



Light-emitting diodes induced *in vitro* regeneration of *Alternanthera reineckii* mini and validation via machine learning algorithms

Muhammad Aasim¹ · Seyid Amjad Ali² · Pınar Bekiş¹ · Muhammad Azhar Nadeem¹

Received: 8 August 2022 / Accepted: 10 October 2022 / Published online: 1 November 2022 / Editor: Yong Eui Choi
© The Society for In Vitro Biology 2022

Abstract

Optimization of *in vitro* regeneration protocol using multiple input variables is highly significant, and can be achieved by validating the data using machine learning algorithms. Shoot tip and nodal segment explants of *Alternanthera reineckii* mini were inoculated on Murashige and Skoog (MS) medium enriched with different concentrations of benzylaminopurine (BAP), and cultured under five different monochromatic light-emitting diodes (LEDs). The attained results were validated through the application of four different supervised machine learning models (RF, XGBoost, KNN, and GP). The prediction of the data were validated by using regression coefficient (R^2), mean squared error (MSE), and mean absolute percentage error (MAPE) performance metrics. Results revealed R^2 values of 0.61 and 0.59 for shoot counts and shoot length, respectively. The results of MSE were registered between 3.48–5.42 for shoot count and 0.40–0.74 for shoot length, whereas, 28.9–35.1% and 13.2–18.4% MAPE values were recorded for both shoot count and shoot length. Among the utilized models, the RF model validated and predicted the results more accurately, followed by the XGBoost model for both output variables. The results confirm that ML models can be used for data validation, and opens a new era of employing ML modeling in plant tissue culture of other economically important plants.

Keywords *Alternanthera reineckii* · Aquatic plant · Supervised machine learning · ANOVA

Introduction

Aquatic plants are known as hydrophytes as they spend their whole life cycle or some part of their life cycle in water (Saini *et al.* 2010). These aquatic plants are the primary producers of the aquatic ecosystem by releasing oxygen into the water (Aasim *et al.* 2019). They provide shelter, an ideal environment to lay eggs for fish, and a good food source for aquatic animals due to their richness in vitamins, protein, carbohydrates, and minerals. These plants are highly nutritive and consequently are considered an alternative food source for humans (Aasim *et al.* 2018a). Some of the aquatic plants are also used as medicinal plants in folk medicines (Gonsalves 2010). In recent years, an increase in indoor aquariums and outdoor ponds also increased the demand

for aquatic plants. *Alternanthera reineckii* mini is an important ornamental aquatic plant with variable leaf colors ranging from green to red, and pink. It is the dwarf version of *Alternanthera reineckii* and is ideal for nano-planting.

The commercial propagation of aquatic plants can be achieved either by classical methods or with the aid of a modern biotechnological approach of plant tissue culture. The optimization of *in vitro* regeneration protocol requires different approaches like the selection of proper explants, plant growth regulators, and culture conditions like lighting source, *etc.* (Al-Tanbouz and Abu-Qaoud 2016; Aasim *et al.* 2018a). The lighting system in plant tissue culture regulates the whole *in vitro* organogenesis process (Sotthikul *et al.* 2015). In general, growth rooms or growth cabins are equipped with fluorescent lamps, and light-emitting diodes (LEDs) are also employed commercially for *in vitro* organogenesis (Bello-Bello *et al.* 2016). The most commonly used LEDs include the cool white, daylight, and red and blue LEDs used as single lighting sources or in different combinations (Karataş *et al.* 2016; Aasim *et al.* 2018b). Likewise, the explant is another significant factor that regulates *in vitro* regeneration (Karataş *et al.* 2016).

✉ Muhammad Aasim
mshazim@gmail.com

¹ Department of Plant Protection, Faculty of Agricultural Sciences and Technology, Sivas University of Science and Technology, Sivas, Turkey

² Department of Information Systems and Technologies, Bilkent University, Ankara, Turkey

Machine learning modeling is a sub-branch of artificial intelligence that uses computer-based programs for learning and extracting the best possible knowledge from the given dataset for making predictions or intelligent decisions. The entire ML process is comprised of three major parts: (i) data input, (ii) modeling, and (iii) generalization. The major application of ML-based modeling is to solve complex problems, and its application in the field of precision agriculture has already been established (Sharma *et al.* 2020). The ML algorithms are generally categorized into three major sub-classes of (i) supervised learning, (ii) unsupervised learning, and (iii) reinforcement learning. Among these algorithms, supervised ML works with the labeled dataset by generating input–output relationships for predicting the outputs for unseen inputs (Hesami *et al.* 2021). The prediction can be done by using either classification algorithms for categorical values or regression-based algorithms for quantitative values (Sharma *et al.* 2020). Application of ML algorithms for precision agriculture and commercial farming is prevailing for better prediction, whereas ML algorithms in plant tissue culture studies are also gaining popularity among researchers to optimize their results or protocols more accurately. The use of different ML models for *in vitro* sterilization (Hesami *et al.* 2019), *in vitro* germination and phenotype analysis (Hesami *et al.* 2021; Aasim *et al.* 2022b), *in vitro* organogenesis (Aasim *et al.* 2022a; Kirtis *et al.* 2022), embryogenesis (Hesami *et al.* 2020), cell culture (Farhadi *et al.* 2020), *in vitro* elicitation (Salehi *et al.* 2020), and media formulation (Jamshidi *et al.* 2020) have already been documented, recently. In these studies, different ML algorithms and performance metrics were employed for the prediction of the output variables. In this study, four different supervised ML algorithms were used for optimizing the *in vitro* shoot regeneration of the aquatic ornamental plant *A. reineckii mini*. Two different explants, five different LED lightings, and three different levels of BAP were used as input variables to measure output variables (shoot count, shoot length).

