت مقادیر ضریب تعیین، انحراف جذر میانگین مربعات خطا و ریشه میانگین مربعات خطا نرمال شده به ترتیب برابر با ۰/۸۵، ۳/۱۲ و ۱۲/۸۹ برای ظرفیت زراعی بدست آمده است. طبق تعاریف معیار ارزیابی NRMSE، سطح تمام سناریوهای حاصله از رگرسیون بردار پشتیبان برای ظرفیت زراعی خوب توصیف می‌شود.

همان‌طور که در جدول (۷) مشاهده می‌شود، در بهترین حالت مقادیر ضریب تعیین، انحراف جذر میانگین مربعات خطا و ریشه میانگین مربعات خطا نرمال شده به ترتیب برابر ۰/۸۳، ۱/۵۸ و ۱۴/۸۴ برای نقطه پژمردگی دائم، به‌دست آمده است. سطح مدل‌های یک تا چهار، خوب و سناریوی پنج، متوسط می‌باشد.

در مقایسه عملکرد مدل‌های حاصله از روش شبکه عصبی با روش رگرسیون بردار پشتیبان، مشهود است که روش رگرسیون بردار پشتیبان عملکرد بهتری داشته است. مقادیر بالاتر ضریب تعیین در سناریوهای حاصله از این روش، نشان از صحت و کارایی بالاتر این مدل در برآورد نقاط مهم رطوبتی داشته است. همچنین می‌توان بیان نمود که روش رگرسیون بردار پشتیبان در مقایسه با روش‌هایی همچون شبکه عصبی و روش رگرسیون آماری، نتایج را بهبود می‌دهد. همانگونه که Lin et al. (2009) بیان داشتند، روش SVR در مقایسه با شبکه‌های عصبی مصنوعی، نتایج بهتری در مرحله آموزش برای هر دو نقطه رطوبتی حاصل داشت که این مهم، عامل تعیین‌کننده در برآوردهای مرحله آزمون نیز می‌باشد، برای ظرفیت زراعی و نقطه پژمردگی دائم به ترتیب

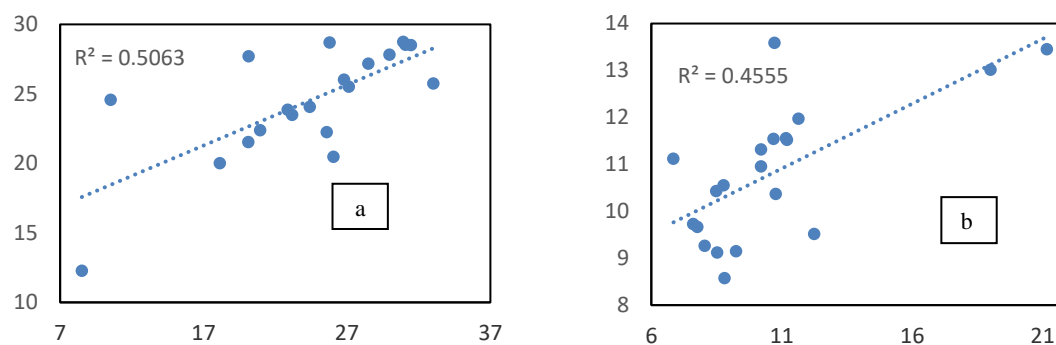


Fig. 3- Diagram fitting the data measured with predictions for crop capacity (a) and permanent wilting point (b) in a multivariate linear regression model

شکل ۳- نمودار برازش داده‌های اندازه‌گیری شده با پیش‌بینی شده برای ظرفیت زراعی (a) و نقطه پژمردگی دائم (b)، در مدل رگرسیون خطی چند متغیره

