


RESEARCH

Open Access



# Predictive modeling of Persian walnut (*Juglans regia* L.) in vitro proliferation media using machine learning approaches: a comparative study of ANN, KNN and GEP models

Mohammad Sadat-Hosseini<sup>1\*</sup> , Mohammad M. Arab<sup>2</sup>, Mohammad Soltani<sup>3</sup>, Maliheh Eftekhari<sup>4</sup>, Amanollah Soleimani<sup>5</sup> and Kourosh Vahdati<sup>2</sup>

## Abstract

**Background:** Optimizing plant tissue culture media is a complicated process, which is easily influenced by genotype, mineral nutrients, plant growth regulators (PGRs), vitamins and other factors, leading to undesirable and inefficient medium composition. Facing incidence of different physiological disorders such as callusing, shoot tip necrosis (STN) and vitrification (Vit) in walnut proliferation, it is necessary to develop prediction models for identifying the impact of different factors involving in this process. In the present study, three machine learning (ML) approaches including multi-layer perceptron neural network (MLPNN), *k*-nearest neighbors (KNN) and gene expression programming (GEP) were implemented and compared to multiple linear regression (MLR) to develop models for prediction of in vitro proliferation of Persian walnut (*Juglans regia* L.). The accuracy of developed models was evaluated using coefficient of determination ( $R^2$ ), root mean square error (RMSE) and mean absolute error (MAE). With the aim of optimizing the selected prediction models, multi-objective evolutionary optimization algorithm using particle swarm optimization (PSO) technique was applied.

**Results:** Our results indicated that all three ML techniques had higher accuracy of prediction than MLR, for example, calculated  $R^2$  of MLPNN, KNN and GEP vs. MLR was 0.695, 0.672 and 0.802 vs. 0.412 in Chandler and 0.358, 0.377 and 0.428 vs. 0.178 in Rayen, respectively. The GEP models were further selected to be optimized using PSO. The comparison of modeling procedures provides a new insight into in vitro culture medium composition prediction models. Based on the results, hybrid GEP-PSO technique displays good performance for modeling walnut tissue culture media, while MLPNN and KNN have also shown strong estimation capability.

**Conclusion:** Here, besides MLPNN and GEP, KNN also is introduced, for the first time, as a simple technique with high accuracy to be used for developing prediction models in optimizing plant tissue culture media composition studies. Therefore, selection of the modeling technique to study depends on the researcher's desire regarding the simplicity of the procedure, obtaining clear results as entire formula and/or less time to analyze.

**Keywords:** Walnut in vitro propagation, Artificial neural network, Gene expression programming, *k*-nearest neighbors, Prediction model

\*Correspondence: m.hosseini@ujiroft.ac.ir

<sup>1</sup> Department of Horticulture, Faculty of Agriculture, University of Jiroft, Jiroft, Iran

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

The walnut is one of the most important nuts in the world. Persian walnuts (*Juglans regia* L.) are the only edible species of walnut which are widely grown for their nuts and timbers [1]. In general, walnut tree propagation is still mainly by using seeds rather than vegetative procedures which results in non-uniform nut quality and irregular yielding [2]. Therefore, in vitro propagation is used to overcome the mentioned problems. But walnuts are considered recalcitrant to in vitro culture which makes difficult the mass propagation of different genotypes while several micropropagation protocols have been published for different genotypes [3–11]. It has been proven that walnut micropropagation results are highly dependent upon genotype [7, 9–11]. In addition to genotype, the formulation of culture medium has a great impact on all micropropagation stages. Up to now, the [3] walnut (DKW) culture medium has been the most employed formulation for walnut tissue culture. Nevertheless, there are some researches reporting improved results using modified DKW or other formulations [6–8, 12–15].

However, to the best of our knowledge, no comprehensive study has been done on the balance of culture media components (mineral nutrients, plant growth regulators (PGRs) and vitamins) and their interaction together and with genotype on walnut in vitro performance to increase the efficiency of the micropropagation process by enhancing proliferation rate and reducing physiological disorders.

Predicting the interaction of mineral nutrients, PGRs, vitamins and genotype on the explant in vitro performance would involve modeling a very complex database, which is very problematic and time-consuming process using classic statistical analyses and needs accurate and advanced modeling procedures [16, 17]. Machine learning (ML) tools allow researchers to perceive the studied process and make proper decisions to develop optimal culture media [17]. In recent years, different ML models like neural networks [18–23] have been successfully applied for prediction and optimization of different plant tissue culture processes. In our previous studies, we described the ML hybrid techniques, combining artificial neural network (ANN) with genetic algorithm (ANN-GA) in *Pyrus* [24] and *Prunus* rootstocks [25–27], rootstocks gene expression programming (GEP) with GA (GEP-GA) [20] and particle swarm optimization (PSO) (GEP-PSO) [28] in *Pyrus* rootstocks as powerful data mining approaches, which allow modeling of complicated databases and finding the factors influencing a given response in micropropagation process.

ANNs are inspired by the functions of human brain [29]. The ANN [multi-layer perceptron neural network (MLPNN) and radial basis function neural network

(RBFNN)] has revealed significant development in complex plant tissue culture systems [20, 24–27]. ANN does not require any previous knowledge regarding the creation or interrelationships between signals of input and output that is one of its profits [16]. Other benefits of ANN are prediction of the plant biomass [30], clustering the micropropagated plantlets and influencing growth and quality of the regenerated plants by controlling light, ventilation, CO<sub>2</sub> and air temperature inside the culture containers which could be of ANN benefits [16].

GEP model is another ML-based optimization technique presented by [31] which comprises useful traits of both genetic programming (GP) and GA. This new model according to an evolving computer programs algorithm was used in our previous studies on *Pyrus* rootstocks micropropagation which precisely detected nonlinear and complicated relationships between input and output [20, 24].

Here, ANN and GEP are compared to *k*-nearest neighbors (KNN) method as one of the simplest machine learning techniques. The KNN technique recognizes the elements amongst the training samples that correspond “current” conditions maximum closely based on some predefined attributes: the neighbors. The prediction value is then specified from the groups of the next values of the neighbors [32]. Comparing to mathematical modeling, the KNN method involves no model development or confirmation and thus can be used without recombining data, contrasting in the case of common data-based models [33]. In spite of the potential advantages, no research has yet been done on the use of this technique in the area of plant micropropagation.

In our previous study [20], we compared the RBFNN and GEP in optimizing the in vitro culture media composition for pear rootstocks. Based on our results GEP was a significantly powerful and more precise technique than RBFNN in prediction of in vitro proliferation quantity and quality. So, GA technique was applied to optimize GEP models [20]. Nevertheless, GA optimized the level of inputs required for each specific output, distinctly. Consequently, in our recent study [28], in order to achieve a complete optimum formulation for culture medium, we compared two algorithms GEP and M5' model tree, to predict the impacts of media minerals and PGRs on in vitro proliferation of pear rootstocks. We found that GEP showed a higher prediction precision than M5' model tree. So, we optimized the GEP prediction models using multi-objective evolutionary optimization algorithms (MOEAs) including GA and PSO methods and compared to the mono-objective GA optimization procedure. The PSO optimized GEP prediction models made the best outputs in both rootstocks [28].

With MOEAs, inputs are evaluated as multi-objective optimization problems (MOPs) and the solutions specify the best probable balance between two reverse functions [34]. Recently, several mathematical methods have been used to solve MOPs, nonetheless the real MOPs applications are specifically nonlinear and also occasionally non-differentiable [35]. This has enhanced interest in metaheuristic methods, and among these procedures, MOEAs are of special interest. Here, PSO as an evolutionary computation technique was used for determining optimized culture media.

The aim of this study is to employ three soft computing methods namely MLPNN, GEP and KNN and to compare the accuracy of their prediction to multiple linear regression (MLR) technique as well as applying PSO algorithm with aim of predicting and optimizing walnut tissue culture media. Briefly, the new contributions of the present research are:

- Comparing the appropriateness of MLPNN, KNN and GEP nonlinear methods for modeling the impacts of mineral nutrients, PGRs and vitamins on *in vitro* culture of walnut.
- Constructing hybrid models in order to assess how Chandler and Rayen explants respond to the culture medium composition according to the new produced shoots attained from the Taguchi design.
- Finding the optimal composition of culture media to maximize the proliferation rate (PR) and minimize callus weight (CW), shoot tip necrosis (STN) and vitrification (Vit) by optimizing the developed model using PSO.

To our knowledge, this study is the first application of MLR, KNN, ANN, GEP and PSO methods for optimizing walnut tissue culture media. In addition, this work is the first use of KNN modeling procedure in plant tissue culture.

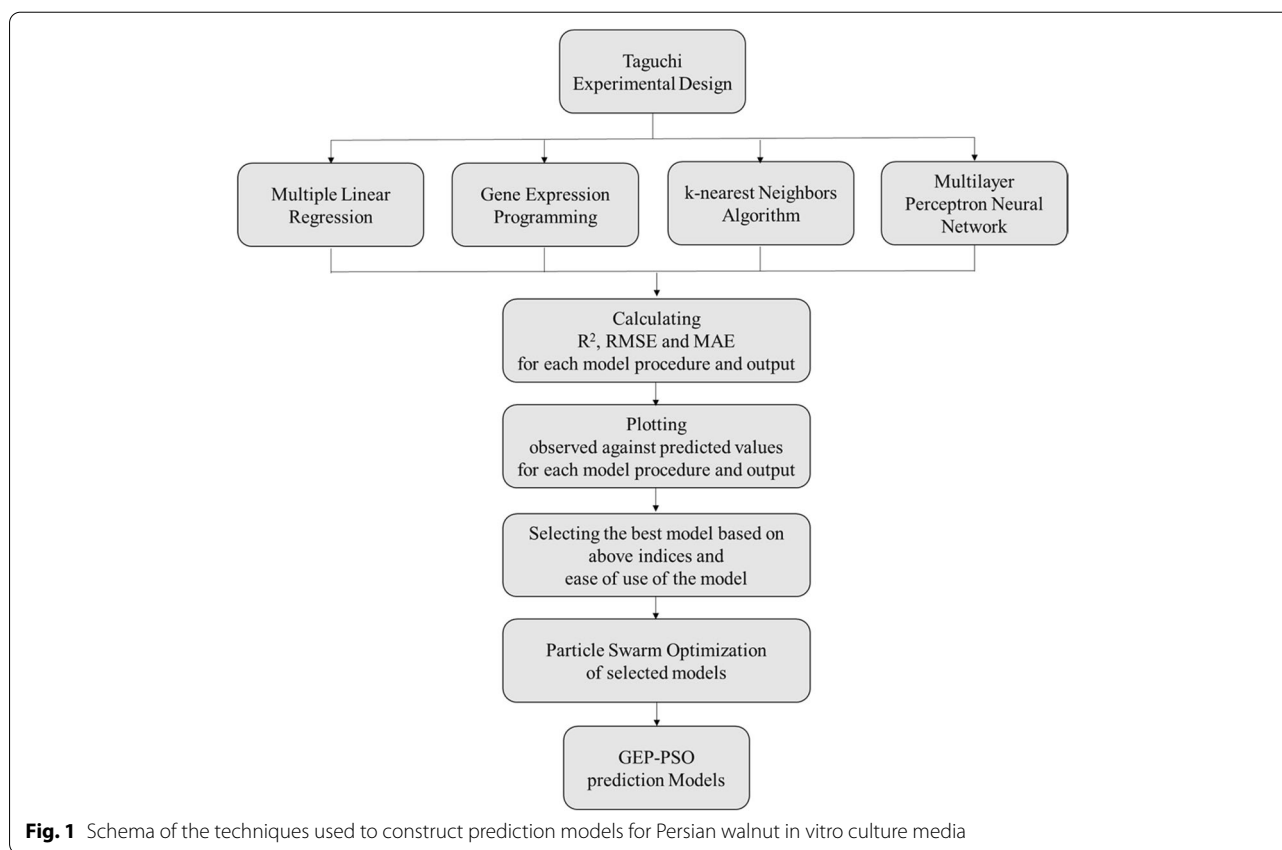
## Results

Our models of the interaction of modifying inputs including nutrients, PGRs and vitamins on outputs including PR, CW, STN and Vit were developed using MLR, MLPNN, KNN and GEP techniques. Here, we assess the developed models' performances through evaluating each modelling method precision to predict the composition of plant micropropagation media for walnut. After that, PSO optimization results of the selected modeling method is investigated to find the most efficient compositions of media for each considered trait. An outline of the techniques used here to achieve the most appropriate model is shown in Fig. 1.

## Comparison of modeling techniques performances

The mathematical equations attained from GEP method, which is showing the best estimate of the explant growth parameters, are shown in Table 1. Moreover, calculated statistics results for output variables (PR, CW, STN and Vit) related to the MLR, MLPNN, KNN and GEP models are given in Table 2. Unlike MLR, the trained MLPNN, KNN and GEP models of PR, CW, STN and Vit resulted in balanced statistic values for both the training and testing subsets (Table 2). For output variables (PR, CW, STN, and Vit) the calculated statistical values corresponding to the KNN, MLPNN and GEP models showed a considerably higher accuracy of prediction than for MLR models as calculated  $R^2$  for MLPNN, KNN and GEP vs. MLR models were: 0.672, 0.695 and 0.802 vs. 0.412 for PR of Chandler; 0.377, 0.354 and 0.428 vs. 0.178 for PR of Rayen; 0.923, 0.931 and 0.844 vs. 0.696 for CW of Chandler; 0.929, 0.930 and 0.839 vs. 0.276 for CW of Rayen; 0.855, 0.915 and 0.807 vs. 0.241 for STN of Chandler; 0.812, 0.831 and 0.808 vs. 0.341 for STN of Rayen; 0.974, 0.975 and 0.853 vs. 0.434 for Vit of Chandler; and 0.977, 0.978 and 0.891 vs. 0.299 for Vit of Rayen, respectively (Table 2).

Comparison of the observed and predicted values of outputs may explain the performance of the developed models according to the studied inputs. A high squared correlation coefficient fitting technique was used to produce plots according to the constructed models derived, to show how each of the four outputs varied as the concentration of media components changed. The plots may be helpful to understand the complete relationship between media components and responses, and to assess the multiple effects of modifying the media components in the DKW medium. The predicted MLR, MLPNN, KNN and GEP models diagrams vs. observed values for the PR, CW, STN and Vit are shown in Figs. 2, 3, 4, 5, 6, 7, 8 and 9. Comparing the fitted simple regression lines of the MLR with ML models showed that MLR resulted in the lowest accordance between the observed and predicted values regarding all considered outputs so that calculated  $R^2$  for MLPNN, KNN and GEP vs. MLR were: 0.696, 0.672 and 0.802 vs. 0.412 for PR of Chandler (Fig. 2); 0.178, 0.359, 0.377 and 0.428 for PR of Rayen (Fig. 3); 0.696, 0.931, 0.924 and 0.844 for CW of Chandler (Fig. 4); 0.276, 0.874, 0.930 and 0.840 for CW of Rayen (Fig. 5); 0.241, 0.916, 0.856 and 0.807 for STN of Chandler (Fig. 6); 0.342, 0.810, 0.813 and 0.809 for STN of Rayen (Fig. 7); 0.435, 0.976, 0.975 and 0.853 for Vit of Chandler (Fig. 8); and 0.300, 0.979, 0.978 and 0.891 for Vit of Rayen (Fig. 9), respectively. Therefore, the ML models were able to accurately predict the outputs while the MLR developed models were not able to describe extensive diversity of growth parameters owing



to the studied variables interaction, that may hide the effects of media components. Figures 2, 3, 4, 5, 6, 7, 8 and 9 may be helpful for realizing the complete relationship between media components and responses, and assessing the combined impacts of modifying the DKW medium components.

