Predicting of Soil Physicochemical Properties and Productivity for Sandy Soil by Using Deep Learning Under Salt Stress Conditions

Diia S.M. Boulos [1]

ABSTRACT

Deep learning is an exciting discipline that has already transformed the way data are analyzed in many fields. This study developed and evaluated artificial neural network (ANN), a type of deep learning algorithm, as a new way to predict the physicochemical properties of sandy soil incorporated with three rates (10, 15 and 20 ton/fed) of farmyard manure (FYM) and compost (COM)] for each treatment with three salinity levels of irrigation water 1000, 1500 and 2000 ppm. These properties were soil bulk density (Bd), available water (AW), cation exchange capacity (CEC), sodium adsorption ratio (SAR), spinach productivity (Pro). Multilayer feedforward ANN with 6 neurons in input layer and 5 neurons in output layer was trained using a back propagation learning algorithm. The ANN model was trained with data collected from previous literatures 555 observations (447 observations for training and 108 observations for testing). The model inputs were [sand, silt, clay, FYM, COM, Ec of irrigation water (EC_{ir})]. Verification of the ANN model in prediction was done using field experimental data which carried out in Ismailia governorate (Data that an ANN model has never seen before). In order to evaluate the ANN model, root mean square error (RMSE) and correlation coefficient (R^2) were calculated. After careful and extensive training, validation and testing for the ANN model were conducted. The RMSE between measured and predicted values for both Bd, AW, CEC, SAR and Pro were 0.00372 $Mg.m^{-3}$, 0.166 %, 0.09903 $Cmol.kg^{-1}$, 0.05975 and 12.63481 kg/fed. The R^2 values were equal to 0.99835, 0.9977, 0.99765, 0.99929 and 0.99916, respectively. The high correlation coefficient for parameters outputs recall indicate for excellent prediction of ANN model for the data has never seen before.

Key Words: Deep Learning, Artificial Neural Network, Soil Physicochemical Properties and Multilayer Feedforward.

INTRODUCTION

Artificial neural network (ANN) is nonlinear computer algorithms that can model the behavior of complicated nonlinear processes. They can learn from history and experience and thereby enhance their efficiency. The ANN method is a robust, powerful, and suitable technique for nonlinear and complex processes (Rajabi-Vandechali et al., 2018). ANNs have various advantages, such as the ability to handle a large amount of data, good globalization of results, the ability to implicitly separate nonlinear independent and nonindependent relationships, and the ability to determine the relative importance of different input parameters.

ANN has the capability of correlating large and complex data sets without any prior knowledge of the relationship between them. It has become powerful tools for modeling a system that had incomplete or a little understanding regarding its governing law (Aru and Okpara, 2018). The strengths of ANN are that it possesses the ability to learn through the means of a set of training data, capability of generalization and association of data as well the fault tolerance in the sense of handling noise and incomplete information. Also, ANN consists of the feature of parallelism which enables computations of multiple neurons simultaneously. ANN is often designed using multilayer feedforward (MLF) back propagation algorithm. MLF network, or commonly known as multilayer perceptron, is one of the most popular neural networks used in the present. In general, MLF network contains an input layer, single or multiple hidden layers and an output layer. To define a MLF network, it is a network whereby the neuron in one layer is connected to the neuron of the subsequent layer, towards the direction of output layer. Typically, the layers are entirely connected in the sense of all neurons. Each layer is connected with all neurons at next layer as shown in Fig. (1) (Abdullah and Tiong, 2008).

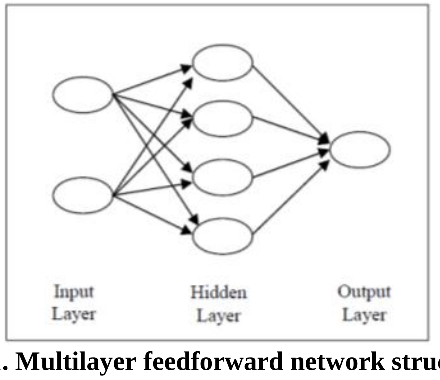

Fig. 1. Multilayer feedforward network structure

Sorour, (2006) found that an ANN similar estimates dry matter loses and storage time of stored wheat under different treatments well compared to measured values. (El Awady et al., 2003). Developed an ANN to study the relative of variables affecting the performance of