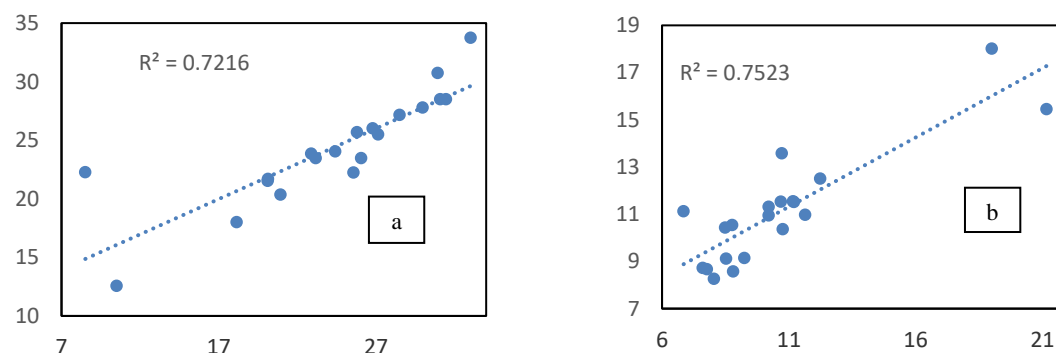


Fig. 4- Diagram fitting the predicted data for crop capacity (a) and permanent wilting point (b) in the ANN model

شکل ۴- نمودار برازش داده‌های اندازه‌گیری شده با پیش‌بینی شده برای ظرفیت زراعی (a) و نقطه پژمردگی دائم (b)، در مدل شبکه‌های عصبی مصنوعی

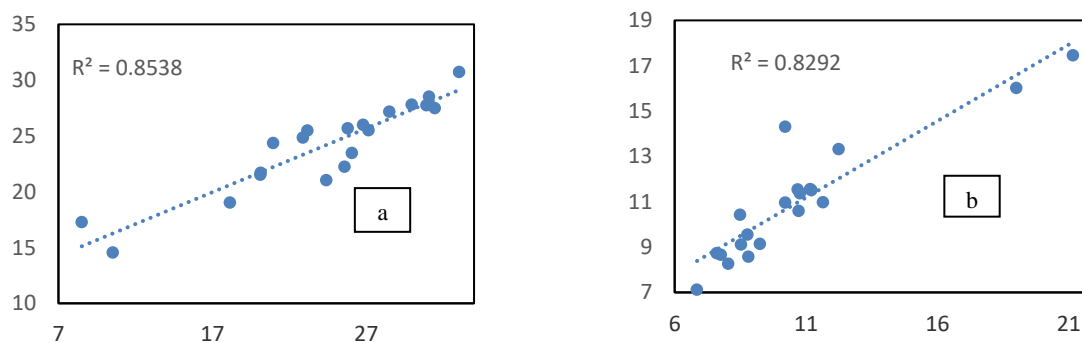


Fig. 5- Diagram of data fitted with predicted crop capacity (a) and permanent wilting point (b), in support vector regression model

شکل ۵- نمودار برازش داده‌های اندازه‌گیری شده با پیش‌بینی شده برای ظرفیت زراعی (a) و نقطه پژمردگی دائم (b)، در مدل رگرسیون بردار پشتیبان

نتیجه گیری

در این تحقیق سعی بر آن شد تا عملکرد روش‌های سنتی رگرسیون خطی و شبکه عصبی مصنوعی با روش نوین ماشین بردار پشتیبان در برآورد نقاط مهم رطوبتی، مورد ارزیابی قرار گیرد، هم چنین با استفاده از روش توابع انتقالی، از داده‌های زودیاقت به عنوان ورودی و از ظرفیت زراعی و نقطه پژمردگی دائم به عنوان داده‌های دیریاقت یا خروجی استفاده شد. نتایج نشان داد:

مدل‌های رگرسیون بردار پشتیبان با تابع کرنل خطی پایه شعاعی قادر خواهند بود که با خطای پایین و ضریب تعیین بالا نقاط مهم رطوبتی خاک را پیش‌بینی کنند، همچنین این مدل‌ها در مرحله آموزش، دارای نتایج بهتر با ضریب تعیین بالاتری بودند که می‌توانند جایگزین بسیار خوبی برای روش‌های سنتی همچون شبکه‌های عصبی و رگرسیون خطی که نیازمند داده‌های با تعداد بالا هستند، باشند. در روش‌های داده مینا به کاررفته، افزایش ورودی باعث بالا رفتن ضریب تعیین و کاهش خطا شد که استفاده از متغیرهای زودیاقت بیشتر پیشنهاد می‌گردد. البته، احتمال اینکه در روش رگرسیون بردار پشتیبان یا شبکه‌های عصبی مصنوعی MLP پدیده بیش‌آموزشی (Overfitting) در مورد داده‌های آموزشی اتفاق بیفتد، وجود دارد. این مسئله نیاز به تحقیق بیشتر و استفاده از پارامترهای ورودی دیگر دارد و در این موارد پژوهشگران پیشنهاد کرده‌اند که از سایر روش‌های بهینه‌سازی برای انتخاب متغیرهای

پیش‌بینی کننده استفاده شود و به عنوان ورودی SVM یا ANN قرار گیرد (Nikbakht Shahbazi et al., 2013).

با توجه به اینکه کارایی مدل ماشین بردار پشتیبان در کاربردهای مختلفی توسط پژوهشگران متعدد به اثبات رسیده است، استفاده از این روش برای ایجاد توابع انتقالی خاک در مناطق مختلف توصیه می‌شود، در ادامه پیشنهاد می‌گردد برای آموزش بهتر مدل‌ها از پایگاه داده وسیع‌تری و همچنین دارای همبستگی اولیه بالا استفاده شود. در تحقیق حاضر، داده‌های مورد استفاده از مرکز تحقیقات کشاورزی استان سمنان و با تعداد محدودی دریافت شد، در صورتی که در کشورهای توسعه‌یافته چنین مطالعاتی با توجه به داشتن داده‌های فراوان و موجود در پایگاه داده‌هایی مانند UNSPDA (پایگاه داده‌های خاک‌های ایالات متحده آمریکا) و HYPRES (پایگاه داده‌های خاک‌های اروپا) صورت می‌گیرد و داده‌های مربوط به خصوصیات خاک برای مناطق خشک و نیمه‌خشک مانند ایران لازم و ضروری می‌باشد.

تشکر و قدردانی

نویسندگان این مقاله از همکاری معاون پژوهشی مرکز تحقیقات و آموزش کشاورزی استان سمنان جناب آقای احمد اخیانی کمال تشکر را دارند.

References

- 1- Baker, L. and Ellison, D., 2008. Optimisation of pedotransfer functions using an artificial neural network ensemble method. *Geoderma*, 144, pp. 212-224.
- 2- Blake, G. and Hartage, K., 1986. Bulk density. In *Methods of Soil Analysis, Part Klute*. ASA Monogor Madison. *Soil Science Society of America*, pp. 363-376.
- 3- Botula, Y. D., Nemes, A., Mafuka, P., Van Ranst, E. and Cornelis, W., 2013. Prediction of water retention of soils from the humid tropics by the nonparametric k-nearest neighbor approach. *Vadose Zone Journal*, 12(2), pp. 1-17.
- 4- Chen, S., Yu, P. and Tang, H., 2010. Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology*, 385, pp. 13-23.
- 5- Gee, G. W. and Bauder, J. W., 1986. Particle-size analysis, hydrometer method. In Klute et al. (eds.) *Method's of Soil Analysis Agron. Soil Science Society of America*, pp. 404-408.
- 6- Hong W., 2011. Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm, *Neurocomputing*, 74, pp. 2096-2107.
- 7- Kaihua, L., Shaohui, X., Jichun, W., Qing, Z. and N. Lesheng., 2014. Using support vector machines to predict cation exchange capacity of different soil horizons in Qingdao City, China. *Journal of Plant Nutrition and Soil Science*, 177, pp. 775-782.
- 8- Kakaeilafdani, E., Moghaddamia, A. and Ahmadi, A., 2013. Daily suspended sediment load prediction using artificial neural networks and support vector machines. *Journal of Hydrology*, 478, pp. 50-62.
- 9- Kisi, O. and Cimen, A., 2011. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *Journal of Hydrology*, 399(2), pp. 132-140.