According to the results presented in Table 2 and Figs. 2, 3, 4, 5, 6, 7, 8 and 9 as well as the above-mentioned results, MLPNN, KNN and GEP models performed accurately in predicting the effect of media components on in vitro performance of Persian walnut. So, in order to select one of these ML modeling techniques to be optimized and achieve final models for in vitro proliferation of Persian walnut, we considered the ease of using model by the end user. In other words, although MLPNN and KNN performed relatively well, none of these models offer explicit mathematical expression. Unlike MLPNN and KNN methods which produce black-box models, GEP can provide the researchers with an opportunity to optimize the extractive equations (optimal values of the variables) by generating explicit mathematical equations between the independent variable and the dependent variable and can be used as an equation for the pre-test stages (initial phase of the study) in designing and developing of their studies. Hence, we selected GEP models to

be optimized and achieve proliferation media formulations of Chandler and Rayen.

#### Optimization of GEP models

Consequently, to achieve the optimized medium resulting in the highest PR and the lowest CW, STN and Vit in walnut, we optimized developed GEP models by using multi-objective PSO technique.

The optimized amounts of the studied factors and the predicted values of growth parameters by the GEP models are shown in Table 3. The PSO optimization of the GEP models revealed that media containing  $1.76 \times \text{NH}_4\text{NO}_3$ ,  $\text{CaNO}_3$  and  $\text{ZnNO}_3$ ,  $1.67 \times \text{KNO}_3$ ,  $0.96 \times \text{K}_2\text{SO}_4$ ,  $0.66 \times \text{MgSO}_4$ ,  $\text{MnSO}_4$  and  $\text{CuSO}_4$ ,  $2.35 \times \text{KH}_2\text{PO}_4$ ,  $\text{H}_3\text{BO}_3$  and  $\text{Na}_2\text{MoO}_4$ ,  $1.64 \times \text{FeEDDHA}$  and  $1.89 \times \text{Thiamine}$ , Nicotinic acid and Glycine concentrations in DKW medium, 0.67 mg/l BAP and 1.30 mg/l TDZ and 1.30 mg/l IBA could lead to optimal PR (23.54), CW (0.12), STN (2.23) and Vit (9.95) in Chandler and media containing  $0.73 \times \text{NH}_4\text{NO}_3$ ,  $\text{CaNO}_3$  and  $\text{ZnNO}_3$ ,  $0.69 \times \text{KNO}_3$ ,  $0.94 \times \text{K}_2\text{SO}_4$ ,  $0.64 \times \text{MgSO}_4$ ,  $\text{MnSO}_4$  and  $\text{CuSO}_4$ ,  $0.83 \times \text{KH}_2\text{PO}_4$ ,  $\text{H}_3\text{BO}_3$  and  $\text{Na}_2\text{MoO}_4$ ,  $1.35 \times \text{FeEDDHA}$  and  $1.52 \times \text{Thiamine}$ , Nicotinic acid and Glycine concentrations in DKW medium, 0.67 mg/l BAP and 1.23 mg/l TDZ and 1.23 mg/l IBA could result

**Table 1** Constructed models using gene expression programming to predict explant growth traits in Persian walnut

Walnut	Equation
Chandler	$PR = \left[ X3 - \left( \left( \cos\left(\frac{X8}{X1}\right) \times 6.178 \right)^3 \times X5 \times X6 \right) \right] + \left[ X6 + \left( 8.010 - \left( \frac{\cos(X4) \times 0.989}{\cos(X8 + X3)} \right) \right) \right] + [-3.539 - \sin(((5.549 + X1) \times (-0.883) \times X4) + (X7 - X9 + (X2 - 3.579)))]$ $+ [X4 + (X7 - (\cos((X2 \times X1) \times (-8.741) \times X7 \times X1 \times X5)))]$ $CW = \left[ 4.719 \times \frac{X1}{(5.715)^{X1} - (X8^{4.879} + \frac{6.250}{X8})} \right] + \left[ \cos(-4.432 - (X9 - X3 - 0.208) \times (X5^{X5} \times 10.590)) \right] + \left[ \left( \frac{7.082 - 6.718X4}{(X6 - 6.287)^3} \right) \times (X3 - X7X4) \right]$ $+ [\cos^2(X9 - (\cos((4.957 - X2) - \cos(X4)))^4)]$ $STN = [-5.448 \times \sin(X7 \times (X7 + ((X5 + 5.490) \times (X1X4))))] + [7.232 \times (3.087 - ((\sin^2(X2 - X6) \times \sin(X5 - 3.272)))]$ $+ [(X7 \times (X7 - X3 + X2)) \times (X8 + \sin(3.793X9))] + [X6 \times (X9 + (X8 + (X7 \times (X8^2 - 6.451))))]$ $Vit = \left[ \sin\left(X6 - \left(X2 \times \left(1.392X1^{F8^3} + 1.334X9\right)\right)\right) \right] + \left[ \sin(X1) + X1 + X1^3 + \left(\frac{X9}{X7} \times 2X8\right) \right]$ $+ \left[ \cos\left(\left(2X1 + X2\right) \times \left(X1 + X4 - X3^3\right)\right) \right] + \left[ X5^{F9} \left( \frac{6.982 - X5}{X2 + X5} \right) \times (\sin(X7^{X1})) \right]$
Rayen	$PR = \left[ \sin\left(\cos\left(X2^3 \times \left(X4 + X6 \times \frac{-2.567}{X9}\right)\right)\right) \right]^9 + [X1 + (\sin(\cos(X3 \times X1)) - (\sin(X3^{X9}))) + 3.955]$ $+ \left[ \left( X6 \times \frac{\sin(X5)}{\sin(X7) + X5^{X9}} \right) + X3 \right] + [\sin(X5 \times (X8^{39.898})) - 0.108X7]$ $CW = \left[ \sin\left(\frac{\sin\left(\frac{X9}{X6} + 1.351X7\right) - \left(\frac{\cos(X5)}{-2.339}\right)\right)}{-1.721}\right) \right] + \left[ \sin\left(\cos\left(\left(\sin(X7) + 0.566\right)^2 \times \left(\frac{X6^3 \times X3}{-7.183}\right)\right)\right) \right]$ $+ \left[ \cos\left(0.089 + \left(\cos\left(\frac{X6}{X4 - 3.919}\right)\right)\right) \right]^{X2} + \left[ \sin\left(X1 - X5 + 0.282\right) \times \left(\frac{\cos(X1)}{-1.095 - X8}\right) \right]$ $STN = [6.002 \times ((\cos(3.862X3 + X9 - X1) \times (\sin(X7))) - X4)] + [X2 \times (X3 + \cos(X2 \times (X8 - X8 + 3.742)))]^2$ $+ \left[ \left( (X9 - 4.711)^2 - (X6 \times X4 \times X1) \right) + 45.170 \right] + \left[ \frac{\cos(F1)}{(X5 - 1.460) - (\cos(X2) \times X2^{X1})} \right]$ $Vit = \left[ \frac{X4}{\sin(F1 \times (9.069 + F4F6 - F4F3))} \right] + [X2 + (X1^3 - (X8 + X5)^{-2.497X9} + X1)] + [8.504 + (X7^X7 \times ((X9^{-1.651} - X6) \times (X3 - X6)))]$ $+ \left[ \frac{X8 \times (\sin(X2 - X6) \times (X9 \times (X2 - X5)))}{\sin(X7)} \right]$

X1–X9: are input factors as presented in Table 3

**Table 2** Evaluation of different developed models using various statistics for PR, CW, STN, and Vit of Persian walnut through *in vitro* proliferation

Walnut	Output	Model	Train Set *			Test Set		
			RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
Chandler	PR	MLR	2.840	1.986	0.264	3.313	2.270	0.412
		KNN (4) **	1.121	0.814	0.909	1.756	1.257	0.672
		MLPNN (9-40-1) ***	1.093	0.796	0.913	1.658	1.186	0.695
	CW	GEP	1.496	1.093	0.805	1.711	1.299	0.802
		MLR	0.225	0.163	0.640	0.190	0.146	0.696
		KNN (4)	0.068	0.046	0.955	0.094	0.063	0.923
	STN	MLPNN (9-15-1)	0.063	0.045	0.960	0.083	0.058	0.931
		GEP	0.113	0.086	0.886	0.124	0.105	0.844
		MLR	7.628	5.787	0.315	7.827	6.614	0.241
	Vit	KNN (5)	3.324	2.432	0.867	3.529	2.767	0.855
		MLPNN (9-30-1)	2.777	1.994	0.917	2.660	2.026	0.915
		GEP	3.980	3.080	0.814	3.910	3.156	0.807
	Rayen	MLR	6.189	4.930	0.435	6.909	5.538	0.434
		KNN (3)	0.992	0.735	0.986	1.232	0.960	0.974
		MLPNN (9-30-1)	0.993	0.699	0.988	1.184	0.938	0.975
Rayen	PR	GEP	2.461	1.901	0.906	3.371	2.584	0.853
		MLR	1.310	1.041	0.181	1.388	1.191	0.178
		KNN (4)	1.049	0.843	0.471	1.159	0.970	0.377
	CW	MLPNN (9-12-1)	1.008	0.832	0.512	1.169	1.000	0.358
		GEP	1.039	0.858	0.459	1.178	0.992	0.428
		MLR	0.400	0.319	0.286	0.460	0.382	0.276
	STN	KNN (4)	0.069	0.050	0.949	0.078	0.057	0.929
		MLPNN (9-30-1)	0.067	0.049	0.952	0.078	0.056	0.930
		GEP	0.207	0.160	0.853	0.223	0.178	0.839
	Vit	MLR	12.757	10.150	0.046	10.170	8.273	0.341
		KNN (5)	2.671	2.087	0.821	3.378	2.805	0.812
		MLPNN (9-30-1)	2.574	2.47	0.837	3.224	2.606	0.831
	Rayen	GEP	4.916	3.822	0.858	5.238	4.242	0.808
		MLR	7.033	5.383	0.370	7.101	5.627	0.299
		KNN (4)	1.228	0.948	0.980	1.211	1.022	0.977
Rayen	MLPNN (9-20-1)	1.197	0.909	0.982	1.192	0.975	0.978	
	GEP	2.849	2.235	0.899	2.937	2.452	0.891	

\* Average of tenfold cross validation

\*\* The number of neighbors (k) leading to the best performance

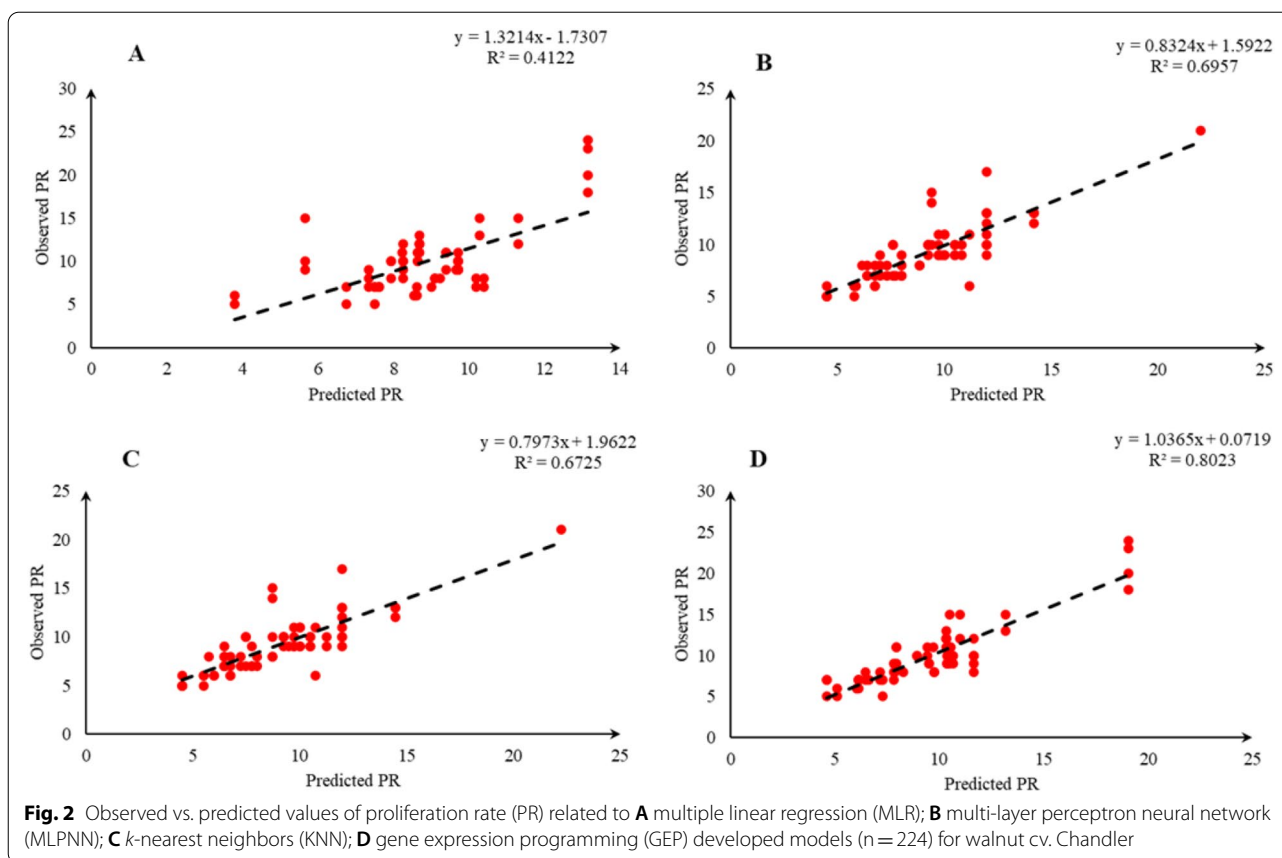
\*\*\* The MLPNN architecture (inputs—hidden layers—outputs)

in optimal PR (24.57), CW (0.64), STN (12.48) and Vit (3.04) in Rayen.