DOI: 10.21608/esm.2024.442234
[1]Soil Chemistry and Physics Department, Desert Research Center, Cairo, Egypt.
Received November 01, 2024, Accepted, December 01, 2024.

chisel plows, Aboukarima et al., 2004 found that correlation coefficients (R^2) were over 0.90 during testing process and the width of plow was the major variable affecting draft and predicted unit draft of tillage machine using statistical and ANN models. The ANN was a MLF network with 11 input and 1 output neurons. The input variables were chisel plow, moldboard plow, disc plow, soil texture, plowing depth, plow width, forward speed, moisture content, soil bulk density, rated tractor power, and plow passes. The output predicted the unit draft of tillage implement. The standard deviations of the errors were 9.38, 6.57 and 8.45 kN/m^2 for moldboard, chisel and disc plows respectively. Also the R^2 between the measured and predicted values were 0.95 for both the ANN and statistical analyses. (Aboukarima, 2007) obtained data for plows in different soil characteristic, width of plow and some operational parameters with the help of ANN model. The variables were depth of plowing, power tractor, forward speed, width of plow, soil texture and water content. The R^2 were 93%. (Akbarzadehe et al., 2009) used alternative methods of ANN for predicting water runoff and particles splash in soils treated with synthetic geotextiles and bare soils. It was found that the ANN had better accuracy than regression analyses for prediction of runoff and splash. (Gholami et al., 2018) found that the ANN can predict soil erosion with an acceptable level (RMSE =0.04, R^2 =0.94). (Warmling et al. 2019) developed an ANN model to predict field capacity (FC), wilting point (WP) and available water (AW) the results between the measured and predicted values had RMSE of 0.01, 0.03 and 0.03 m^3/m^3 and R^2 were 0.99, 0.92 and 0.83 for FC, WP and AW, respectively. The objective of this work was to explore the use of deep learning, specifically artificial neural network (ANN), to predict physicochemical properties of sandy soil treated with organic fertilizers and irrigated with brackish water.

This study aims to achieve the following objectives:

1. construct the optimal structure of an ANN to predict some physicochemical properties [soil bulk density (Bd), available water (AW), CEC, SAR, soil spinach productivity (Pro)] under two different kinds of organic matter [farmyard manure (FYM) and compost (COM)] for three rates [10, 15 and 20 ton/fed] for each treatment with three salinity levels of irrigation water 1000, 1500 and 2000 ppm.
2. Verification of the ANN model in prediction using field experimental data (Data that an ANN model has never seen before) which carried out in sand soil from Ismailia governorate.

MATERIALS AND METHODS

Field experimental

A field experiment was conducted on sandy soil at Ismailia governorate in a private farm. Seeds of spinach (Spinacia oleracea L.) were sown on 21st October in winter season 2023 under drip irrigation system. The recommended rates of NPK mineral fertilizers were as follows: phosphorus as 200 kg/fed of calcium superphosphate (15.5% P_2O_5) which added during soil preparation, potassium as 70 kg/fed of potassium sulphate (48-50% K_2O) were added three weeks after seeding. 250 kg/fed of ammonium sulphate (20.5%N) as a source of N applied in three splits (50 kg during soil preparation, 100 kg three weeks after sowing and 100kg after five weeks from sowing).

Three levels of FYM and compost (10, 15 and 20 ton/fed) for each treatment were mixed in 15 cm soil depth during soil preparation. Area of each plot was 4 m^2. Data of soil analyses according to (Klute, 1986) were tabulated in Table (1). The three salinity levels of irrigation water 1000, 1500 and 2000 ppm were used

Table 1. Some physical and chemical properties of surface soil sample

Soil sample (cm)	Course sand %	Fine sand %	Silt %	Clay %	Texture	Bulk density (Mg.m^{-3})	pH	EC (dS/m)
0-15	13.26	77.71	5.35	3.68	Sandy	1.69	7.65	2.58