Materials and methods

In vitro* regeneration of *A. reineckii The plants for *in vitro* regeneration study were obtained from stock material available at Sivas University of Science and Technology, Molecular Biology and Biotechnology Laboratory, Sivas, Türkiye. Two different explants named shoot tip and nodal segment (1st node below the shoot tip) were used for *in vitro* regeneration studies. Both explants were cultured on Murashige and Skoog (MS) medium (Murashige and Skoog 1962), enriched with 0.5, 1.0, and 1.5 mg/l BAP

(6-benzylaminopurine). The culture medium containing both explants was placed in the growth room equipped with different LED lightings. The characteristics of LEDs used in this study are presented in Table 1. The light illuminance of each LED was measured by placing the Lux meter on top of the culture jar, with an estimated distance of 25 cm between LEDs and jars. The basal medium was prepared by adding MS (4.4 g/L) and sucrose (30.0 g/L), and gelled with agar (6.5 mg/L). The pH of the medium was automated to 5.8 by using 1.0 N HCl or NaOH, followed by autoclaving the basal medium for 15 min at 121 °C, and 1.5 Pa atmospheric pressure. The output variables (shoot regeneration frequency, shoot counts, and shoot length) were analyzed by using one-way ANOVA (SPSS 20.0). The data was transformed into arcsine transformation, and the DMRT test was used to compare the difference among the treatments.

Modeling procedures In this study, interaction *via* explant, LEDs, and BAP concentration were used as input variables. Furthermore, *in vitro* shoot count and shoot length were measured as output variables. Machine learning algorithms of KNN (Peter Hart 1967), GP (Hu *et al.* 2019), XGBoost (Chen and Guestrin 2016), and RF (Aggarwal 2018) were employed to determine the relationship between inputs and each output. The performance of the models was evaluated using the technique of leave-one-out cross-validation (LOO-CV) (Webb *et al.* 2011). Grid search was used to find the optimized hyperparameters to determine the best ML model. The open-source Python programming language (Van Rossum and Drake 2009) was used to code all the supervised ML algorithms using the sklearn library (Pedregosa *et al.* 2011). The models' performance was evaluated by calculating R^2 (coefficient of determination) which measures the strength of the relationship between the model and the dependent variables; MSE which tells us how close a regression line is to the measured data points, and mean absolute percentage error (MAPE) that expresses accuracy as a percentage of the error. The mentioned performance metrics are mathematically represented below in Eqs. 1–3.

Table 1. An overview of different LEDs used for *in vitro* regeneration of *Alternanthera reineckii*

LED type	Wavelength (λ max)	Illuminance (LUX)
Green LEDs (G-LEDs)	520	1110
Orange LEDs (O-LEDs)	595	1188
Blue LEDs (B-LEDs)	448	500
Red LEDs (R-LEDs)	633	165
White LEDs (W-LEDs)	525	2030

$$R^2 = 1 - \left(\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \right) 0 \leq R^2 \leq 1 \tag{1}$$

$$\text{MSE} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} 0 \leq \text{MSE} \leq \infty \tag{2}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \text{ MAPE} \geq 0 \tag{3}$$

where Y_i represents the measured value, \hat{Y}_i denotes the predicted value, \bar{Y} corresponds to the mean of the measured values and n is the count of samples. Although MAPE can take any positive value, values greater than 50% are considered inaccurate predictions (Lewis 1982).

K-nearest neighbors (KNN) is a simple non-parametric supervised ML algorithm that can be used to handle both classification and regression problems. It attempts to approximate the association between input variables and the continuous output variable(s) by averaging the observations in the same neighborhood that minimizes their separation. The distance metrics of ‘euclidean,’ ‘manhattan,’ ‘minkowski,’ ‘chebyshev,’ and ‘wminkowsk’ were used to find the best model fit during the hyperparameter optimization. Moreover, all quantitative inputs were normalized by using the below formula before training and testing the model.

$$X_n = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} 0 \leq X_n \leq 1 \tag{4}$$

where X_i is the measured data, X_n is the normalized data, and X_{\max} and X_{\min} are the maximum and minimum data points, respectively.