## Discussion

Walnuts as one of the important woody plants are considered recalcitrant to *in vitro* culture in which genetic determinism besides other factors such as media components makes more complicated different stages of micropropagation, as well. In the present study, three different ML modeling approaches along with PSO optimization

algorithm were applied to determine and predict the effect of genotype and the media formulation throughout the proliferation of walnut. Walnut micropropagation can be improved by involving different physiological disorders in modeling and optimization processes. The incidence of physiological disorders through micropropagation of walnut has not been comprehensively investigated. Different studies on walnut tissue culture have been focused on introducing some chemicals like phloroglucinol and FeEDDHA to DKW or [36] (MS) basal



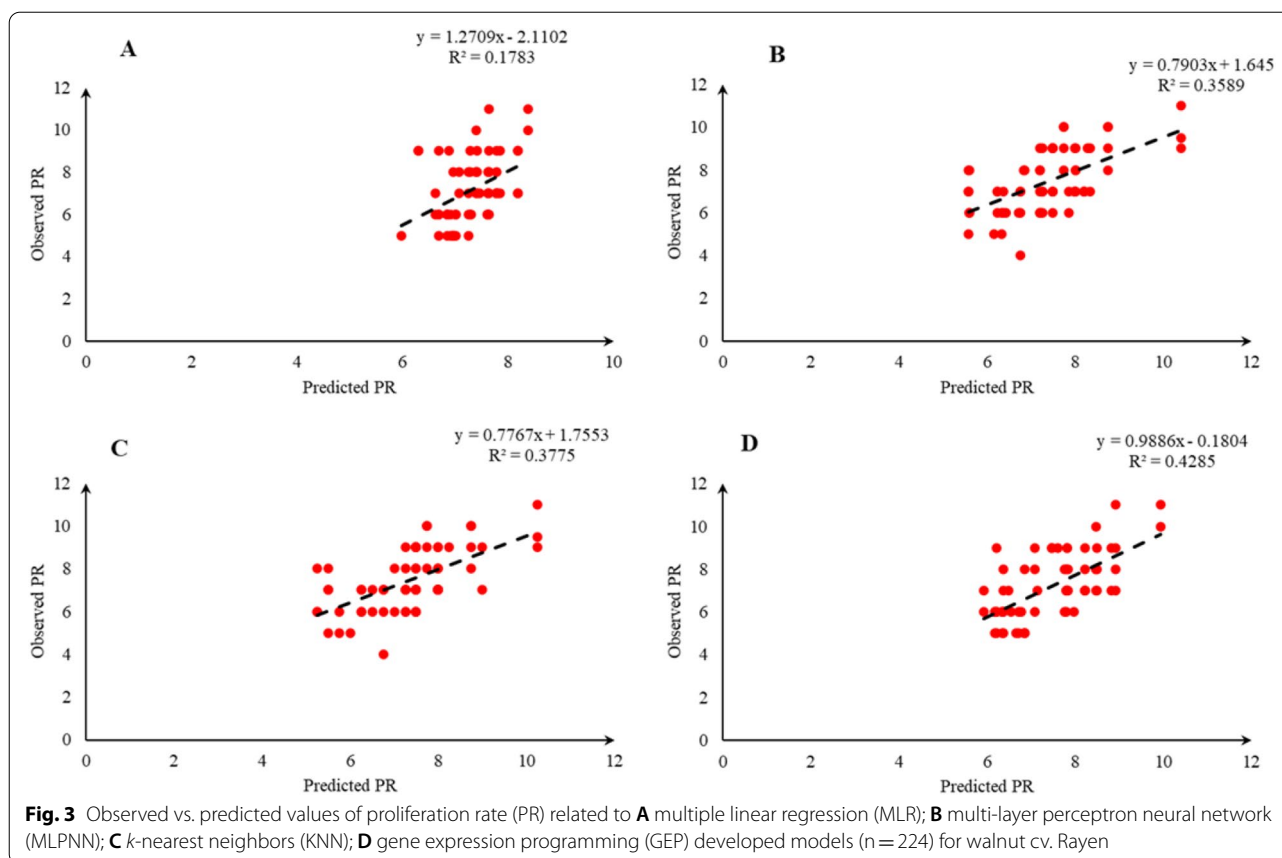
media [11, 15], supplementing media with various concentrations of different PGRs [37–39], removing agar [39], ventilation and reducing sucrose concentration [40], but a few of the studies focused on media components, including mineral nutrients [9, 41], vitamins and PGRs [39] interaction on proliferation quality and quantity.

Here, we concentrated on increasing PR and reducing important abnormalities occurring during this phase, by recording data associated to several designed experiments. The subsequent database including a range of concentrations of each component in culture media allows simultaneous evaluation of the impacts of all minerals, vitamins and PGRs used in media as well as genotype on the explant growth indices only through the ML tools.

Machine learning as a powerful tool has been effectively applied in plant biology studies [42, 43] including plant tissue culture data analysis and accurate prediction of optimal *in vitro* culture media composition [20, 24–28]. The development of *in vitro* plant tissues is controlled by minerals, vitamins and PGRs in the culture media. To achieve maximum explant performance, the prediction of the most efficient media composition is highly useful since the optimization of the type and

concentration of minerals, vitamins and PGRs in media is a time-consuming, expensive and laborious job [9, 41].

In our previous studies, we successfully performed constructing neural models using ANN technique to study the effects of different combinations of minerals and PGRs on *in vitro* proliferation and rooting of  $G \times N15$  *Prunus* rootstock [25–27]. Our study on comparing ANN with MLR modeling to forecast the optimum concentrations of macronutrients for OHF 69 and Pyrodwarf *Pyrus* rootstocks *in vitro* media showed ANN as a precise and promising technique [24]. The important benefit of ANN-based methods is that they do not need a prior identification of proper fitting function consequently; they have an overall approximation ability to calculate all kinds of non-linear functions in practice. This trait may help the modeler to develop the most possible precise model. Despite the fact that ANN is a good alternative for MLR, it does not provide us any equations including the relationships between input and output variables. Moreover, the ANN technique needs a time-consuming process of trial and error to find network parameters like number of neurons and hidden layers [44–46]. ANNs as the most extensively used ML model, can efficiently solve



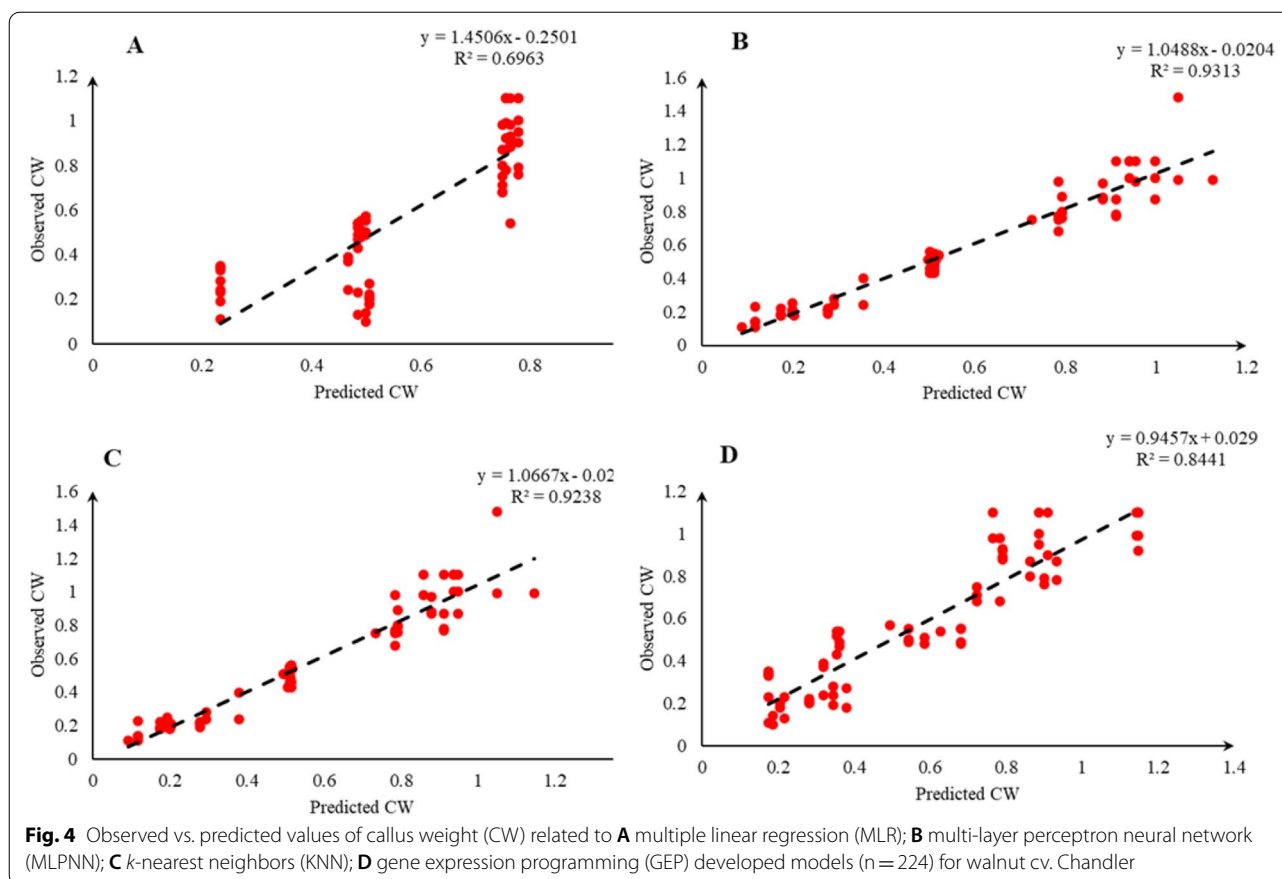
different multivariate, non-linear and nonparametric problems via an unidentified “black box” training [47]. Nevertheless, there are also some drawbacks with ANN “black box” nature [48]. In general, ANN is unable to clarify its logical process and this constraint makes ANN application unfriendly in natural science studies, as it can just simulate the change process according to experimental data, without helping us to understand the reason of the change.

Considering these restrictions in using ANN models, in another study on *Pyrus* rootstocks in vitro proliferation [20] we compared the power of GEP technique to ANN (RBFNN) and MLR in predicting the optimal media. RBFNN and GEP exhibited higher performance precision towards the MLR, and the GEP resulted in the most precise model as well as being practical [20]. In our recent research [28], we used two algorithms, GEP and M5' model tree to overcome the ANN method weaknesses and simplify forecast of the media components interactions on in vitro proliferation of *Pyrus* rootstocks. Again, we found GEP as a more accurate technique than M5' model tree [28].

Consequently, in the present study, we applied GEP as the most precise modeling procedure found by [20, 28],

MLPNN as an ANN technique that its models are easier to give precise prediction than RBFNN when input data are randomly distributed [49] and KNN as one of the simplest machine learning approaches which can also be used for regression problems [50]. The MLR was also applied as a linear modeling method to be compared with above-mentioned ML procedures in predicting the optimum in vitro proliferation media composition of walnut to achieve the most appropriate outcomes. The accuracy of the developed prediction models was evaluated using MAE, RMSE and  $R^2$  statistics and correlation coefficient between observed and predicted values of each output. To our knowledge, KNN algorithms have not ever been applied to predict the plant tissue culture media composition. The advantage of KNN algorithm is that it does not require specific assumptions about the predictors' distribution. The samples of KNN are classified according to the  $k$  neighbor responses mean values in a space of predictor [51]. The examples of training are defined by  $n$  traits. Each example means a point in a space with  $n$ -dimension. So, all examples of training will be kept in a space with the pattern on  $n$ -dimension. Here, the number of neighbors ( $k$ ) leading to the best results for each model are presented in Table 3.



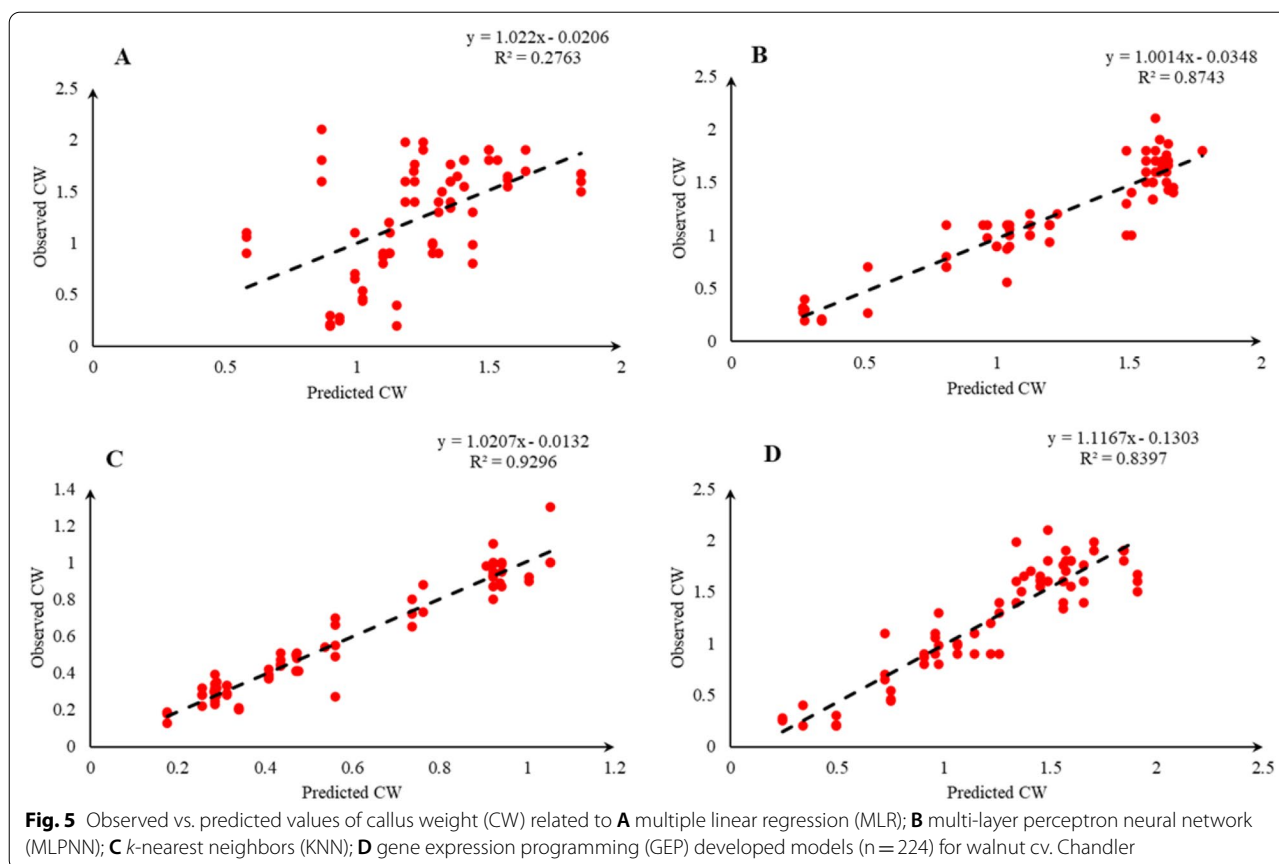


A key advantage of GP-based procedures such as GEP, toward other methods is that they do not need any hypothesis for preceding form of the relationship to produce prediction equations. GP and its deviations have been applied in many researches to find any complicated relationships which fit different experimental data [52–54]. An individuals' population is employed in this technique and afterwards, better individuals are chosen by using genetic variations and fitness function. The genetic variations are introduced by genetic operators. Machine learning approaches including GEP have been programmed to learn the variables' relationships in data collections. GEP difference with GA and GP as its precursors is in the method of individual programming so that in GEP, individuals are programmed as chromosomes i.e. fixed length linear strings which are presented finally as a simple diagram called expression tree. Whereas, in GA and GP, individuals are expressed, as nonlinear entities with different shapes (parse trees) and sizes and chromosomes, respectively. One of the GEP strengths over GA and GP is that genetic operators work very simple at the level of chromosome in GEP making development of genetic diversity. GEP unique, multi-genic nature is another important point which allows more complicated

programs with multiple sub-programs to be developed. The advantages of both GA and GP are collected in GEP, whereas some of their constraints are met [55].

Based on our results presented in Table 2, KNN, MLPNN and GEP models were much more accurate than MLR. On the other hand, in most cases, the MLPNN method provided better fit calculation than KNN and GEP. But based on the results of our aforementioned studies [20, 28], the optimized GEP method provides better fit calculation than other approaches. Furthermore, GEP is preferred over ANN models, as ANN is a black-box model, whereas GEP explains the constructed prediction models with mathematical Eqs. [54].