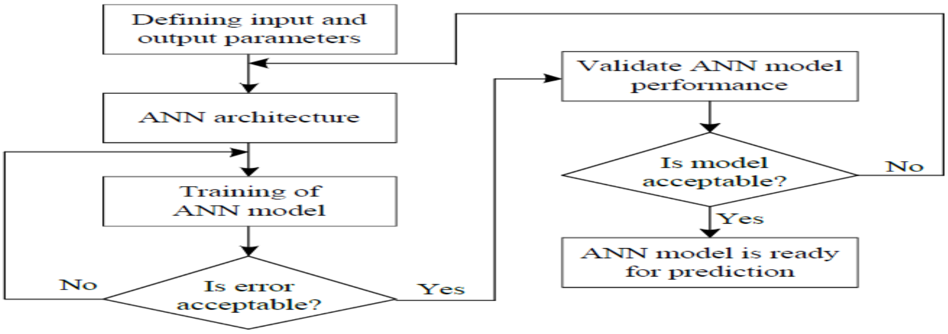

Fig. 2. Neural network modelling procedure flowchart

The ANN model

The steps of construct ANN model were: the first step was conceptualized the inputs and outputs to be used. Second, gathering data to be used for training (learning) the model. Third, create the ANN model. Fourth, test the model with some cases. Finally, validate the model or examine how the ANN model performs with the test data. The aim of the learning procedure is to determine the optimal set of weights and biases that produce the correct output for any input. The output of the network is compared with the target response to produce an error. Once the ANN is properly trained, it can be generalized to similar situations that are unprecedented. ANNs usually consist of three layers (input layer, hidden layers, output layer) (Noor et al., 2016). A flowchart of the ANN modeling procedure is shown in Fig. (2).

The ANN model backpropagation with two hidden layers were used in this study. This type of ANN is a nonlinear data transformation structure consisting of input and output nodes connected to hidden nodes by adaptable coefficients. Hidden nodes depend on the complexity of the underlying problem and were determined empirically by calibrating ANN with different numbers of hidden nodes. Both the hidden and output nodes contain transfer function of sigmoid that provides the ANN with nonlinear capabilities. The accuracy of the network was evaluated by the RMSE and R^2 (Abdullah and Tiong, 2008). R^2 is represents the actual data sets, it can vary from 0 to 1. An R^2 value close to 1 indicates that the ANN model perfectly predicts the output.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}$$

Where:
n = number of data points during testing process
x_i = value from measured
y_i = value from predicted

RESULTS AND DISCUSSION

The ANN model was trained with data collected from previous literatures 555 observations. The collected data were separated into two groups 447 of observations were set for training and 108 for testing. Statistical measures for the entire collected dataset are presented in Table (2). The inputs were (sand, silt, clay, farmyard manure (FYM), compost (COM), EC of irrigation water (EC_{ir})) for prediction of (soil bulk density (Bd), available water (AW), CEC, SAR and spinach productivity (Pro)). The data used in this study as shown in Table (2) has a wide range of soil particle size distribution. Sand content ranging from 81.98 to 64.35%, silt content ranging from 3.65 to 9.33% and clay content ranging from 1.97 to 8.69%. The addition different types of organic matter (FYM, COM) with different three rates ranging from 0 to 20 ton/fed for both. Also, EC of irrigation water (EC_{ir}) ranging from 0 to 2000 ppm.

Several ANN models were trained with various design parameters including number of hidden layers and number of nodes in each hidden layer. The selection of the optimum model was based on minimizing the difference between the ANN predicted and measured values outputs Fig (3).

The best model consisted of hidden layers with 7 and 14 nodes in the first and second hidden layer. The architecture of the developed ANN model is depicted in Fig. (4).

The RMSE decreased with increasing of learning iterations for 5 outputs. The training network gave achieved the best results at 50,000 training runs with RMSE at 0.0115 and R^2 at 0.9975 Fig. (3). Also, the ANN model was tested with testing data set (108 observations) where RMSE and R^2 equal to 0.0219 and 0.9915 respectively. According to these results, outcomes were acceptable during the training and testing stages.