Gaussian process (GP) regression is another non-parametric supervised learning model that is capable of learning nonlinear maps from inputs to continuous output(s) using a kernel function that builds the covariance matrix among all data pairs. It uses a Bayesian approach to regression and classification problems by applying multivariate Gaussian distribution with two parameters, namely, a mean function that corresponds to the mean vector and a positive definite covariance or kernel function corresponding to a positive definite covariance matrix. It is well suited to work proficiently with small datasets with accuracy, ease of calculation, and consistency (Hu *et al.* 2019). The approach is presented in Eq. 5 for each input x and output y produced by this function.

$$y_i = f(x_i) + \epsilon \tag{5}$$

The Extreme Gradient Boosting (XGBoost) is a very popular decision-tree-based ML algorithm that can be used for

supervised learning tasks such as regression, classification, and ranking (Chen and Guestrin 2016). It is an efficient and improved implementation of the stochastic gradient boosting ensemble algorithm, which uses decision tree models and adds trees one after another to the previous models to decrease the prediction error. Equation 6 indicates the XGBoost objective function, and Eq. 7 shows the model of XGBoost at iteration j that we need to be minimized.

$$y_i = F(x_i) = \sum_{(d=1)}^D f_d(x_i), f_d \in F, i = 1, \dots, n \tag{6}$$

$$L_j = \sum_{(i=1)}^n l(y_i, \hat{y}_i^{(j-1)} + f_j(x_i)) + \Omega(f_j) \tag{7}$$

The random forest (RF) model is another widely used supervised ensemble learning method based on decision trees that can be used for regression, classification, and ranking problems (Breiman 2001). It differs from XGBoost in two key ways. Firstly, it builds each tree independently while XGBoost builds one tree at a time. Secondly, contrary to XGBoost, it combines results at the end of the process by using either the averaging or ‘majority rules’. It is among one of the most widely used ML models due to its simplicity in design, high efficiency, less susceptibility to overfitting, handling the noise, and ability to manage a large number of inputs. Although different distance metrics are supported by RF, the regression models mostly use MSE to measure the distance between the nodes to define which branch is better for the forest. The following Eq. 8 describes this concept (Pavlov 2019).

$$y = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x, x_i) + b \tag{8}$$

where y is the value of the data point and n is the number of samples.

In this work, four different supervised ML algorithms were used to predict shoot count and shoot length with the aid of different input features that were selected while conducting the experimental work.

Results

In vitro regeneration of *A. reineckii* This study presents the successful *in vitro* shoot regeneration of an important ornamental aquatic plant, *A. reineckii* mini, using three different input variables named explant, LED source, and BAP concentration. To optimize the outputs, one-way ANOVA results were analyzed for individual input factors (explant,

LEDs, BAP) and their combination (explant \times LEDs \times BAP). Results illustrated a significant impact of all input factors and resulted in 100% *in vitro* shoot regeneration and callus induction, and, hence, not subjected to statistical analysis. On the other hand, shoot counts and shoot length were affected by corresponding input factors (Fig. 1a, b).

The comparison of two different explants tested in this study played a significant role on shoot counts ($p > 0.01$) of *A. reineckii*. It is apparent from the results that explant type is highly significant for yielding more shoots rather than shoot length. Nodal segment explants generated more shoots (7.19) compared to shoot tip explants and were recorded as 5.14 shoots per explant. Contrarily, almost similar shoot length was documented for both explants and was recorded as 3.77 cm for the nodal segment and 3.67 cm for the shoot tip explant (Fig. 2; Supplementary Table 1). The comparison of three different BAP doses significantly controlled the shoot counts ($p > 0.01$) and shoot length ($p > 0.05$). The maximum value for both output variables was documented on different BAP concentrations. Shoot counts increased gradually with elevated BAP concentration, and maximum shoot counts were observed on a medium augmented with 1.5 mg/L BAP. Instead, the variable impact of BAP concentration was observed and minimum shoot length was documented on MS medium enriched with 1.0 mg/L BAP. Supplementation of both 0.5 and 1.5 mg/L BAP induced statistically similar shoot lengths (Fig. 2; Supplementary Table 1). The response of five different LED types was statistically significant for shoot counts ($p > 0.05$) and shoot length ($p > 0.01$) of *A. reineckii*. The shoot counts were recorded in the descending order

of O-LEDs (7.59) > G-LEDs (6.51) > W-LEDs (6.41) > B-LEDs (5.99) > R-LEDs (4.32), whereas shoot length was significant at $p > 0.01$, and followed the order of B-LEDs (4.93 cm) > O-LEDs (3.72 cm) > G-LEDs (3.58 cm) > W-LEDs (3.31 cm) > R-LEDs (3.06 cm). The R-LEDs were less responsive than other LEDs (Fig. 2; Supplementary Table 1).