- 10-Koekkoek, E. J. and Bootink, H., 1999. Neural network models to predict soil water retention. *European Journal of Soil Science*, 50, pp. 489-495.
- 11-Lin, G., Chen, G., Huang, P. and Chou, Y., 2009. Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods. *J. of Hydrology*, 372, pp. 17-29.
- 12-MahdaviMeymand, A. and Ahadian, J., 2015. Comparison of Statistical, Experimental, neural network and fuzzy neural network methods in estimation of air Overflow needed. *Journal of Irrigation Science and Engineering*, 38 (3), pp. 51-61. (In Persian).
- 13-Minasny, B. and McBratney, A., 2002. The Neuro-m method for fitting neural network parametric pedotransfer functions. *Soil Science of Society America*, 66, pp. 352-361.
- 14-Nemes, A., Schaap, M. and Wosten, J., 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection, *Journal of Soil Science*, 67, pp. 1093-1102.
- 15-Nguyen, P.M., De Pue, J., Van, K. L. and Cornelis, W., 2015. Impact of regression methods on improved effects of soil structure on soil water retention estimates, *Journal of Hydrology*, 29, pp. 598-606.
- 16-Noori, R., Karbassia, A., Moghaddamniac, D., Hand, M.H., Zokaei-Ashtianie, A., Farokhniab, F. and GhafariGoushehc, M., 2013. Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. *Journal of Hydrology*, 401 (3), pp. 177-189.
- 17-NikbakhtShahbazi, A., Zahraie, B. and Naseri, M., 2013. Seasonal Meteorological Drought Forecasting Using Support Vector Machines. *Journal of Water and Wastewater*, 2, pp. 73-85. (In Persian).
- 18-Shirani, H., 2011. Estimation of some soil moisture characteristic curve points including FC and PWP using soil transfer functions and regression method in Kerman. *Journal of Agricultural Science and Technology Soil and Water Sciences*, 59 (16), pp. 141-150. (In Persian).
- 19-Schaap, M. G. and F. Leij., 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil and Tillage Research*, 47, pp. 37-42.
- 20-Shukri, Q., Sadeghi, M. And Ahmadi Marwash, M., 2013. Presentation of a Combined Data Preprocessing Method in Regression Vector Machine to Predict the Quality of Refined Oil. *Journal of Petroleum Research*, 75 (23), pp. 102-116. (In Persian).
- 21-Ungaro, F., Calzolari, C. and Busoni, E., 2005. Development of pedotransfer functions using a group method of data handling for the soil of the Pianura Padano-Veneta region of North Italy. *Water Retention Properties Geoderma*, 124, pp. 293-317.
- 22-Vapnik, V. N. and Cortes, C., 1995. Support vector networks. *Machine Learning*, 20, pp. 273-297.
- 23-Vali, A., Moiri, M. and Movahediniya, N., 2009. Comparative analysis of artificial neural networks performance and suspended sediment prediction regression models. *Natural Geography Research*, 71, pp. 21-30. (In Persian).
- 24-Yin, J. and Log, P., 2011. Prediction for blocked tripe tides with amino acids descriptors (HMLP) by multiple linear regression and support vector regression”, *Procedia Environmental Sciences*, 8, pp. 173-178.
- 25-Yoon, H., Jun, S. C., Hyun, Y., Bae, G. O. and Lee, K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *Journal of Hydrology*, 396 (1), pp. 128-138.

-
- 26-JafariGilandeh, S., Khodaverdillo, H. and Rasolzadeh, A., 2017. Application and comparison of parametric transfer functions of Van Genuchten model in simulating unsteady water flow in cultivated soil. *Soil Knowledge Journal*, 25 (2), pp. 82 - 92. (In Persian).
- 27-Zhang, Y., 2007. Artificial neural networks based on principal component analysis input selection for clinical pattern recognition analysis. *Talanta*, 73 (1), pp. 68-75.