Table 2. Statistical measures for the entire collected dataset

Parameter	Minimum	Maximum	Mean	Standard deviation
Sand (%)	81.98	94.35	89.34	3.52
Silt (%)	3.65	9.33	5.85	1.73
Clay (%)	1.97	8.69	4.80	1.89
Farmyard manure (ton/fed)	0	20	7.30	8.03
Compost (ton/fed)	0	20	7.30	8.03
EC of irrigation water (ppm)	0	2000	1459.46	47.88

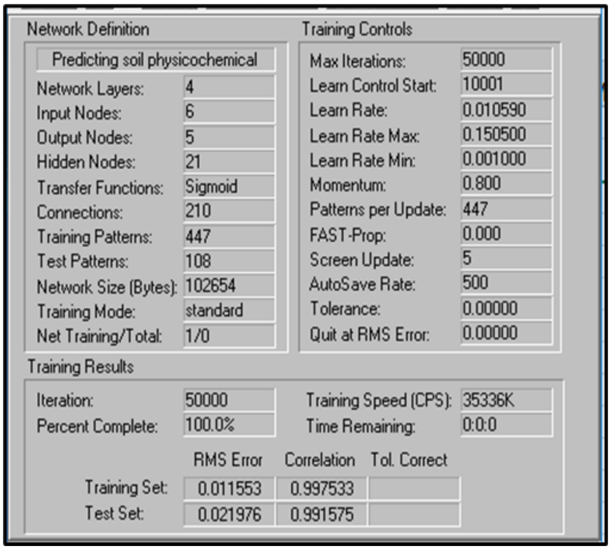

Fig. 3. Network definition for ANN model

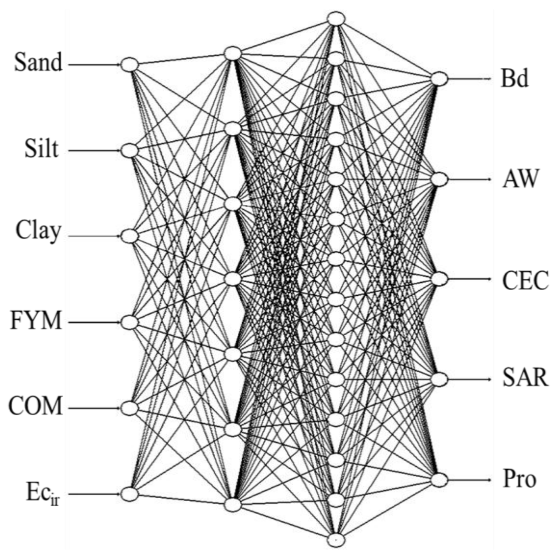

Fig. 4. The architecture of the developed ANN model

After network training and optimization, we carried out the verification stage for the optimized network. This was conducted through the comparison between the measured values (Data that an ANN model has never seen before) from field experimental and the predicted values from ANN model, and the results are shown in Figs. (5-9).

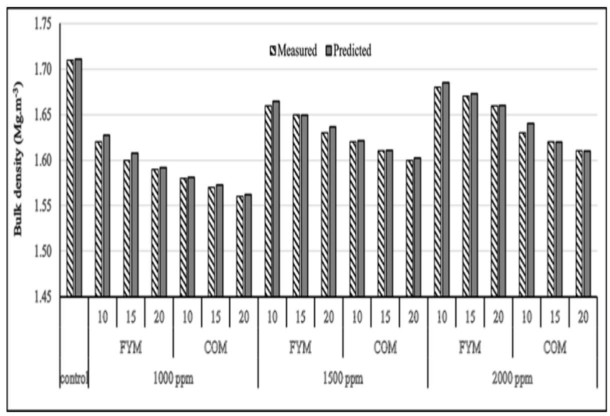

Fig. 5. Measured and predicted ANN model for bulk density under different treatments

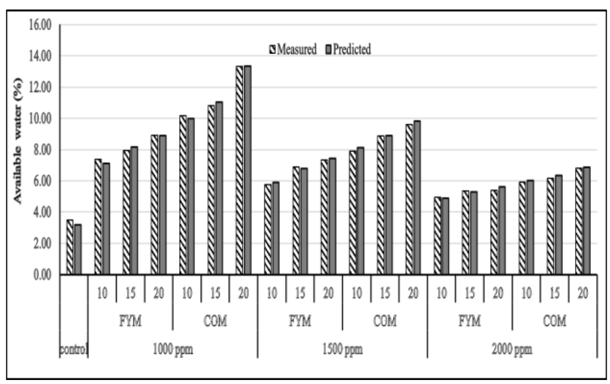

Fig. 6. Measured and predicted ANN model for available water under different treatments