The results of the individual input variable significantly affected the shoot counts and shoot length of *A. reineckii*. A similar impact was also observed when their combinations were taken into account. The maximum shoot counts of the nodal segment (14.80) and shoot tip explant (10.40) were recorded from the combination of 1.0 mg/L BAP \times G-LEDs. Besides, minimum shoot counts were recorded as 4.20 (NS \times B-LEDs \times 0.5 mg/L BAP) and 2.53 (ST \times R-LEDs \times 0.5 mg/L BAP). In general, more shoot counts were attributed to nodal segment explants compared to shoot tip explants for all LEDs \times BAP combinations. The results on shoot length revealed 2.62–6.47 cm for the nodal segment and 2.45–5.38 cm for shoot tip explant in combination with LEDs \times BAP. The maximum shoot length for both explants was registered from B-LEDs \times 0.5 mg/L BAP combination (Figure 3; Supplementary Table 2). Supplementation of all input variables resulted in spontaneous rooting (Fig. 1a, b) from all culture media, and plantlets were directly acclimatized in the water.

Application of supervised machine learning algorithms In this work, data regarding shoot count and shoot length were subjected to four different ML models (RF, XGBoost, KNN, and GP) to predict the results. The validation of each model was carried out by using three different performance metrics, namely R^2 , MSE, and MAPE. The value of R^2 ranged from 0 to 1, and values closer to 1 reflect the better performance of the model. The values of R^2 are generally closely linked with MSE values. A low value of MSE reflects better performance of the model, and therefore, high R^2 with low MSE values exemplifies better performance of the model. Similarly, a low value of MAPE also corresponds to better performance of the model and *vice versa*.

Figure 4 provides plots illustrating the performance of the four different supervised ML algorithms that were used to predict the shoot count and shoot length for different experimental input parameters during the study. The scatter plots use circles to represent the shoot tip, whereas a triangle is used to identify the nodal segment of the explant. The colors (green, orange, blue, red, and white) show the predicted values due to different LEDs, and the size of the dots identifies three concentration levels of BAP (0.5, 1.0, and 1.5 mg/L). The 1:1 line (dashed line) also called the identity line is a 45° line representing perfect predictions under ideal conditions.



Figure 1. Multiple shoot induction of *A. reineckii* mini under B-LEDs from (a) nodal segment, (b) shoot tip explant

Figure 2. Impact of explant type, LEDs, and BAP concentration on shoot count and shoot length of *A. reineckii*