Through the previous years, GEP has been applied extensively in other areas because of its high efficiency and effectiveness. GEP applications are so wide and are rapidly enhancing [55]. GEP is one of the most effective function mining algorithms which has been widely used in classification, pattern recognition, prediction, and other research areas. This algorithm can mine an ideal function to deal with further complex tasks [56]. GEP has been used to determine the quality and stress of water on lakes or rivers as a result of the wastewater pollutants [31]. The problem of missing values in data set



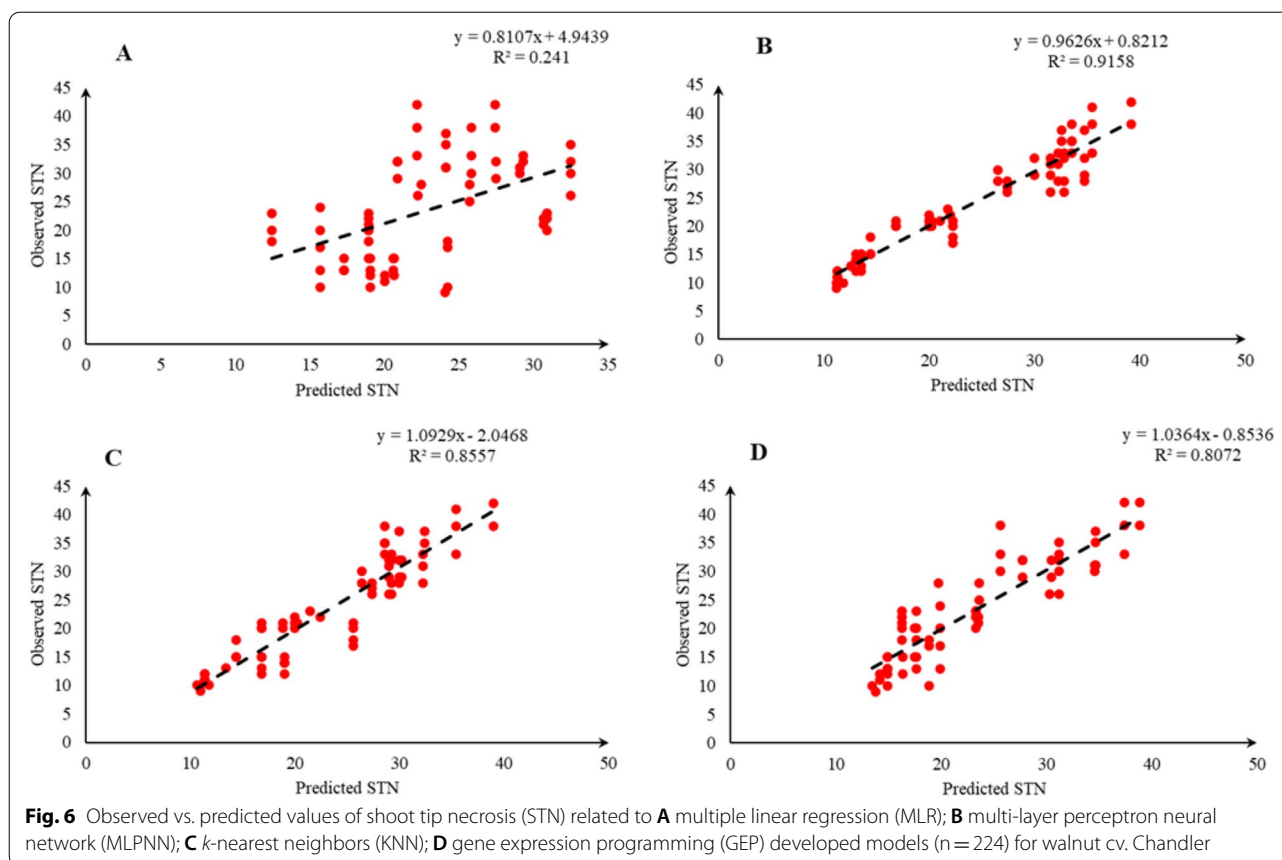
due to the measurement conditions can simply be solved by employing GEP [31]. Results based on actual data set confirmed that the multiple GEP and fuzzy expert system outperforms detection methods in medical field by attaining high prediction precision [57].

Our previous studies [20, 24, 28] on pear rootstocks using ML-based modeling showed that there are different responses to the concentrations of macronutrients and PGRs based on genotypes, as we found here in Persian walnut varieties. Regarding the complex interactions, detection of the optimum levels of minerals and PGRs for a certain plant genotype is complicated [58]. Furthermore, the incidence of physiological disorders like Vit and STN throughout the proliferation phase of walnut needs improvement of media for optimal growth of explants. Constructing optimized and effective media by using authentic mathematical modeling and optimization methods have been performed previously on different plant species [17, 20, 24–28, 59–61]. Here, we consequently suggested use of ML-based modeling to recognize concentrations of minerals and PGRs that would maximize PR while minimizing CW, STN and Vit [24]. As we found here (Table 3), our previous results on pear [20, 24] showed that ANN prediction models had

higher precision than MLR models and MLR could not be a trustworthy method for assessing nonlinear or non-polynomial relationships among variables.

It has been revealed from our recent study on pear rootstocks micropropagation [28] that the most efficient optimization method for optimizing GEP models was multi-objective PSO. Therefore, here, we used multi-objective PSO method for optimization of selected GEP models. Our GEP-PSO optimized models could give us intact optimized formula for proliferation of Chandler and Rayen (Table 3).

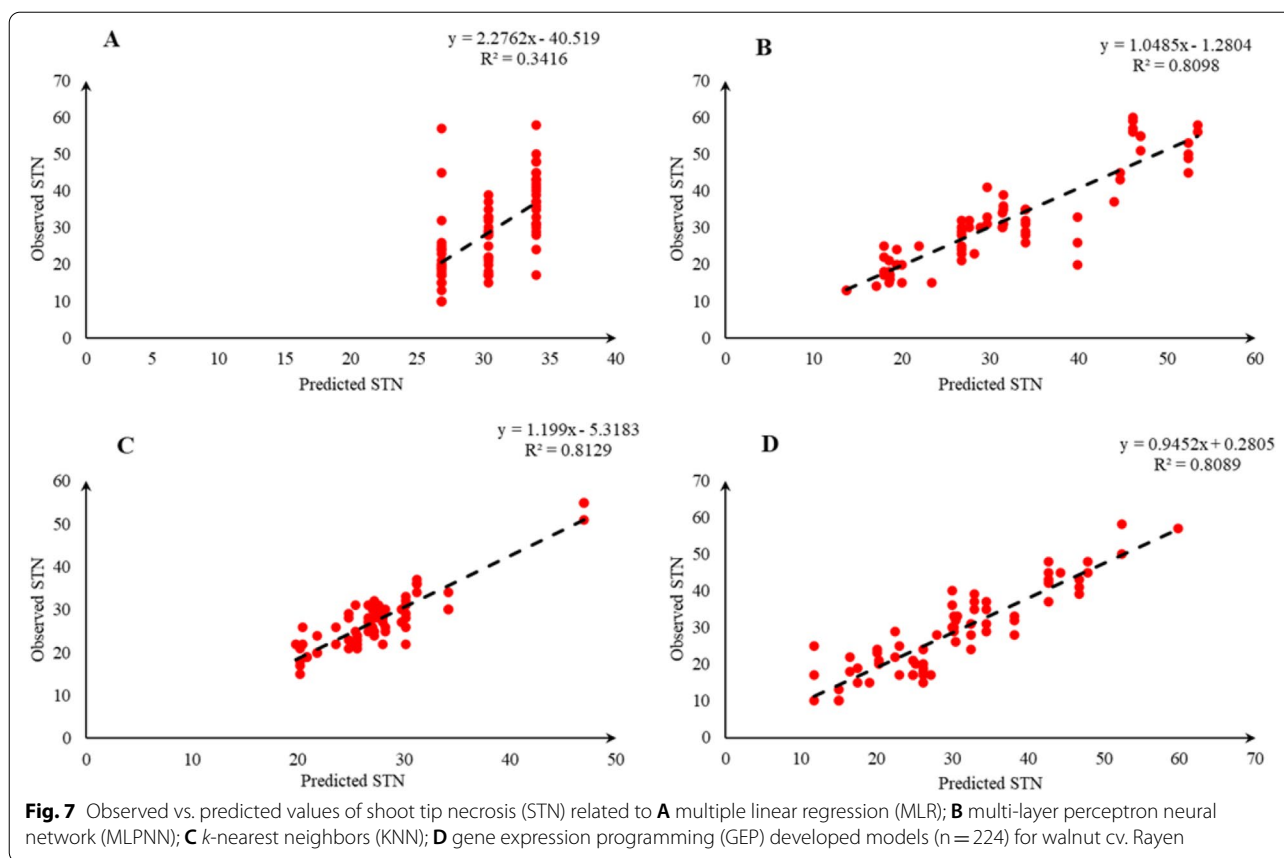
The mono-objective GA optimized MLPNN and RBFNN-based models obtained in our previous studies [20, 24] on Pyrodwarf and OHF *Pyrus* rootstocks showed the significance of some minerals such as  $\text{NH}_4^+$  and  $\text{NO}_3^-$  and/or PGRs for explants proliferation. Our previous research [25] on  $G \times N15$  *Prunus* rootstock by using mono-objective GA optimized ANN models found the higher importance of  $\text{NO}_3^-$ ,  $\text{NH}_4^+$ ,  $\text{Ca}^{2+}$ ,  $\text{K}^+$ , and  $\text{PO}_4^{2-}$  towards  $\text{Mg}^{2+}$ ,  $\text{Cl}^-$  and  $\text{SO}_4^{2-}$  for in vitro proliferation. Our recent study [28] on *Pyrus* rootstocks using mono-objective GA optimization of GEP models indicated that high PR may cause low quality plantlets. In accordance with it, our study [25] on  $G \times N15$  using mono-objective



GA optimization of ANN models also predicted that increasing the  $\text{NH}_4^+$  concentration will enhance shoot number and length with higher number of non-healthy shoots but decreasing amount of  $\text{NH}_4^+$  will enhance the plantlets quality. Our results [28] on pear rootstock using RBFNN and GEP modeling procedures also indicated that a lower content of nitrogen will result in higher quality plantlets.  $\text{NH}_4^+$ ,  $\text{NO}_3^-$  and  $\text{K}^+$  interaction has been the main subject of most in vitro studies [62] but using ML models, [63] reported interaction of  $\text{K}^+$ ,  $\text{EDTA}^-$  and  $\text{SO}_4^{2-}$  with critical effect of  $\text{K}^+$  on PR of pistachio; as low and high concentrations of  $\text{K}^+$  resulted in the highest and lowest PR, respectively. Study on *Prunus* sp. also showed that  $\text{K}^+$  at low concentration promotes PR [64]. Nezami-Alanagh et al. [63] concluded that either low or too high amounts of  $\text{K}^+$ ,  $\text{EDTA}^-$  and  $\text{SO}_4^{2-}$  ions result in low quality plantlets. Considering macro- and micro-elements, our multi-objective PSO optimized GEP models in Chandler showed that increasing  $\text{NH}_4^+$ ,  $\text{NO}_3^-$  and  $\text{SO}_4^{2-}$  increased PR and Vit while decreasing CW and STN. But the results in Rayen showed that increasing  $\text{SO}_4^{2-}$  except  $\text{K}_2\text{SO}_4$  as well as increasing  $\text{NO}_3^-$  except  $\text{KNO}_3$  increased PR and CW while decreasing STN and Vit (Table 3).

Reed et al. [65] emphasized on the optimization of nitrogen components content of the culture media to stimulate high number of elongated shoots and reduced amount of callus, in different pear species. Nezami-Alanagh et al. [66] suggested avoiding high content of  $\text{NH}_4^+$  to reduce callus formation in the in vitro pistachio shoots. Low amounts of some of the MS medium components such as  $\text{KNO}_3$ ,  $\text{MgSO}_4$ ,  $\text{KH}_2\text{PO}_4$ ,  $\text{CaCl}_2$ , and  $\text{NH}_4\text{NO}_3$  have been reported to contribute to STN promotion in some *Pyrus* species [67]. Whereas based on our results, lower concentrations of  $\text{K}_2\text{SO}_4$ ,  $\text{MgSO}_4$ ,  $\text{MnSO}_4$ ,  $\text{CuSO}_4$  in Chandler and  $\text{K}_2\text{SO}_4$  and  $\text{KNO}_3$  in Rayen reduced the occurrence of STN. The results of [63] using neurofuzzy logic showed that low amount of  $\text{K}^+$  and mid to high concentrations of  $\text{SO}_4^{2-}$  inhibit the STN in pistachio explants with lower signs in UCB1 than in Ghazvini which refers to the genotypes differences as we found in our study. Ion confounding problem again prevents determining exact relationship between a given mineral and the physiological disorder.

The neurofuzzy logic procedure show a linear positive effect of nicotinic-acid and pyridoxine-HCl on pistachio parameters of shoot multiplication [68], but, to our knowledge, there is no study about the impact of

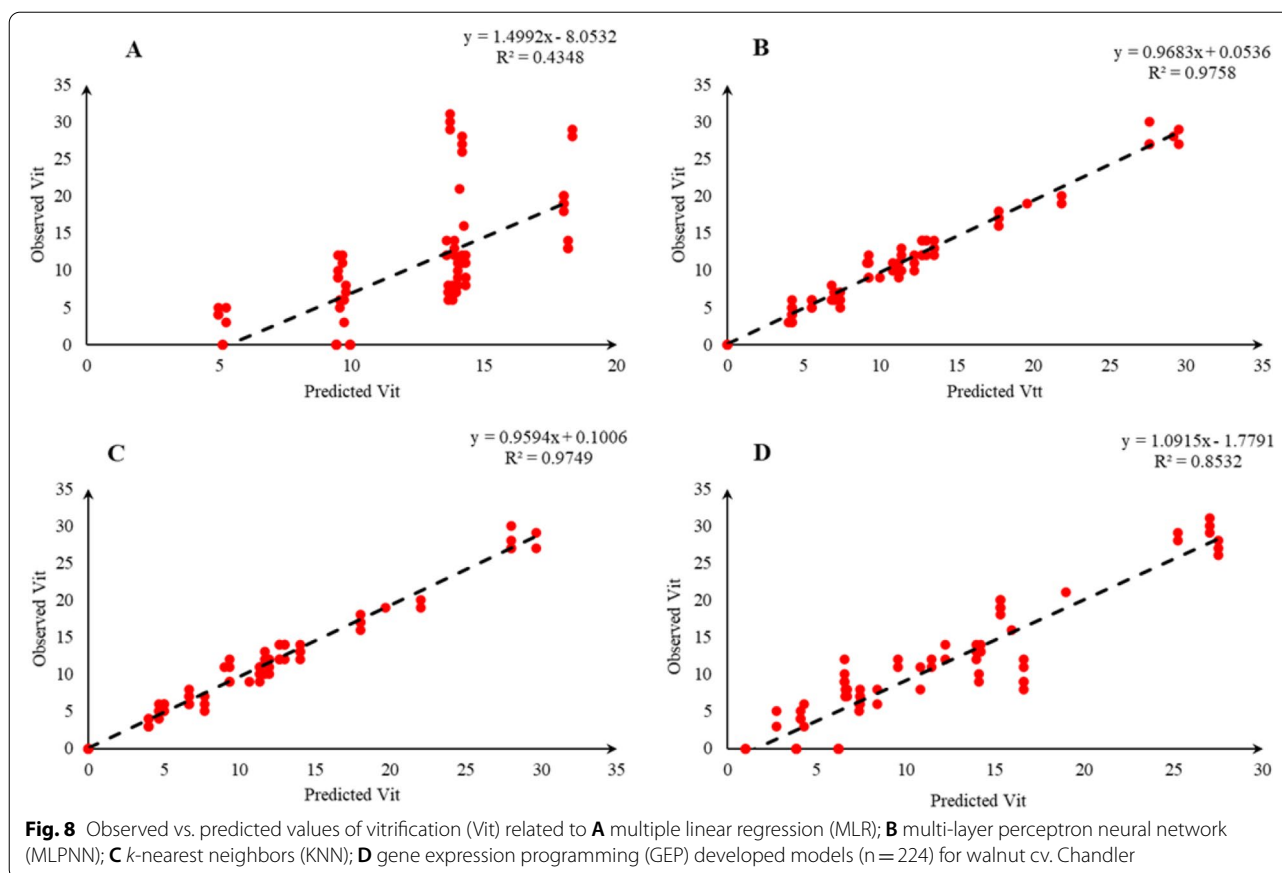


vitamins on the proliferation of walnut. Nezami-Alanagh et al. [66] showed that the glycine and thiamin-HCl affected differently on some in vitro disorders of pistachio. They showed that increasing glycine content highly reduced the development of callus. Our study showed that higher content of vitamins reduced CW in Chandler (Table 3) and reduced vitamins content in Rayen which caused higher CW (Table 3). Rayen was more recalcitrant to micropropagation than Chandler, hence, achieving higher PR and lower incidence of STN and Vit can cover the low increase in CW.