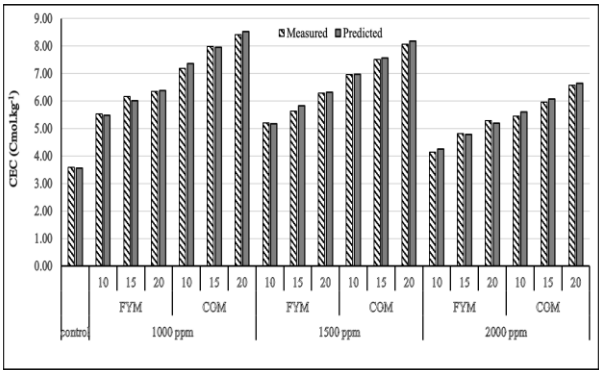

Fig. 7. Measured and predicted ANN model for CEC under different treatments

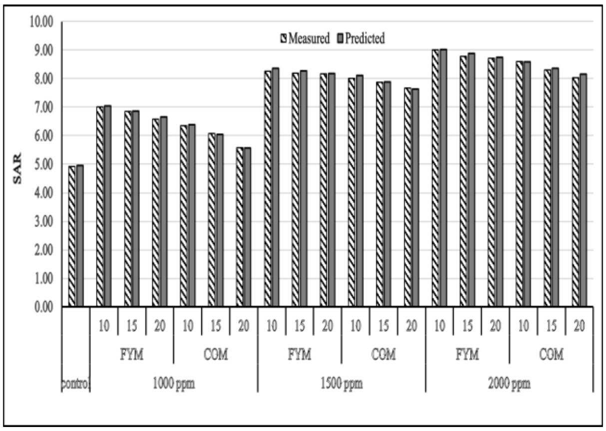

Fig. 8. Measured and predicted ANN model for SAR under different treatments

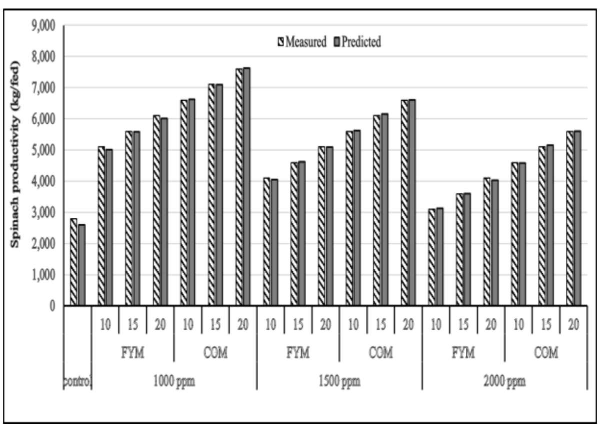

Fig. 9. Measured and predicted ANN model for spinach productivity under different treatments

Table (3) showed the values of the RMSE between measured and predicted for Bd, AW, CEC, SAR and Pro were 0.00372 $Mg.m^{-3}$, 0.166 %, 0.09903 $Cmol.kg^{-1}$, 0.05975 and 12.63481 kg/fed. While the R^2 were equal to 0.99835, 0.9977, 0.99765, 0.99929 and 0.99916, respectively. The high correlation coefficient for parameters outputs recall indicate for excellent prediction of ANN model for data has never seen before. These findings suggest that ANN model are a promising way of enhancing the prediction skill for soil physicochemical properties and productivity agreement with (Slater et al., 2023).

Table 3. The ANN model recall statistical

parameter	RMSE	Maximum Error	Correlation Coefficient (R^2)
Bulk density ($Mg.m^{-3}$)	0.00372	0.00738	0.99835
Available water (%)	0.166	0.30196	0.9977
CEC ($Cmol.kg^{-1}$)	0.09903	0.20363	0.99765
SAR	0.05975	0.11492	0.99929
Spinach productivity (kg/fed)	12.63481	20.84106	0.99916