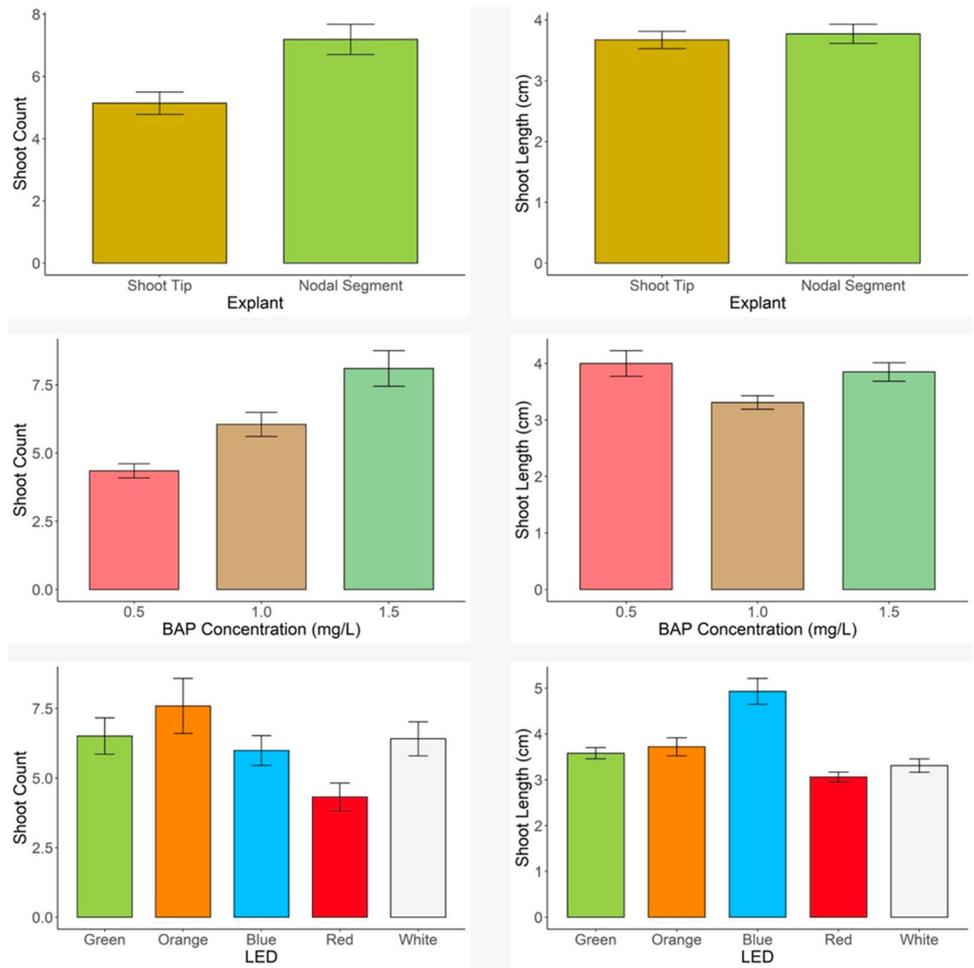


Figure 3. Impact of explant type \times BAP \times LEDs on shoot count and shoot length of *A. reineckii*

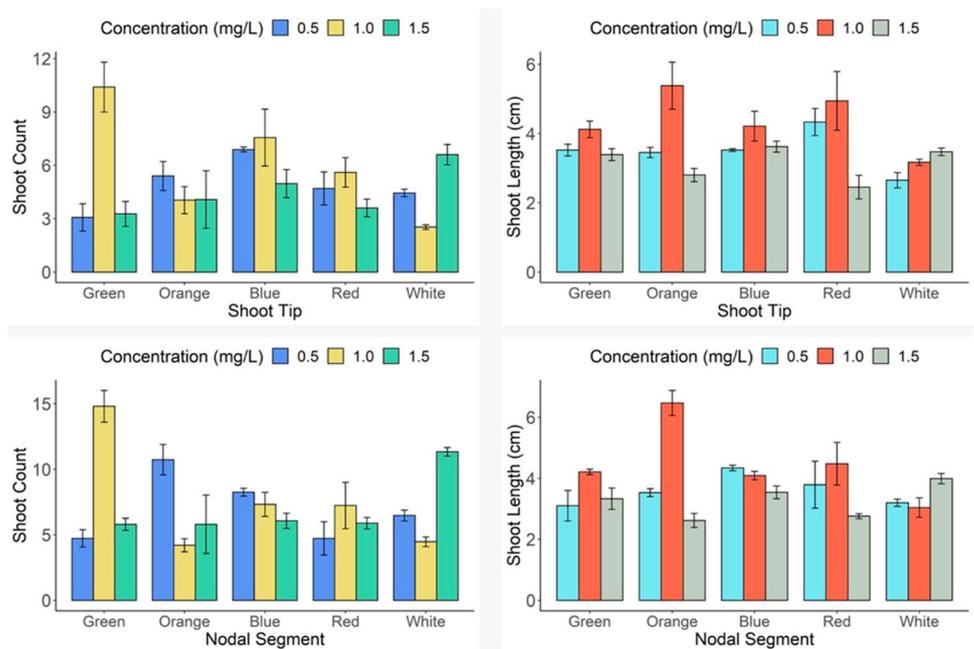


Figure 4. The relationship between the predicted and measured values for shoot count (A) and shoot length (B) for different ML models