Genotype is an important factor influencing the occurrence of physiological disorders in walnut which is in agreement with reports of [63, 66] in pistachio. Similarly, other researches on pear [20, 24, 28, 67] explained that the in vitro physiological disorders incidence caused by unbalanced mineral nutrition differed among genotypes.

The purpose of our study was to present an ML approach with high accuracy for prediction of optimized culture media. We applied techniques of MLPNN, KNN and GEP combined with PSO to walnut proliferation data sets to achieve the most appropriate proliferation results. Comparison of our results with the previous ones [20, 24–28, 63, 64, 66] indicates that using at least

two methods together results in more precise consequences. So that, comparing the results of the used methods showed the effect of media components enhancing or reducing the measured parameters (Table 3). The efficiency of the developed optimized media was compared to DKW. The media constituents proposed by our PSO optimized GEP models related to Chandler showed that decrease in  $K_2SO_4$ ,  $MgSO_4$ ,  $MnSO_4$ ,  $CuSO_4$  and BAP besides increase in other nutrients, PGRs and vitamins increased PR as well as Vit while reducing CW, and STN. Nevertheless, it was slightly different for Rayen as decrease in  $K_2SO_4$ ,  $KNO_3$ , vitamins and BAP along with increase in remained nutrients and PGRs caused higher PR and CW but lower STN and Vit (Table 3). The use of macro- and micro-nutrients as factors, in many micropropagation studies [20, 24, 28, 63, 68], indicates the ion confounding problem, being problematic to recognize precisely corresponding ion(s) affecting the studied parameter [69]. Our results in comparison to previous studies on walnut [11, 15, 37–39] which were about minerals and/or PGRs effects, showed for the first time that not only the effects of minerals depend on the used PGRs concentration but vitamins concentration affects the explant response. The interaction of minerals, PGRs

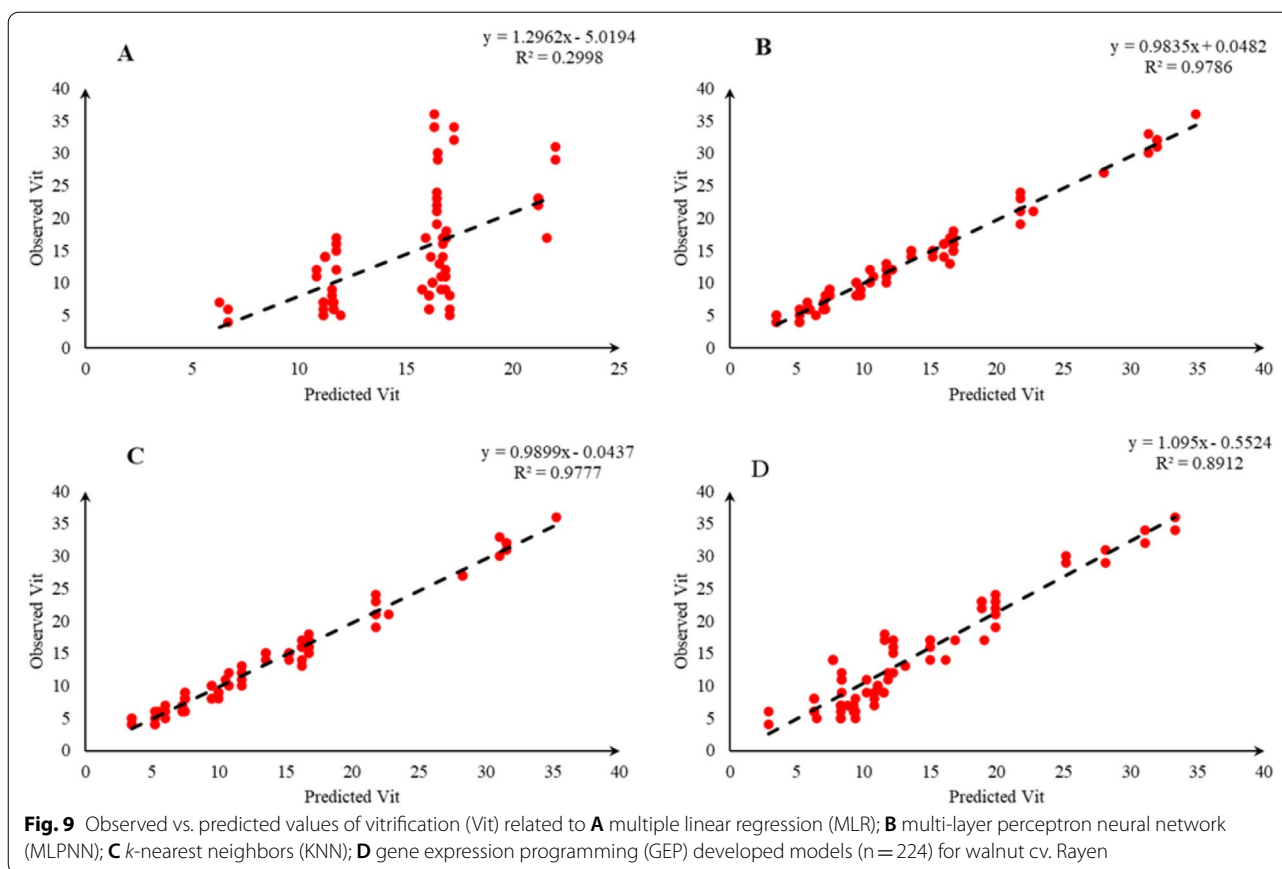


and vitamins could determine the quantity and quality of proliferated plantlets. The plant species and genotype are also highly important in predicting the explant growth response to the minerals, PGRs and vitamins interaction.

Plant PGRs interactions make a critical complication in regulating the processes of plant growth, as well. Cytokinin controls cell proliferation [70] and auxin enhances the sensitivity of apical meristem less mitotically active cells to cytokinin [71]. Cytokinin to auxin ratio is a key signal which controls phenotype [72]. As auxin and cytokinin have roles in DNA replication and cell cycle regulation, respectively [73]. PGRs effects may vary with plant species. Ref. [26] results on *Prunus* rootstock indicated that applying cytokinin and auxin together will result in higher PR than employing each one alone. According to their results, PGRs concentration and interaction are also important. According to these results and [74] and [75] findings, we used various concentrations of BAP, TDZ and IBA in our experiments. Our adverse results can be attributed to the interaction of genotype and culture medium constituents [76] with PGRs [20]. Type and concentration of cytokinin highly affected in vitro growth and survival of black walnut [39]. Ref. [37] reported that lower concentrations of zeatin was better than BAP for

fast shoot elongation of black walnut nodal explants, while higher levels of zeatin and BAP led to shoot necrosis. Using TDZ at 0.01–0.02 mg/l in the medium resulted in an enhanced rate of morphological disorders [37]. But higher levels of TDZ (1.30 and 0.52 mg/l in Chandler and Rayen, respectively) in our present study resulted in reduction of STN in both Chandler and Rayen. *Juglans regia* was successfully micropropagated using 0.1–2.01 mg/l BAP [4, 8, 12, 77–81]. Our used BAP concentrations (0.67 and 0.99 mg/l for Chandler and Rayen, respectively) are also in this range. There is no result in the literature about the effect of BAP on the incidence of walnut in vitro physiological disorders. But according to the results of in vitro studies on other plant species like pistachio [64, 82, 83], addition of adequate amount of BAP strongly decreases the incidence of STN.

Therefore, in the present study, we evaluated the interaction of cytokinin and auxin PGRs and medium components including nutrients and vitamins on proliferation of walnut to achieve the most efficient protocol with a reasonable range of PGRs. Our analyses using PSO optimized GEP modeling technique showed that this method can be used as an efficient procedure for evaluating the



interaction of different factors on walnut explant growth indices in proliferation phase. Therefore, for the first time GEP is introduced as a great tool in optimizing higher quality and efficiency walnut tissue culture protocols in less time.

Callus development during explant proliferation is a common problem in walnut micropropagation which has been reduced here by increasing PR in Chandler while enhanced by increasing PR in Rayen (Table 3). Yegizbayeva et al. [15] reported that callus formation is not correlated with PR in walnut. Callus formation has been attributed to certain concentrations of different mineral nutrients in various plant species like  $\text{KH}_2\text{PO}_4$ ,  $\text{CaCl}_2$  and  $\text{MgSO}_4$  in some *Prunus* cultivars [67],  $\text{NO}_3^-$  in germplasm of *Robus* [84] or  $\text{MgSO}_4$  in *Prunus armeniaca* [85]. Akin et al. [86] reported  $\text{NH}_4^+$  and after that genotype and  $\text{SO}_4^{2-}$  as significant factors affecting callus formation in hazelnut in vitro proliferation using CHAID analysis. Nezami-Alanagh et al. [63] using neurofuzzy logic predicted that high and low concentrations of  $\text{Fe}^{2+}$  and  $\text{SO}_4^{2-}$ , respectively, result in the lowest callus formation in pistachio rootstocks explants. They suggested that lower concentration of  $\text{SO}_4^{2-}$  in MS reduces shoot tip necrosis and callus

development in pistachio in vitro proliferation. While our results showed that lower concentrations of both FeEDDHA and minerals containing  $\text{SO}_4^{2-}$  in DKW caused lower CW in both Chandler and Rayen. Bosela et al. [37] showed that the high-salt media i.e. DKW and MS resulted in lower Vit vs. WPM and 1/2X DKW media in walnut.

## Conclusions

Walnut micropropagation is a problematic process with lots of in vitro drawbacks including necrosis, callusing and vitrification. The present study demonstrated the efficiency of plant in vitro proliferation predictive models by using advanced ML modeling procedures. Therefore, a regression model i.e. MLR and three advanced ML models including MLPNN, KNN, and GEP were constructed to predict walnut in vitro PR and associated physiological disorders under the effect of culture medium constituents and genotype. According to the results, following conclusions and suggestions are presented:

- Advanced computational models are the highly precise approaches which can be applied to control

**Table 3** Multi-objective PSO optimization of GEP models to achieve the highest quantity and quality through in vitro proliferation of walnut

Walnut	Medium	inputs										Outputs				
		NH <sub>4</sub> NO <sub>3</sub> , CaNO <sub>3</sub> , ZnNO <sub>3</sub>	KNO <sub>3</sub>	K <sub>2</sub> SO <sub>4</sub>	MgSO <sub>4</sub> , MnSO <sub>4</sub> , CuSO <sub>4</sub>	KH <sub>2</sub> PO <sub>4</sub> , H <sub>3</sub> BO <sub>3</sub> , Na <sub>2</sub> MoO <sub>4</sub>	FeEDDHA	Thiamine, Nicotinic acid, Glycine	BAP (mg/l)	TDZ, IBA (mg/l)	PR	CW (g)	STN	Vit		
Chandler	Optimized	1.76	1.67	0.96	0.66	2.35	1.64	1.89	0.67	1.30	23.54	0.12	2.31	9.95		
	DKW	1	1	1	1	1	1	1	1	0.1	7.12	0.3	10.75	0		
Rayen	Optimized	1.78	0.85	0.93	1.43	1.95	1.59	0.53	0.99	0.52	7.42	0.70	11.29	12.99		
	DKW	1	1	1	1	1	1	1	1	0.1	6.25	0.5	23.75	16.12		

and predict walnut explant *in vitro* performance. They can also be employed as an alternative technique for linear regression and usual statistical analysis methods with noteworthy performance among them. The KNN model has been used for the first time in this study for predicting plant *in vitro* performance. The optimized models should be applied to predict walnut PR in experimental designs for controlling undesirable physiological disorders.

- All ML models performed accurately for forecasting PR, CW, STN and Vit. Nevertheless, the accuracy of the GEP models were mostly higher than ANN and KNN models. So, the GEP models were selected to be optimized by PSO technique in order to achieve optimal culture media.
- Using above-mentioned ML models is extremely useful for reducing time and cost for formulating efficient walnut tissue culture media.
- The ML-designed media for walnut can not only raise PR (especially about Chandler) but, simultaneously, reduce CW, STN and Vit.
- Genotype is a very important factor which affects the *in vitro* performance and based on our results, it seems that Rayen as a not bred genotype is more recalcitrant to *in vitro* propagation than the bred cultivar Chandler.
- Other factors such as sucrose along with our studied medium components and their interaction on PR and occurrence of physiological disorders also need to be incorporated into the predicting model to control the PR comprehensively.

## Methods

MLR, MLPNN, KNN and GEP modeling techniques were applied to make models using various arrangements of minerals, vitamins and PGRs with different concentrations as inputs and different proliferation indices as outputs. The selected models were used to achieve the optimized models using PSO. Two case studies were done using walnut cultivar Chandler and genotype Rayen which have explained details of the used procedures to understand the optimized inputs combinations as follows.

### Case studies

*In vitro* established nodal cultures of Chandler and Rayen were sub-cultured in altered DKW media supplemented with various auxin and cytokinin PGRs concentrations, 30 g/l sucrose and 3 g/l Gelrite. The media were dispersed into jam jars (250 ml) with polyethylene caps

after adjusting pH to 5.5. Then, the distributed media were autoclaved for 15 min at  $1 \text{ kg cm}^{-2} \text{ s}^{-1}$  (121 °C). The cultures were kept under 16-h white fluorescent ( $80 \mu\text{mol m}^{-2} \text{ s}^{-1}$ ) light at  $25 \pm 2$  °C for 30 days. Subsequently, parameters comprising PR, CW, STN and Vit were measured. In each experiment set, every treatment included 8 replicates (jam jars) for both Chandler and Rayen.