CONCLUSIONS

In this study, developed and evaluated artificial neural network (ANN), a type of deep learning algorithm, as a new way to predict soil physicochemical properties (soil bulk density (Bd), available water (AW), cation exchange capacity (CEC), sodium adsorption ratio (SAR), spinach productivity (Pro)) under two different kinds of organic matter [farmyard manure (FYM) and compost (COM)] with three rates [10, 15 and 20 ton/fed] for each treatment with three salinity levels of irrigation water 1000, 1500 and 2000 ppm. The ANN model inputs were [sand, silt, clay, FYM, COM, Ec of irrigation water (EC_{ir})]. The architecture of optimal ANN model consisted of two hidden layers with 7 and 14 nodes in the first and the second hidden layers respectively. After network training and optimization, we carried out the verification stage for the optimized network. This was conducted through the comparison between the measured values (Data that an ANN model has never seen before) from field experimental and the predicted values from ANN model. The RMSE between measured and predicted for soil bulk density, available water, CEC, SAR and spinach productivity were 0.00372 $Mg.m^{-3}$, 0.166 %, 0.09903 $Cmol.kg^{-1}$, 0.05975 and 12.63481 kg/fed. While the R^2 were equal to 0.99835, 0.9977, 0.99765, 0.99929 and 0.99916 respectively. The high correlation coefficient for parameters outputs recall indicate for excellent prediction of ANN model for data has never seen before. These findings suggest that ANN model are a promising way of enhancing the prediction skill for soil physicochemical properties and productivity.

REFERENCES

Abdullah, S. and Tiong, E. C. (2008) Prediction of palm oil properties using artificial neural network. Inter. J. of Com. Sci. and Net. Sec., 8 (8): 101-106.

Aboukarima, A.M. (2007) Draft models of chisel plow based on simulation using artificial neural networks. Misr J. Ag. Eng., 24(1): 42-61.

Aboukarima, A.M., El Awady, M.N., El-Kabany, A.G. and Kabeel, M.H. (2004) Predicting unit draft of tillage implements using statistical models and neural networks. The 12th Conf. of Misr Soc. of Ag. Eng., 21(4): 239-249.

Akbarzadeh, A., Mehrjardi, R.T., Rouhipour, H., Gorji, M. and Refahi, H.G. (2009) Estimating of soil erosion covered with rolled erosion control systems using rainfall simulator (Neuro-fuzzy and Artificial Neural Network Approaches). J. of Ap. Sci. Res., 5(5): 505-514.

Aru, O. E. and Okpara, M. (2018) Review of artificial neural network: an emerging problem solving tool in computing. American Journal of Engineering Research, 7 (5): 152-161.

El Awady, M.N., El-Kabany, A.G., Kabeel, M.H. and Aboukarima, A.M. (2003) Plows performance under Egyptian conditions depicted by artificial neural networks. The 11th Conf. of Misr Soc. of Ag. Eng., 20(4): 919-936.

Gholami, V., Booij, M.J., Tehrani, E.N. and Hadian, M.A. (2018) Spatial soil erosion estimation using an artificial neural network (ANN) and field plot data. CATENA (163): 210–218. (Gholami et al., 2018)

Klute, A. (1986) Methods of soil analysis. Part 1. Physical and microbiological methods. 2nd edition. Agronomy. Monogr. 9. ASA, Madison, WI. USA, pp: 228-235.

Noor, C.W.M., Mamat, R., Najafi, G., Yasin, M.H.M., Ihsan, C.K. and Noor, M.M. (2016) Prediction of marine diesel engine performance by using artificial neural network model. J. of Mechanical Eng. and Sci., 10 (1):1917-1930.

Rajabi-Vandechali, M., M.H. Abbaspour-Fard and Rohani A. (2018) Development of a prediction model for estimating tractor engine torque based on soft computing and low cost sensors. Measurement (121): 83-95.

Slater, L. J., and Coauthors, 2023: Hybrid forecasting: Blending climate predictions with AI models. Hydrology and Earth System Sciences (27): 1865−1889

Sorour, H.M. (2006) Prediction of wheat deterioration during storage period using artificial neural network. Misr J. Ag. Eng., 23(3): 637-657.

Tofah, R. (2015) Effect of organic manure kinds on growth، yield and quality of Spinach plant. Tishreen University Journal for Research and Scientific Studies-Biological Sciences 37 (2): 279-291.