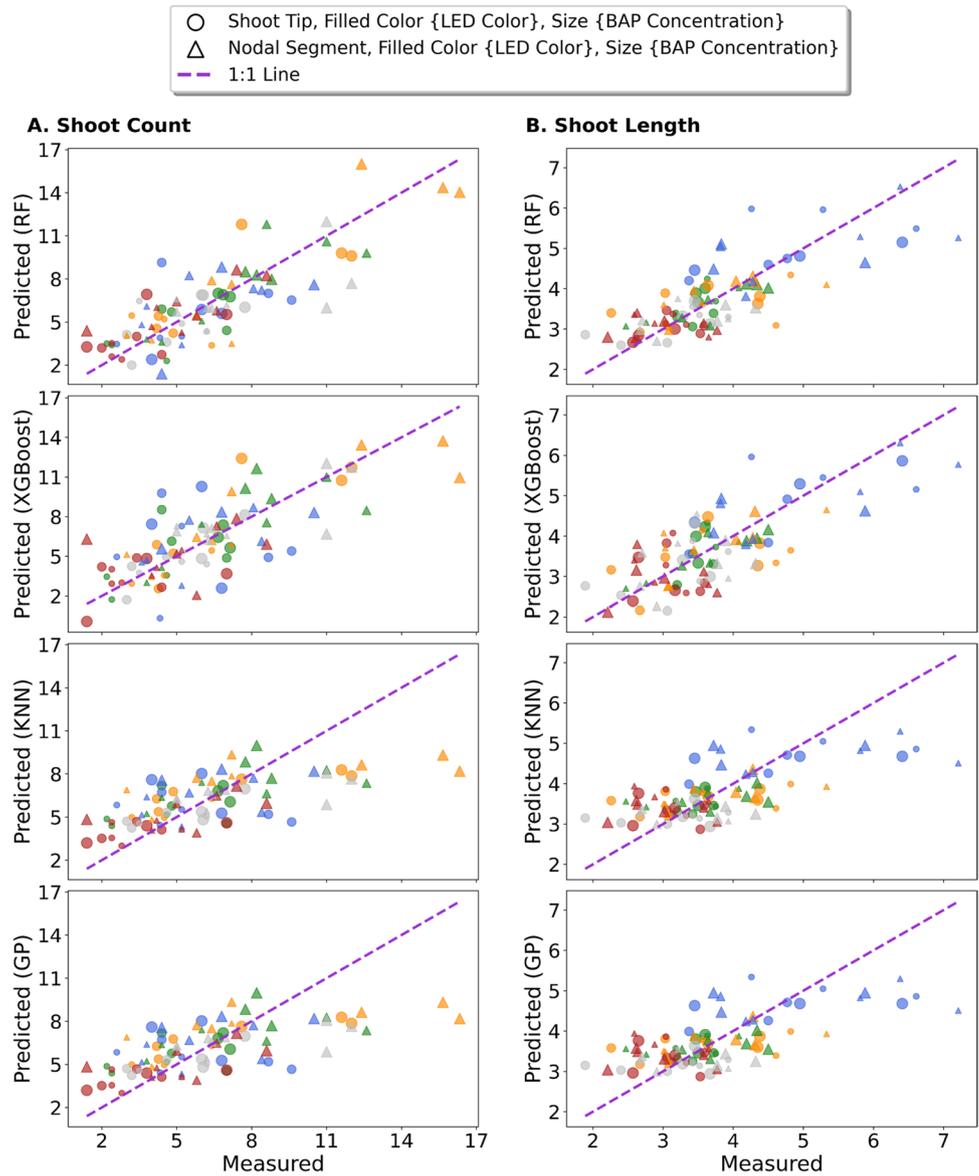


Table 2 depicts values of R^2 , MSE, and MAPE of four different models for shoot counts and shoot length. The results of performance metrics for all models exhibited a similar response for the shoot counts. The results on the R^2 performance metric revealed the order of RF ($R^2=0.61$) > XGBoost ($R^2=0.48$) > KNN ($R^2=0.40$) = GP ($R^2=0.40$). Considering the values of MSE performance metrics, the

results were in order of RF (MSE = 3.48) < XGBoost (MSE = 4.67) < GP (MSE = 5.42) < KNN (MSE = 5.45). Instead, MAPE exhibited the variable order of RF (MAPE = 29.70%) < XGBoost (MAPE = 33.9%) < KNN (MAPE = 35.1%) < GP (MAPE = 35.9%). Comparing the results of all models and performance metrics, RF models predicted the results more accurately followed by

Table 2. Performance metrics for different ML models

	Shoot count			Shoot length (cm)		
	R^2	MSE	MAPE	R^2	MSE	MAPE
RF	0.61	3.48	29.7%	0.59	0.40	13.2%
XGBoost	0.48	4.67	33.9%	0.51	0.48	16.0%
KNN	0.40	5.45	35.1%	0.44	0.55	16.1%
GP	0.40	5.42	35.9%	0.25	0.74	18.9%

the XGBoost model. The results on shoot length exhibited a similar pattern in all performance metrics for the tested ML models. The results on the R^2 performance metric exhibited the order of RF ($R^2=0.59$) > XGBoost ($R^2=0.51$) > KNN ($R^2=0.44$) > GP ($R^2=0.25$). Similarly, the MSE and MAPE values for all models were found as RF (MSE=0.40; MAPE=13.20%) < XGBoost (MSE=0.48; MAPE=16.0%) < KNN (MSE=0.55; MAPE=16.10%) < GP (MSE=0.74; MAPE=18.9%). The results on shoot length again depicted that the RF model is the best model when compared to other models due to its relatively high R^2 and low MSE and MAPE values. *Vice versa*, the GP model provided the lowest performance due to low R^2 along with relatively high MSE and MAPE values.

Discussion

In vitro plant regeneration is the manipulation of a series of input factors (plant, explant, culture medium, and culture conditions). All these factors either individually or in combination lead to *in vitro* regeneration. The type of explant is based on the presence or absence of meristem in the explants. Meristem-containing explants are generally regarded as more efficient for yielding axillary shoot regeneration. In this study, both explants used contained meristematic regions, and already reported for successful *in vitro* regeneration of other aquatic plants (Karatas *et al.* 2014; Dogan 2018; Doğan 2019). Both explants induced shoots efficiently; the nodal segment explant exhibited better performance compared to the shoot tip explant. An investigation on aquatic plants revealed the variable response of both explants possibly due to different genotypes, explant age, and culture conditions (Karatas *et al.* 2014; Dogan *et al.* 2016). Similarly, variable response on shoot length also recorded from both explants confirmed the previous findings in *Ceratophyllum demersum* (Karatas *et al.* 2014).