### Taguchi experimental design for optimization of explant proliferation

Taguchi design is a strong and effective tool for the process of optimization that functions constantly and optimally through different conditions. Evaluating numerous factors with limited runs is possible via Taguchi designs i.e. orthogonal arrays. In this design, factors are not weighted more or less in the same experiment and therefore all factors are analyzed independently to each other. Deviation of a product efficient characteristics from their target values is produced by some noise factors such as human errors. Based on orthogonal arrays of Taguchi's, a standard orthogonal array  $L_{27} (3^5)$  27 experiments by  $26^\circ$  of freedom were applied for each of Chandler and Rayen to evaluate the effect of nine factors according to Table 4, on PR, CW, STN, and Vit. For each experiment, three different levels of factor variations were based on various coefficients  $\times$  DKW basal medium nutrients and different PGRs concentrations (Tables 5). Every nutrient and PGR concentration treatment includes at least 8 replicates. 157 experimental sets (70% of data lines) among 224 sets were randomly chosen for training the modeling methods and the rest 67 sets (30% of data lines) were applied for testing the model's generalization capacity. In all ML models, k-fold ( $k=10$ ) cross validation method [87, 88] was used for training to maintain and grantee the generalizability of constructed models.

## Modeling techniques

### Multiple linear regression

MLR analysis is a multivariate statistical method to assess the relationship between multiple independent variables and an individual dependent variable. Two important purposes of MLR are prediction and explanation. The MLR prediction comprises the level to which the independent variables can predict the dependent variables. The mentioned description of MLR estimates the coefficients of regression, their sign, magnitude and statistical interface, for each independent variable [89]. Linear regression is considered as the first statistical method in regression and assumed to be an index technique to be used by new methods. As other regression methods, the relationships between a dependent variable and multiple



**Table 4** The components and levels of factors used for walnut micropropagation and measured traits mean values applied to characterize it

Walnut	Culture medium	Nutrients (× DKW basal medium) and PGRs concentrations										PR	CW (g)	STN	Vit			
		F1	F2	F3	F4	F5	F6	F7	F8	F9								
		NH <sub>4</sub> NO <sub>3</sub> , CaNO <sub>3</sub> , ZnNO <sub>3</sub>	KNO <sub>3</sub>	K <sub>2</sub> SO <sub>4</sub>	MgSO <sub>4</sub> , MnSO <sub>4</sub> , CuSO <sub>4</sub>	KH <sub>2</sub> PO <sub>4</sub> , H <sub>3</sub> BO <sub>3</sub> , Na <sub>2</sub> MoO <sub>4</sub>	FeEDDHA	Thiamine, Nicotinic acid, Glycine	BAP	TDZ, IBA (mg/l)								
Chandler	1	0.5	2	2	2.5	0.5	0.5	0.5	2	0.5	0.5	2	2	9.5±0.18	0.79±0.01	30.5±0.9	17.3±0.32	
	2	1.25	2	0	1.5	1.5	2	0.5	0.5	1.25	0.5	0.5	1.25	8.1±0.44	0.20±0.01	20.3±0.6	11.7±0.36	
	3	2	2	1	0.5	2.5	1.25	0.5	0.5	1.25	0.5	1.25	0.5	6.75±0.2	0.49±0.02	27.2±0.4	13.1±0.44	
	4	2	1	0	2.5	1.5	0.5	2	0.5	0.5	2	1.25	2	10±0.32	0.89±0.05	31.2±1.1	12.8±0.29	
	5	1.25	0	1	2.5	1.5	2	0.5	0.5	1.25	0.5	1.25	2	8.6±0.18	0.97±0.06	33.3±1	21.2±0.45	
	6	2	0	2	1.5	2.5	1.25	0.5	0.5	1.25	0.5	1.25	2	7.3±0.7	1.09±0.06	30.1±0.4	29.1±0.51	
	7	1.25	2	0	1.5	2.5	0.5	1.25	1.25	0.5	1.25	1.25	2	6.2±0.31	0.89±0.01	30.7±0.7	16.5±0.32	
	8	0.5	2	2	2.5	1.5	1.25	0.5	0.5	0.5	0.5	1.25	0.5	10.5±1.1	0.28±0.01	20.2±0.9	0	
	9	0.5	1	1	1.5	0.5	0.5	0.5	0.5	0.5	0.5	1.25	1.25	6.3±0.3	0.5±0.01	21±0.8	0	
	10	2	2	1	0.5	1.5	0.5	2	2	0.5	2	0.5	2	10.7±0.9	0.99±0.03	36.1±1.1	11.5±0.42	
	11	0.5	0	0	0.5	1.5	1.25	1.25	1.25	0.5	1.25	1.25	1.25	11.7±1	0.5±0.01	33.1±0.9	5.5±0.32	
	12	1.25	1	2	0.5	1.5	2	0.5	0.5	2	0.5	2	0.5	10.5±0.3	0.49±0.01	39.3±1.1	10.7±0.36	
	13	1.25	1	2	0.5	1.5	2	0.5	0.5	2	0.5	2	0.5	10±0.32	0.49±0.01	34.3±0.8	6.7±0.35	
	14	0.5	0	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	4.8±0.3	0.24±0.01	20.6±0.7	4.6±0.37	
	15	1.25	0	1	2.5	0.5	1.25	1.25	2	0.5	2	0.5	1.25	7.6±0.18	0.2±0.01	12.6±0.6	9.8±0.39	
	16	0.5	0	0	0.5	2.5	2	2	2	2	2	2	2	5.7±0.31	0.89±0.01	31.6±0.8	27.1±0.45	
	17	2	1	0	2.5	0.5	2	1.25	1.25	2	1.25	1.25	0.5	9.7±0.28	0.13±0.01	13.3±0.3	13.2±0.33	
	18	0.5	1	1	1.5	2.5	2	2	2	2	2	0.5	0.5	6.7±0.25	0.18±0.01	11.3±0.3	3.8±0.29	
	19	1.25	1	2	0.5	0.5	1.25	2	2	1.25	2	1.25	2	8±0.32	0.52±0.01	27.1±0.5	6.8±0.25	
	20	1.25	1	2	0.5	2.5	0.5	0.5	0.5	1.25	1.25	0.5	0.5	7.3±0.41	0.20±0.01	15±0.6	6.8±0.39	
	21	0.5	1	1	1.5	1.5	1.25	1.25	1.25	0.5	1.25	2	2	5.8±0.3	0.73±0.01	22±0.5	9.3±0.32	
	22	1.25	0	1	2.5	0.5	0.5	1.25	1.25	0.5	1.25	2	0.5	9.1±0.29	0.2±0.01	15.7±1	7.1±0.29	
	23	2	2	1	0.5	2.5	2	2	2	2	1.25	1.25	1.25	5.8±0.3	1.1±0.03	21.8±0.6	19.5±0.32	
	24	0.5	2	2	2.5	2.5	2	2	2	2	2	1.25	1.25	21.8±0.7	0.09±0.01	11.1±0.7	3.1±0.12	
	25	2	0	2	1.5	0.5	2	1.25	1.25	2	1.25	0.5	2	13.6±0.3	0.98±0.01	18.1±0.8	10.7±0.45	
	26	2	1	0	2.5	2.5	1.25	0.5	0.5	1.25	0.5	0.5	2	12.1±0.3	0.8±0.02	13.6±0.6	9.7±0.44	
	27	2	0	2	1.5	1.5	0.5	2	2	0.5	2	1.25	0.5	9.7±0.31	0.51±0.01	11.6±0.6	29±0.46	
	DKW	1	1	1	1	1	1	1	1	1	1	1	0.1	7.12±0.35	0.3±0.03	10.75±0.45	0	
Rayen	1	0.5	2	2	2.5	0.5	0.5	0.5	0.5	0.5	0.5	2	2	8.5±0.42	0.92±0.03	33.8±0.8	21.7±0.59	

**Table 4** (continued)

Walnut Culture medium	Nutrients (× DKW basal medium) and PGRs concentrations										PR	CW (g)	STN	Vit
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F9				
2	1.25	2	0	1.5	1.5	2	0.5	0.5	0.5	1.25	7.7±0.31	0.28±0.02	25.6±1.1	14±0.26
3	2	2	1	0.5	2.5	1.25	0.5	0.5	0.5	1.25	6.3±0.37	0.48±0.01	49.5±1.4	15.7±0.36
4	2	1	0	2.5	1.5	0.5	2	2	2	1.25	8±0.37	0.94±0.02	25.3±1.5	16.8±0.29
5	1.25	0	1	2.5	1.5	2	0.5	0.5	0.5	2	7.7±0.31	0.99±0.02	27.2±1.3	27.7±0.45
6	2	0	2	1.5	2.5	1.25	0.5	0.5	0.5	1.25	8±0.32	0.79±0.03	29.6±0.6	31.3±0.54
7	1.25	2	0	1.5	2.5	0.5	1.25	1.25	1.25	2	5.8±0.29	1.07±0.04	31.2±1	17.3±0.37
8	0.5	2	2	2.5	1.5	1.25	1.25	1.25	1.25	0.5	6.3±0.26	0.31±0.01	21.8±1.1	6±0.46
9	0.5	1	1	1.5	0.5	0.5	0.5	0.5	0.5	1.25	6.3±0.46	0.52±0.01	27±0.4	6.2±0.36
10	2	2	1	0.5	1.5	0.5	2	2	2	2	7±0.53	1.05±0.04	33.1±1.2	15±0.37
11	0.5	0	0	0.5	1.5	1.25	1.25	1.25	1.25	1.25	8±0.42	0.45±0.01	28.5±0.4	8±0.26
12	1.25	1	2	0.5	1.5	2	0.5	0.5	0.5	2	8.2±0.49	0.17±0.01	28±0.6	10.6±0.46
13	1.25	1	2	0.5	1.5	2	0.5	0.5	0.5	2	9±0.46	0.39±0.01	27.3±0.9	9.2±0.36
14	0.5	0	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	6.1±0.39	0.27±0.01	23.3±0.5	4±0.26
15	1.25	0	1	2.5	0.5	1.25	2	2	2	1.25	7±0.26	0.29±0.01	21.7±1.3	12.1±0.29
16	0.5	0	0	0.5	2.5	2	2	2	2	2	6±0.46	0.72±0.02	29±1.5	31.8±0.44
17	2	1	0	2.5	0.5	2	1.25	1.25	1.25	0.5	7.3±0.37	0.26±0.01	18.1±1.4	16.1±0.37
18	0.5	1	1	1.5	2.5	2	2	2	2	0.5	6.3±0.53	0.31±0.01	27.2±1.3	5±0.32
19	1.25	1	2	0.5	0.5	1.25	2	2	2	2	7.5±0.5	0.46±0.01	27.8±0.5	7.1±0.29
20	1.25	1	2	0.5	2.5	0.5	1.25	1.25	1.25	1.25	7.5±0.32	0.29±0.01	20.1±0.6	6.7±0.25
21	0.5	1	1	1.5	1.5	1.25	1.25	1.25	1.25	2	6±0.26	0.94±0.02	27.1±0.7	10.7±0.36
22	1.25	0	1	2.5	2.5	0.5	1.25	1.25	1.25	0.5	6.5±0.32	0.16±0.01	22±0.9	6±0.37
23	2	2	1	0.5	0.5	2	1.25	1.25	1.25	2	6.6±0.25	0.91±0.03	27.2±0.9	22.5±0.5
24	0.5	2	2	2.5	2.5	2	2	2	2	1.25	7±0.31	0.17±0.01	27.5±0.6	6±0.26
25	2	0	2	1.5	0.5	2	1.25	1.25	1.25	2	10.2±0.28	0.55±0.02	21.6±0.8	9.3±0.41
26	2	1	0	2.5	2.5	1.25	0.5	0.5	0.5	2	7.2±0.36	0.94±0.03	29.5±0.7	11.6±0.37
27	2	0	2	1.5	1.5	0.5	2	2	2	1.25	8.3±0.37	0.46±0.01	27.3±0.3	35±0.46
DKW	1	1	1	1	1	1	1	1	1	1	6.25±0.4	0.5±0.06	23.75±0.59	16.12±1.1

**Table 5** The components of factors and experimental runs ranges based on DKW medium

Code of Factors	Culture medium components	Level of factors
X1	NH <sub>4</sub> NO <sub>3</sub> , CaNO <sub>3</sub> , ZnNO <sub>3</sub>	0.5–1.25–2×
X2	KNO <sub>3</sub>	0–1–2×
X3	K <sub>2</sub> SO <sub>4</sub>	0–1–2×
X4	MgSO <sub>4</sub> , MnSO <sub>4</sub> , CuSO <sub>4</sub>	0.5–1.5–2.5×
X5	KH <sub>2</sub> PO <sub>4</sub> , H <sub>3</sub> BO <sub>3</sub> , Na <sub>2</sub> MoO <sub>4</sub>	0.5–1.5–2.5×
X6	FeEDDHA	0.5–1.25–2×
X7	Thiamine, Nicotinic acid, Glycine	0.5–1.25–2×
X8	BAP	0.5–1.25–2 mg l <sup>-1</sup>
X9	TDZ, IBA	0.5–1.25–2 mg l <sup>-1</sup>

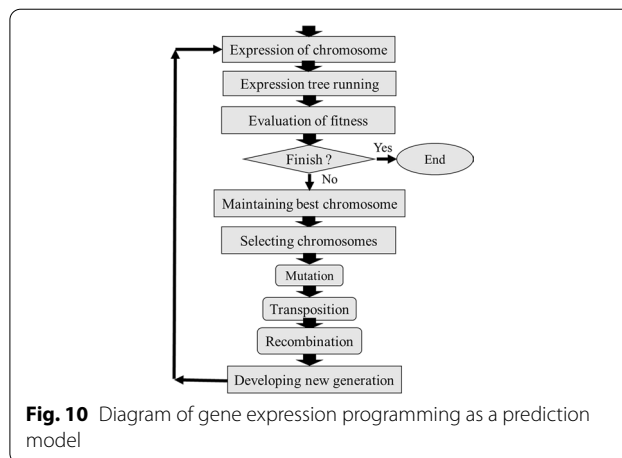
independent variables are modeled by MLR and a linear equation is fitted to the experimental data. MLR technique makes relationship between independent variable *k* value and the dependent variable *M* value. The regression equation of *n* input variables  $\times 1, \times 2, \dots, \times n$  is according to the following:

$$M = \alpha_0 + \alpha_1 k_1 + \dots + \alpha_n$$

in which the dependent variable is *M*, *k* ( $\times 1, \dots, \times n$ ) denotes a vector of input variables,  $\alpha_0$  indicates intercept (a constant), and  $\alpha$  is the coefficient of regression vector, each of which is for each expository variable. *Y* experimental values have various meanings and are supposed with the identical standard deviation  $\epsilon$ . The SPSS 19 software package was used for the MLR modeling.

**Multilayer perceptron neural network**

The neural network is divided into various types based on the transfer functions basis. In the present study, we used multi-layer perceptron (MLPNN) network. The MLPNN model is the most common and widely used type of artificial neural network [90]. This model generally contains an input layer and an output layer. One or more hidden layers can be placed between these two layers. Each neuron in this structure has a number of inputs and a number of outputs. A neuron calculates its output responses based on the weighted sum of all its inputs, performed by a stimulus or transmission function. In the MLPNN model, starting from the input information in the first layer (independent variables), the information flows in only one direction and enters the output layer (dependent variable) by transferring from the hidden layer. The training process of MLPNN model involves adjusting and modifying the weights of the interface between neurons using different network training methods [91]. In this study, Broyden-Fletcher-Goldfarb-Shanno (BFGS) training algorithm has been used. Also, stimulus functions;



**Fig. 10** Diagram of gene expression programming as a prediction model

The tangent hyperbolic (Tanh), sigmoid function (Logs), exponential function (Exp), relu function (Relu) in the hidden layer and linear function (Idn) in the output layer were compared and evaluated and the best function was selected. The number of hidden layers was also determined by trial and error by reaching the minimum error rate. See [91, 92] for more information.