Warmling, M., Mendes, R., Albuquerque, J. and Bortolini, D. (2019) Artificial neural networks to predict soil water retention of Santa Catarina - Brazil. In book: Terra - Mudanças Climáticas e Biodiversidade, Publisher: Barlavento, pp: 393-403.

الملخص العربي

التنبؤ بالخواص الطبيعية الكيميائية للتربة والانتاجية لارض رملية باستخدام التعلم العميق تحت ظروف الاجهاد الملحى

ضياء سعيد منير بولس

يُعد التعلم العميق تخصصًا مثيرًا للاهتمام، وقد أحدث نقلة نوعية في طريقة تحليل البيانات في العديد من المجالات. وقد طورت هذه الدراسة وقيّمت الشبكة العصبية الاصطناعية (ANN)، وهي نوع من خوارزميات التعلم العميق، كطريقة جديدة للتنبؤ بالخصائص الفيزيائية والكيميائية للتربة [الكثافة الظاهرية للتربة (Bd)، والماء الميسر (AW)، وسعة التبادلية الكاتيونية (CEC)، ونسبة امتصاص الصوديوم (SAR)، وإنتاجية السبانخ (Pro)] باستخدام نوعين مختلفين من المواد العضوية [سماد المزرعة (FYM) والكمبوست (COM)] لثلاثة معدلات اضافة [١٠، ١٥، ٢٠ طن/فدان] لكل معاملة، مع ثلاثة مستويات ملوحة لمياه الري (EC_{ir}): ١٠٠٠، ١٥٠٠، ٢٠٠٠ جزء في المليون. استخدمت الشبكة العصبية الاصطناعية متعددة الطبقات ذات التغذية الامامية حيث تتكون من ٦ خلايا عصبية فى طبقة المدخلات، ٧ خلايا عصبية فى الطبقة المختفية الاولى، ١٤ خلية عصبية فى الطبقة المختفية الثانية و ٥ خلايا عصبية فى طبقة المخرجات وتم تدريبها بطريقة خوارزمية التعلم بالانتشار الخلفي. تم تعليم وتدريب واختبار الشبكة العصبية الاصطناعية ببيانات تم جمعها من الدرسات السابقة ٥٥٥ مشاهدة (٤٤٧ مشاهدة للتدريب و١٠٨ مشاهدة للاختبار). كانت المدخلات للشبكة العصبية الاصطناعية هى (نسبة الرمل، نسبة السلت، نسبة الطين، (FYM)، (COM)، (EC_{ir})). وكانت مخرجات الشبكة العصبية الاصطناعية هى (الكثافة الظاهرية، الماء الميسر، CEC، SAR، انتاجية السبانخ). تم التحقق من كفاءة نموذج الشبكة العصبية الاصطناعية فى التنبؤ باستخدام بيانات التجربة الحقلية التى اجريت فى محافظة الاسماعيلية (بيانات لم يسبق لنموذج الشبكة العصبية الاصطناعية رؤيتها من قبل). بعد التدريب الدقيق والمكثف والتحقق من صحة واختبار نموذج الشبكة العصبية الاصطناعية. من أجل تقييم نموذج الشبكة العصبية الاصطناعية، تم استخدام جذر متوسط مربع الخطأ (RMSE) ومعامل الارتباط (R^2). وكانت قيم جذر متوسط مربع الخطأ بين القيم المقاسة من التجربة الحقلية والقيم المتنبأ بها من الشبكة العصبية الاصطناعية لكلا من الكثافة الظاهرية، الماء الميسر، CEC، SAR، انتاجية السبانخ كانت ٠,٠٠٣٧٢ ميجاجرام/م٣، ٠,١٠٥٢٨%، ٠,٠٩٩٠٣ سنتيمول/كجم، ٠,٠٥٩٧٥ ، ١٢,٦٣٤٨١ كجم/فدان. بينما كان معامل الارتباط يساوى ٠,٩٩٨٣٥، ٠,٩٩٧٧، ٠,٩٩٧٦٥، ٠,٩٩٩٢٩، ٠,٩٩٩١٦ على الترتيب. يشير معامل الارتباط المرتفع للمخرجات الى التنبؤ الممتاز للشبكة العصبية الاصطناعية للقيم التى لم يسبق للشبكة التعرض لها من قبل.