Culture conditions like temperature, light, *etc.* are the physical factors that regulate the *in vitro* organogenesis. Among these factors, lighting source, intensity, and photoperiod are highly significant. Application of LEDs in the plant tissue culture is prevailing due to certain advantages like specific light illumination, which controls the organogenesis in some aquatic plants like *Bacopa monnieri* (Karataş *et al.* 2016, 2018; Aasim *et al.* 2018a, b), *Limnophila aromatica* (Dogan 2018, 2020), and *Rotala rotundifolia* (Dogan 2020). However, specific light color and illumination are some of the factors which regulate the whole *in vitro* organogenesis (Aasim *et al.* 2018b; Karataş *et al.* 2018; Dogan 2020). The results achieved in this study illustrated the better impact of O-LEDs and G-LEDs on shoot counts as compared to W-LEDs, B-LEDs, or

R-LEDs. The better performance of both O-LEDs and G-LEDs might be the light illumination that falls between 440 and 670 nm used by plants for photosynthesis. On the other hand, supplementation of B-LEDs yielded longer shoots than other LED lights. The results are contrary to the findings on *B. monnieri* (Karataş *et al.* 2016, 2018), possibly due to different explant and culture conditions. The type and concentration of PGRs in the culture medium are one of the most significant factors for *in vitro* organogenesis. The results revealed a gradual increase in shoot count with the respective increase in BAP concentration.

Supplementation of light (type, illumination, photoperiod) along with cytokinins (type and concentration) regulated the *in vitro* organogenesis of *A. reineckii*. Results exhibited the specific LEDs × BAP combination for both explants and for both output variables. Investigating the results of each input factor exhibited a significant impact on shoot counts and shoot length, and maximum performance was associated with O-LEDs or G-LEDs, in combination of 1.5 mg/L BAP. Investigating the combination of all input variables confirmed the similar response, and the combination of O-LEDs × 1.5 mg/L yielded more shoot counts for both explants. On the other hand, supplementation of B-LEDs × 0.5 mg/L BAP and W-LEDs × 1.0 mg/L BAP yielded longer and shorter shoots, respectively, for both explants. The better shoot counts under O-LEDs (Rocha *et al.* 2010), might be due to more endogenous production of cytokinin in response to provided light wavelength (Stirk *et al.* 2011). On the other hand, B-LEDs in combination with low BAP concentrations generated longer shoots. These results revealed the significance of LEDs × BAP combination on shoot counts and shoot length. The studies carried out on aquatic plants using LED lighting systems demonstrated the significance of cytokinin type and concentration on shoot length (Karataş *et al.* 2016; Aasim *et al.* 2018a, b; Dogan 2020). The application of specific light wavelengths by LEDs (Chang *et al.* 2003; Li *et al.* 2010) triggered the photosynthetic pigments (Lian *et al.* 2002), which in turn promotes plant growth. It is also documented in some reports that supplementation of B-LEDs either alone or in combination with R-LEDs leads to more shoot length compared to other LED lights (Karataş *et al.* 2016).

Optimization of plant tissue culture is the manipulation of input factors (triggers), transcriptional cellular responses to the triggers, epigenetic, and molecules stem cell niche (Sugimoto *et al.* 2019), which led to non-deterministic and non-linear developmental patterns in plant cells and tissues (Prasad and Gupta 2008). At the end of the experiment, the output variables tabulated in response to input variables are tested by computer-based statistical software programs. The analysis of variance (ANOVA) and linear regression models are the most commonly employed techniques for checking

the correlation between independent (input) and dependent (output) factors. The outcome of the results is generally interpreted by using different performance tests, and the most commonly used are the least significant difference (LSD), Duncan's multiple range test (DMRT), Tukey's HSD, *etc.* (Ayuso *et al.* 2019). Plant tissue culture techniques are case-sensitive and, hence, need more efficient approaches to optimize or predict the results. The major concerns related to classical computer-based software are due to their inefficiency for rather complex and non-linear inputs, and relatively high probability (Jafari and Shahsavari 2020; Hesami and Jones 2021; Yoosefzadeh-Najafabadi *et al.* 2021a, b).