**k-nearest neighbors’ algorithm**

The *k*-nearest neighbors (KNN) model is a non-supervised learning machine algorithm for data classification. In this model, each data represents a coordinate position in a vector-space model that the information of each particular section must have similar properties as well as be close to each other. In the KNN algorithm, determining the number of neighbors (*k*) as well as the method based on which the distance between them is calculated is of particular importance. If *k* is considered too small, then neighboring points that do not appear in the classification will reduce the accuracy of the results. On the other hand, if *k* is considered too large, the results of the same classifications may be merged as the computational volume increases [93]. The nearest neighbor was evaluated and selected from different values to find the best value of *k* and to achieve the highest model accuracy. Distances between neighboring points were determined using various geometric methods. In this study, the methods of Euclidean Distance, Chebyshev Distance, Manhattan Distance and Minkowski Distance were studied and the best method was selected.

**Gene expression programming**

GP is a modeling approach used to model the structural engineering complications behavior. It is an extension of genetic algorithm that utilizes a program space for

searching, rather than using a data space. An important benefit of applying GP-based techniques toward other methods is their capability to produce equations of prediction without using any hypothesis for previous relationship form. Many researchers have applied GP and GP-based methods to find any complicated relationships fitting different experimental data [44, 94, 95]. GEP has been introduced as an effective substitute method to the conventional GP [31, 46]. GEP have established many computer programs, by getting encoded in linear chromosomes with constant length, each of which included several encoding genes [31, 96]. GEP is originated of evolutionary algorithms such as GA and GP. In this technique, an individual population is applied and afterwards, fitness function and genetic variations are used to select better individuals. The genetic variations are presented by genetic operators. GEP is a learning machine which is assumed to learn the variables relationship in datasets. The individual programming technique is different in GEP and its predecessors GP and GA since GEP programs individuals as linear strings (chromosomes) with fixed length which are finally displayed by expression trees as unsophisticated diagram. While, GP and GA express individuals in the form of linear strings (chromosomes) with fixed length and nonlinear entities of diverse forms (parse trees) and dimensions, respectively. One of the strongpoints of GEP towards GP and GA is that genetic operators run very easily at the level of chromosome in GEP producing genetic diversity creation. Another strength of GEP is its exclusive, multi-genic nature letting more complicated programs with numerous subprograms to be developed. Both GP and GA advantages are collected in GEP, whereas some of their constraints are met [57, 97, 98].

The real GEP chromosome phenotype is the illustration in Fig. 10 and the genotype would be simply described of the phenotype as represented in Eq. (1)

$$(a + b) * (c - d) \tag{1}$$

Functional steps of the GEP are represented in Fig. 10 [31]. According to this diagram, the GEP start point is a population of chromosomes. After that, the chromosomes genes are expressed, and each individual fitness is analyzed. Then, the individuals are defined according to their fitness to reproduce with alteration. The same development process is run on the new individuals' generation. Overall, this technique is replicated for a particular number of generations or it is performed until reaching a termination condition. Roulette wheel sampling with elitism is employed by GEP system to ensure that the top individuals, according to the fitness, are remained and copied to the next generation. Once genetic operator(s) are performed on chosen chromosomes, comprising mutation, cross over and rotation, diversity is developed into the population.

The GeneXpro software package was applied to perform the GEP models. The parameters employed in the GEP models are represented in Table 6.

In this study, the selected functions and mathematical operators are rational and not definite so that the plant modeling designer is free to select such functions according to the studied problem anatomy. The functions and operators were selected with a viewpoint of invocating simpleness of the advanced model assuring quicker convergence. The size of the population (chromosomes number) adjusts the programs number into the population. The larger the population, the longer it takes for an iteration run. High chromosomes number were tried to

**Table 6** Parameters of training GEP model

Parameter	Parameter description	Parameter setting
1	Function set	+, -, ÷, *, sin, cos, x^2, x^3, power,
2	Chromosomes	50
3	Head size	10
4	Number of genes	4
5	Linking function	addition
6	Fitness function	Root relative square error (RRSE)
7	Mutation rate	0.044
8	Inversion rate	0.1
9	One-point recombination rate	0.1
10	Two-point recombination rate	0.3
11	Gene recombination rate	0.1
12	Gene transportation rete	0.1

realize minimum error models. The program running continued to reach no significant rectification in the models' performance. Here, we aimed to achieve obvious relationship between decision variables and response variables. GEP clear formulations were obtained for PR, CW, STN and Vit as a function of experimental parameters including Y1, Y2, Y3 and  $Y4 = f(X1, X2, X3, X4, X5, X6, X7, X8 \text{ and } X9)$  (Table 1).

Input data were normalized in the range of 0 and 1 according to the Eq. 2:

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (2)$$

where  $X_n$  is normalized dimensionless data,  $X_i$  is observed data,  $X_{min}$  is the minimum amount of observed data, and  $X_{max}$  is its maximum value.

### Comparison of the performance of developed models

To evaluate the precision of created models, we used different statistical indices including coefficient of determination ( $R^2$ ), root mean square error (RMSE) and mean absolute error (MAE) based on Eqs. 5, 6 and 7:

$$R^2 = \frac{\left(\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right)^2}{\sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (6)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (7)$$

where  $y$  and  $\bar{y}$  are observed values and their mean and  $\hat{y}$  and  $\bar{\hat{y}}$  are predicted values and their mean, respectively, as well for  $N$  samples. Analyses of parameters were performed together to achieve an accurate medium composition.

In addition, the predicted values by the developed models were plotted against the corresponding observed values to evaluate the ability of models for prediction.

### Particle swarm optimization of GEP models

PSO is a method of evolutionary calculation and swarm intelligence algorithm according to population to solve the pervasive problem of optimization that was developed by [99]. It is a method of mathematical computation that starts with the swarm (a population of grain) and mostly based on social models, such as the swarm theory, fish schooling and bird flocking [20]. PSO key factors are with behavior of swarm i.e. keeping optimum distances between different members and their

neighbors. To optimize each particle location, their position is modified as arranged for the objective function within the search area. Thus, PSO key factor is a particle velocity which is compared to the previous one in each repetition to lead the particle to its optimal position. The best solution (fitness) every particle in a swarm achieves so far in each repetition, named pbest. Extra "best" value that a particle is attained in the population up to now followed by the particle swarm optimizer which is global best, named gbest. Each particle velocity in a swarm is estimated by Eqs. 3 and 4 [99].

$$V_{i+1} = wV_i + c_1r_1(pBest_i - X_i) + c_2r_2(gBest_i - X_i) \quad (3)$$

$$X_{i+1} = X_i + V_{i+1} \quad (4)$$

in which,  $V_{i+1}$  is each particle new velocity based on prior velocity ( $V_i$ ),  $w$  is inertial coefficient (0.8–1.2),  $c_1$  and  $c_2$  are cognitive and social coefficients, respectively (0–2),  $r_1$  and  $r_2$  are random values for each velocity update (0–1) and  $X_{i+1}$  is new location for each particle according to the prior location ( $X_i$ ).

### Abbreviations

PGRs: Plant growth regulators; STN: Shoot tip necrosis; Vit: Vitrification; MLL: Machine learning; MLPNN: Multi-layer perceptron neural network; KNN: k-Nearest neighbors; GEP: Gene expression programming; MLR: Multiple linear regression; PSO: Particle swarm optimization; DKW: Driver and Kuniyuki walnut; ANN: Artificial neural network; GA: Genetic algorithm; RBFNN: Radial basis function neural network; GP: Genetic programming; MOEA: Multi-objective evolutionary optimization algorithm; MOPs: Multi-objective optimization problems; PR: Proliferation rate; CW: Callus weight; MS: Murashige and Skoog.

### Acknowledgements

We also would like to thank University of Jiroft, University of Tehran, Tarbiat Modares University (TMU), Iran National Science Foundation (INSF), Iran's National Elite Foundation (INEF) and Center of Excellence for Walnut Improvement and Technology of Iran, for their supports.

### Authors' contributions

MSH: performing the experiments, gathering data, interpreting data and revising the manuscript. MMA: designing the experiments, interpreting data, summing up and revising the manuscript. MS: statistical analyzing. ME: interpreting experiments, writing and revising the manuscript. AS: performing experiments. KV: revising manuscript. All authors listed, had direct, considerable and intellectual contribution to the research. All authors read and approved the final manuscript.

### Funding

No funding was received.

### Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Horticulture, Faculty of Agriculture, University of Jiroft, Jiroft, Iran. <sup>2</sup>Department of Horticulture, College of Aburaihan, University of Tehran, Tehran, Iran. <sup>3</sup>Department of Irrigation and Drainage Engineering, College of Aburaihan, University of Tehran, Tehran, Iran. <sup>4</sup>Department of Horticultural Science, Faculty of Agriculture, Tarbiat Modares University (TMU), Tehran, Iran. <sup>5</sup>Department of Agronomy and Plant Breeding, Faculty of Agriculture, University of Jiroft, Jiroft, Iran.

Received: 16 January 2022 Accepted: 7 March 2022

Published online: 11 April 2022

### References

- McGranahan G, Leslie C. Walnut. In: Fruit breeding. Springer: Boston; 2012. p. 827–846.
- Skender A, Kurtovic M, Drkenda P, Becirspahic D, Ebrahimi A. Phenotypic variability of autochthonous walnut (*Juglans regia* L) genotypes in north-western Bosnia and Herzegovina. *Turk J Agric For*. 2020;44(5):517–25.
- Driver JA, Kuniyuki AH. In vitro propagation of Paradox walnut rootstock. *HortScience*. 1984;19(4):507–9.
- McGranahan GH, Driver JA, Tulecke W. Tissue culture of *Juglans*. In: Cell and tissue culture in forestry; 1987. Springer, Dordrecht. p. 261–271.
- Cornu D, Jay-Allemand C. Micropropagation of hybrid walnut trees (*Juglans nigra* x *Juglans regia*) through culture and multiplication of embryos. In: *Annales des sciences forestières* 1989 (Vol. 46, No. Supplement, pp. 113s-116s). EDP Sciences.
- Pijut PM. Micropropagation of *Juglans cinerea* L. (butternut). In: *High-Tech and Micropropagation V 1997*; Springer, Berlin. p. 345–357.
- Bourrain L, Navatel JC. Plant production of walnut *Juglans regia* L. by in vitro multiplication. In: *IV International Walnut Symposium* 544 1999. p. 465–471.
- Vahdati K, Jariteh M, Niknam V, Mirmasoumi M, Ebrahimzadeh H. Somatic embryogenesis and embryo maturation in Persian walnut. In *V International Walnut Symposium*. 2004;705:199–205.
- Vahdati K, Najafian Ashrafi E, Ebrahimzadeh H, Mirmasoumi M. Improved micropropagation of walnut (*Juglans regia* L.) on media optimized for growth based upon mineral content of walnut seed. In: *I International Symposium on Biotechnology of Fruit Species: BIOTECHFRUIT2008* 839 200. p. 117–124.
- Gotea R, Gotea I, Sestras RE, Vahdati K. In vitro Propagation of Several Walnut Cultivars. In: *Bulletin of the University of Agricultural Sciences & Veterinary Medicine Cluj-Napoca. Horticulture*. 2012. p. 69.
- Licea-Moreno RJ, Contreras A, Morales AV, Urban I, Daquinta M, Gomez L. Improved walnut mass micropropagation through the combined use of phloroglucinol and FeEDDHA. *Plant Cell Tissue Organ Culture*. 2015;123(1):143–54.
- Gruselle R, Boxus P. Walnut micropropagation. *Int Symp Walnut Production*. 1989;284:45–52.
- Lone IA, Misger FA, Banday FA. Effect of different growth regulator combinations on the per cent media browning in walnut in vitro studies using MS medium. *Asian J Soil Sci*. 2017;12(1):135–42.
- Sadat-Hosseini M, Vahdati K, Leslie CA. Germination of Persian walnut somatic embryos and evaluation of their genetic stability by ISSR fingerprinting and flow cytometry. *HortScience*. 2019;54(9):1576–80.
- Yegizbayeva TK, García-García S, Yausheva TV, Kairova M, Apushev AK, Oleichenko SN, Licea-Moreno RJ. Unraveling Factors Affecting Micropropagation of Four Persian Walnut Varieties. *Agronomy*. 2021;11(7):1417.
- Gago J, Martínez-Núñez L, Landín M, Gallego PP. Artificial neural networks as an alternative to the traditional statistical methodology in plant research. *J Plant Physiol*. 2010;167(1):23–7.
- Gallego PP, Gago J, Landín M. Artificial neural networks technology to model and predict plant biology process. In: Suzuki K. *Artificial Neural Networks-Methodological and Biomedical Applications* (Croatia: Intech Open Access Publisher). 2011. p. 197–216.
- Barone JO. Use of multiple regression analysis and artificial neural networks to model the effect of nitrogen in the organogenesis of *Pinus taeda* L. *Plant Cell Tissue Organ Culture*. 2019;137(3):455–64.
- Hesami M, Naderi R, Tohidfar M. Modeling and optimizing medium composition for shoot regeneration of *Chrysanthemum* via radial basis function-non-dominated sorting genetic algorithm-II (RBF-NSGAI). *Sci Rep*. 2019;9(1):1–1.
- Jamshidi S, Yadollahi A, Arab MM, Soltani M, Eftekhari M, Sabzalipoor H, Sheikhi A, Shiri J. Combining gene expression programming and genetic algorithm as a powerful hybrid modeling approach for pear rootstocks tissue culture media formulation. *Plant Methods*. 2019;15(1):1–8.
- García-Pérez P, Lozano-Milo E, Landin M, Gallego PP. Machine Learning unmasked nutritional imbalances on the medicinal plant *Bryophyllum* sp. cultured in vitro. *Front Plant Sci*. 2020;11:1887.
- Salehi M, Farhadi S, Moieni A, Safaie N, Ahmadi H. Mathematical modeling of growth and paclitaxel biosynthesis in *Corylus avellana* cell culture responding to fungal elicitors using multilayer perceptron-genetic algorithm. *Front Plant Sci*. 2020;11:114.
- Hesami M, Naderi R, Tohidfar M. Introducing a hybrid artificial intelligence method for high-throughput modeling and optimizing plant tissue culture processes: The establishment of a new embryogenesis medium for chrysanthemum, as a case study. *Appl Microbiol Biotechnol*. 2020;104(23):10249–63.
- Jamshidi S, Yadollahi A, Ahmadi H, Arab MM, Eftekhari M. Predicting in vitro culture medium macro-nutrients composition for pear rootstocks using regression analysis and neural network models. *Front Plant Sci*. 2016;29(7):274.
- Arab MM, Yadollahi A, Shojaeiyan A, Ahmadi H. Artificial neural network genetic algorithm as powerful tool to predict and optimize in vitro proliferation mineral medium for Gx N15 rootstock. *Front Plant Sci*. 2016;19(7):1526.
- Arab MM, Yadollahi A, Ahmadi H, Eftekhari M, Maleki M. Mathematical modeling and optimizing of in vitro hormonal combination for Gx N15 vegetative rootstock proliferation using Artificial Neural Network-Genetic Algorithm (ANN-GA). *Front Plant Sci*. 2017;1(8):1853.
- Arab MM, Yadollahi A, Eftekhari M, Ahmadi H, Akbari M, Khorami SS. Modeling and optimizing a new culture medium for in vitro rooting of Gx N15 *Prunus* rootstock using artificial neural network-genetic algorithm. *Sci Rep*. 2018;8(1):1–8.
- Jamshidi S, Yadollahi A, Arab MM, Soltani M, Eftekhari M, Shiri J. High throughput mathematical modeling and multi-objective evolutionary algorithms for plant tissue culture media formulation: Case study of pear rootstocks. *PLoS ONE*. 2020;15(12):e0243940.
- Kosiński W, Prokopowicz P, Ślęzak D. Fuzzy reals with algebraic operations: Algorithmic approach. In *Intelligent Information Systems 2002*. Heidelberg: Physica; 2002. p. 311–20.
- Albiol J, Campmajó C, Casas C, Poch M. Biomass estimation in plant cell cultures: a neural network approach. *Biotechnol Prog*. 1995;11(1):88–92.
- Ferreira C. Gene expression programming: a new adaptive algorithm for solving problems. *arXiv preprint cs/0102027*. 2001.
- Chu Y, Urquhart B, Gohari SM, Pedro HT, Kleissl J, Coimbra CF. Short-term reforecasting of power output from a 48 MWe solar PV plant. *Sol Energy*. 2015;1(112):68–77.
- Tongal H. Nonlinear forecasting of stream flows using a chaotic approach and artificial neural networks. *Earth Sci Res J*. 2013;17(2):119–26.
- Antonio LM, Coello CA. Use of cooperative coevolution for solving large scale multiobjective optimization problems. In: *2013 IEEE Congress on Evolutionary Computation 2013*. p. 2758–2765.
- Durillo JJ, Nebro AJ, Coello CA, García-Nieto J, Luna F, Alba E. A study of multiobjective metaheuristics when solving parameter scalable problems. *IEEE Trans Evol Comput*. 2010;14(4):618–35.
- Murashige T, Skoog F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol Plant*. 1962;15(3):473–97.
- Bosela MJ, Michler CH. Media effects on black walnut (*Juglans nigra* L) shoot culture growth in vitro: evaluation of multiple nutrient formulations and cytokinin types. *Vitro Cell Develop Biol-Plant*. 2008;44(4):316–29.
- Kepenek KA, Kolağasi Z. Micropropagation of Walnut (*Juglans regia* L). *Acta Phys Pol, A*. 2016;130:1.

39. Stevens ME, Pijut PM. Rapid in vitro shoot multiplication of the recalcitrant species *Juglans nigra* L. *In Vitro Cell Develop Biol-Plant*. 2018;54(3):309–17.
40. Hassankhah A, Vahdati K, Lotfi M, Mirmasoumi M, Preece J, Assareh MH. Effects of ventilation and sucrose concentrations on the growth and plantlet anatomy of micropropagated Persian walnut plants. *Int J Horticult Sci Technol*. 2014;1(2):111–20.
41. Ashrafi EN, Vahdati K, Ebrahimzadeh H, Mirmasoumi M. Analysis of in-vitro explants mineral contents to modify medium mineral composition for enhancing growth of Persian walnut (*Juglans regia* L.). *J Food Agric Environ*. 2010;8:325–9.
42. Eftekhari M, Yadollahi A, Ahmadi H, Shojaeiyan A, Ayyari M. Development of an artificial neural network as a tool for predicting the targeted phenolic profile of grapevine (*Vitis vinifera*) foliar wastes. *Front Plant Sci*. 2018;19(9):837.
43. Sheikhi A, Mirdehghan SH, Arab MM, Eftekhari M, Ahmadi H, Jamshidi S, Gheysarbigi S. Novel organic-based postharvest sanitizer formulation using Box Behnken design and mathematical modeling approach: A case study of fresh pistachio storage under modified atmosphere packaging. *Postharvest Biol Technol*. 2020;160:111047.
44. Gandomi AH, Alavi AH. A new multi-gene genetic programming approach to nonlinear system modeling Part I: materials and structural engineering problems. *Neural Computing Appl*. 2012;21(1):171–87.
45. Gandomi AH, Alavi AH, Arjmandi P, Aghaeifar A, Seyednour R. Genetic programming and orthogonal least squares: a hybrid approach to modeling the compressive strength of CFRP-confined concrete cylinders. *J Mech Mater Struct*. 2010;5(5):735–53.
46. Gandomi AH, Roke DA. Assessment of artificial neural network and genetic programming as predictive tools. *Adv Eng Softw*. 2015;1(88):63–72.
47. De Giorgi MG, Congedo PM, Malvoni M, Laforgia D. Error analysis of hybrid photovoltaic power forecasting models: a case study of mediterranean climate. *Energy Convers Manage*. 2015;1(100):117–30.
48. Kannangara M, Dua R, Ahmadi L, Bensebaa F. Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. *Waste Manage*. 2018;1(74):3–15.
49. Bagheri M, Bazvand A, Ehteshami M. Application of artificial intelligence for the management of landfill leachate penetration into groundwater, and assessment of its environmental impacts. *J Clean Prod*. 2017;15(149):784–96.
50. Quiros AR, Bedruz RA, Uy AC, Abad A, Bandala A, Dadios EP, Fernando A. A KNN-based approach for the machine vision of character recognition of license plate numbers. *InTENCON 2017–2017 IEEE Region 10 Conference 2017*. p. 1081–1086.
51. Mahmoudzadeh H, Matinfar HR, Taghizadeh-Mehrjardi R, Kerry R. Spatial prediction of soil organic carbon using machine learning techniques in western Iran. *Geoderma Reg*. 2020;21:e00260.
52. Dikmen E. Gene expression programming strategy for estimation performance of LiBr–H<sub>2</sub>O absorption cooling system. *Neural Comput Appl*. 2015;26(2):409–15.
53. Nazari A. RETRACTED ARTICLE: Application of gene expression programming to predict the compressive damage of lightweight aluminosilicate geopolymer. *Neural Comput Appl*. 2019;31(2):767–76.
54. Esha R, Imteaz MA. Pioneer use of gene expression programming for predicting seasonal streamflow in Australia using large scale climate drivers. *Ecohydrology*. 2020;13(8):e2242.
55. Zhong J, Feng L, Ong YS. Gene expression programming: A survey. *IEEE Comput Intell Mag*. 2017;12(3):54–72.
56. Xu L, Huang Y, Shen X, Liu Y. Parallelizing gene expression programming algorithm in enabling large-scale classification. *Sci Program*. 2017;20:2017.
57. Ferreira C. Gene expression programming: mathematical modeling by an artificial intelligence. New York: Springer; 2006.
58. Nas MN, Read PE. A hypothesis for the development of a defined tissue culture medium of higher plants and micropropagation of hazelnuts. *Sci Hortic*. 2004;101(1–2):189–200.
59. Gago J, Landín M, Gallego PP. A neurofuzzy logic approach for modeling plant processes: A practical case of in vitro direct rooting and acclimatization of *Vitis vinifera* L. *Plant Sci*. 2010;179(3):241–9.
60. Gago J, Landín M, Gallego PP. Artificial neural networks modeling the in vitro rhizogenesis and acclimatization of *Vitis vinifera* L. *J Plant Physiol*. 2010;167(15):1226–31.
61. Gago J, Pérez-Tornero O, Landín M, Burgos L, Gallego PP. Improving knowledge of plant tissue culture and media formulation by neurofuzzy logic: a practical case of data mining using apricot databases. *J Plant Physiol*. 2011;168(15):1858–65.
62. Niedz RP, Hyndman SE, Evens TJ, Weathersbee AA. Mineral nutrition and in vitro growth of *Gerbera hybrida* (Asteraceae). *In Vitro Cell Develop Biol-Plant*. 2014;50(4):458–70.
63. Nezami-Alanagh E, Garoosi GA, Landín M, Gallego PP. Combining DOE with neurofuzzy logic for healthy mineral nutrition of pistachio rootstocks in vitro culture. *Front Plant Sci*. 2018;9:1474.
64. Nezami-Alanagh E, Garoosi GA, Haddad R, Maleki S, Landín M, Gallego PP. Design of tissue culture media for efficient *Prunus* rootstock micropropagation using artificial intelligence models. *Plant Cell Tissue Organ Culture (PCTOC)*. 2014;117(3):349–59.
65. Reed M, Sugae W, Randall P. Niedz & Barbara. *In Vitro Cell Dev Biol-Plant*. 2015;51:19–27.
66. Nezami-Alanagh E, Garoosi GA, Landín M, Gallego PP. Computer-based tools provide new insight into the key factors that cause physiological disorders of pistachio rootstocks cultured in vitro. *Sci Reports*. 2019;9(1):1–5.
67. Plant IV. Barbara M. Reed, Sugae Wada, Jeanine DeNoma & Randall P. Niedz.
68. Nezami-Alanagh E, Garoosi GA, Maleki S, Landín M, Gallego PP. Predicting optimal in vitro culture medium for *Pistacia vera* micropropagation using neural networks models. *Plant Cell Tissue Organ Culture* 2017;129(1):19–33.
69. Niedz RP, Evens TJ. A solution to the problem of ion confounding in experimental biology. *Nat Methods*. 2006;3(6):417.
70. Vanstraelen M, Benková E. Hormonal interactions in the regulation of plant development. *Annu Rev Cell Dev Biol*. 2012;10(28):463–87.
71. Schaller GE, Bishopp A, Kieber JJ. The yin-yang of hormones: cytokinin and auxin interactions in plant development. *Plant Cell*. 2015;27(1):44–63.
72. George EF, Hall MA, De Klerk GJ. Plant growth regulators I: Introduction; auxins, their analogues and inhibitors. In: *Plant propagation by tissue culture 2008* (pp. 175–204). Springer, Dordrecht.
73. Pasternak T, Miskolczi P, Ayaydin F, Mészáros T, Dudits D, Fehér A. Exogenous auxin and cytokinin dependent activation of CDKs and cell division in leaf protoplast-derived cells of alfalfa. *Plant Growth Regul*. 2000;32(2):129–41.
74. Hepaksoy S. Investigations on micropropagation of some cherry rootstocks I shoot development and proliferation. *Ziraat Fakültesi Dergisi*. 2004;41(3):11.
75. Ružić D, Vujović T, Nikolić D, Cerović R. In vitro growth responses of the 'Pyrodwarf' pear rootstock to cytokinin types. *Romanian Biotechnological Letters*. 2011;16(5):6631.
76. Molassiotis AN, Dimassi K, Therios I, Diamantidis G. Fe-EDDHA promotes rooting of rootstock GF-677 (*Prunus amygdalus* × *P. persica*) explants in vitro. *Biol Plant*. 2003;47(1):141–4.
77. Chalupa V. Clonal propagation of broad-leaved forest trees in vitro. *Communications Institut Forestalis Cechosloveniae*. 1982;12:9.
78. Revilla MA, Majada J, Rodríguez R. Walnut (*Juglans regia* L.) micropropagation. In *Annales des Sciences Forestières 1989* (Vol. 46, No. Supplement, pp. 149s-151s). EDP Sciences.
79. Jay-Allemand C, Capelli P, Cornu D. Root development of in vitro hybrid walnut microcuttings in a vermiculite-containing gelrite medium. *Sci Hortic*. 1992;51(3–4):335–42.
80. Heloir MC, Kevers C, Hausman JF, Gaspar T. Changes in the concentrations of auxins and polyamines during rooting of in-vitro-propagated walnut shoots. *Tree Physiol*. 1996;16(5):515–9.
81. Saadat YA, Hennerty MJ. The effects of different in vitro and ex vitro treatments on the rooting performance of Persian walnut (*Juglans regia* L.) microshoots. In: *IV International Walnut Symposium 544 1999*. p. 473–480.
82. Yang Z, Lüdders P. In vitro propagation of Pistachio (*Pistacia vera* L.). *Gartenbauwissenschaft*. 1994;59(1):30–4.
83. Onay A. In Vitro organogenesis and embryogenesis of pistachio, *Pistacia vera* L. (Doctoral dissertation, University of Edinburgh).

84. Poothong S, Reed BM. Optimizing shoot culture media for *Rubus* germplasm: the effects of NH<sub>4</sub><sup>+</sup>, NO<sub>3</sub><sup>-</sup>, and total nitrogen. *In Vitro Cell Devel Biol-Plant*. 2016;52(3):265–75.
85. Kovalchuk IY, Mukhitdinova Z, Turdiyev T, Madiyeva G, Akin M, Eyduran E, Reed BM. Modeling some mineral nutrient requirements for micropropagated wild apricot shoot cultures. *Plant Cell Tissue Organ Culture*. 2017;129(2):325–35.
86. Akin M, Eyduran E, Niedz RP, Reed BM. Developing hazelnut tissue culture medium free of ion confounding. *Plant Cell Tissue Organ Culture*. 2017;130(3):483–94.
87. Fushiki T. Estimation of prediction error by using K-fold cross-validation. *Stat Comput*. 2011;21(2):137–46.
88. Shiri J, Marti P, Karimi S, Landeras G. Data splitting strategies for improving data driven models for reference evapotranspiration estimation among similar stations. *Comput Electron Agric*. 2019;1(162):70–81.
89. Hair Jr JF, Black WC, Babin BJ, Anderson RE. *Multivariate Data Analysis A Global Perspective*. New Jersey: Person Education.
90. Moghaddamnia A, Gousheh MG, Piri J, Amin S, Han D. Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. *Adv Water Resour*. 2009;32(1):88–97.
91. Hassanpour Kashani M, Montaseri M, Lotfollahi Yaghin MA. Flood estimation at ungauged sites using a new hybrid model. *J Appl Sci*. 2008;8(9):1744–9.
92. ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. Artificial neural networks in hydrology. II: Hydrologic applications. *J Hydrol Eng*. 2000;5(2):124–37.
93. Wei CC. Predictions of surface solar radiation on tilted solar panels using machine learning models: A case study of Tainan city. *Taiwan Energies*. 2017;10(10):1660.
94. Dey P, Das AK. A utilization of GEP (gene expression programming) metamodel and PSO (particle swarm optimization) tool to predict and optimize the forced convection around a cylinder. *Energy*. 2016;15(95):447–58.
95. Gholampour A, Gandomi AH, Ozbakkaloglu T. New formulations for mechanical properties of recycled aggregate concrete using gene expression programming. *Constr Build Mater*. 2017;15(130):122–45.
96. Mitchell M. *An introduction to genetic algorithms*. New York: MIT press; 1998.
97. Ferreira C. Gene expression programming in problem solving. In: *Soft computing and industry*; 2002. Springer, London. pp. 635–653
98. Ferreira SC, Bruns RE, Ferreira HS, Matos GD, David JM, Brandão GC, da Silva EP, Portugal LA, Dos Reis PS, Souza AS, Dos Santos WN. Box-Behnken design: an alternative for the optimization of analytical methods. *Anal Chim Acta*. 2007;597(2):179–86.
99. Kennedy J, Eberhart R. Particle swarm optimization. In: *Proceedings of ICNN'95-international conference on neural networks 1995*; Vol. 4, p. 1942–1948.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