The latest developments in the field of data science (Ramazan *et al.* 2015; Kul *et al.* 2020) have opened ways for new possibilities to test the output variables with the aid of modern high throughput technologies like ML algorithms and ANN models to optimize, predict, or validation of the results (Katirci *et al.* 2021). In plant tissue culture, the application of these data science technologies (ML, ANN) is relatively new and challenging due to low data set, and expertise. Still, a reasonable number of researchers covering the different aspects of plant tissue culture techniques have been documented in the last few years. In these studies, the selection of algorithms, hyperparameters, cross-validation methods, and performance metrics is highly significant (Kirtis *et al.* 2022). In this study, four different supervised ML models were employed for the prediction of output variables. Results revealed a similar trend in all models, and the RF model predicted more precisely followed by XGBoost, KNN, and GP, respectively. A recent study also revealed the better performance of the MLP model over RF and other ML models for predicting *in vitro* regeneration of chickpea (Kirtis *et al.* 2022). An investigation of ML models exhibited the hypothesis that the prediction of ML models is case sensitive and may vary with a close relationship between input and output variables (Salehi *et al.* 2021; Kirtis *et al.* 2022).

The selection of performance metrics is another important factor in supervised ML models. Previous studies revealed the use of variable performance metrics like R^2 , MSE, RMSE, MAE, *etc.* In this study, three different performance metrics namely R^2 , MSE, and MAPE were employed. The values of R^2 ranged from 0 to 1, and high R^2 values reflect better coherence and association between input and output variables. *Vice versa*, low R^2 values present the low compatibility between input and output variables, but it does not mean that the outcome of output is either non-significant or less significant. The results attained revealed the significant R^2 values for both output variables by using the RF model. Previous reports also highlighted the variable R^2 values for different output variables in plant tissue culture studies. The results of these studies revealed the R^2 values of 0.94 (Hesami *et al.* 2019), 0.56–0.85 (Salehi *et al.* 2020), 0.70 (Hesami *et al.* 2021), and 0.98–1.0 (Kirtis *et al.* 2022).

The advantage of using multiple performance metrics is to validate results more accurately. The MSE is one of the most powerful metrics used in ML modeling along with R^2 . The low MSE values exhibited a better prediction of the models and *vice versa*. The results depicted the relatively low MSE values for shoot length compared to shoot count, but still recorded low enough, and showed a better prediction of all tested models due to low error between actual and predicted values (Kirtis *et al.* 2022). The MAPE is another metric used to assert the production of the models. The value of MAPE below 50.0% is considered as good, but values near zero are more preferred. Results revealed MAPE values below 50.0%, which showed that all models predicted the outputs reasonably. Considering all three performance metrics, the RF model predicted more precisely shoot count due to relatively low MSE and MAPE values as compared to shoot length.

The optimization of plant tissue culture is relatively difficult due to the complex nature of input and output variables. The results revealed that supervised ML algorithm-based models are a powerful tool and can be employed in plant tissue culture studies. It is also recommended to use more models and performance metrics to predict the results more precisely.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1007/s11627-022-10312-6>.

Author contribution MA: conceived the idea and designed the research, data analysis, and article writing; SAA: performed machine learning modeling and article writing; PB: conducted research and data tabulation; MAA: data analysis and article writing.

Data availability The whole datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable and acceptable request.

Declarations

Conflict of interest Not applicable.

References

- Aasim M, Bakhsh A, Sameeullah M, et al (2018a) Aquatic plants as human food. In: Global perspectives on underutilized crops. Springer, pp 165–187
- Aasim M, Karatas M, Bakirci S, Sevinc C (2018b) In vitro adventitious shoot regeneration of water hyssop (*Bacopa monnieri* L. PENNEL) under light emitting diodes (LEDs). *J Glob Innov Agric Soc Sci* 6:129–133
- Aasim M, Katirci R, Akgur O et al (2022b) Machine learning (ML) algorithms and artificial neural network for optimizing *in vitro* germination and growth indices of industrial hemp (*Cannabis sativa* L.). *Ind Crops Prod* 181:114801
- Aasim M, Katirci R, Baloch F, et al (2022a) Innovation in the breeding of common bean through a combined approach of *in vitro* regeneration and machine learning algorithms. *Front Genet* 13:

- Aasim M, Khawar KM, Ahmed SI, Karataş M (2019) Multiple uses of some important aquatic and semiaquatic medicinal plants. In: Plant and human health, Volume 2. Springer, pp 541–577
- Aasim M, Özcan SF, Khawar KM, Özcan S (2012) Comparative studies on the competence of axillary shoot regeneration on unsliced and longitudinally sliced cotyledon nodes of *vigna unguiculata*. *Turk J Botany* 36 <https://doi.org/10.3906/bot-1101-38>
- Aggarwal CC (2018) Neural networks and deep learning
- Al-Tanbouz R, Abu-Qaoud H (2016) In vitro regeneration of chickpea (*Cicer arietinum* L.). *Plant Cell Biotechnol Mol Biol* 17:
- Ayuso M, García-Pérez P, Ramil-Rego P et al (2019) In vitro culture of the endangered plant *Eryngium viviparum* as dual strategy for its ex situ conservation and source of bioactive compounds. *Plant Cell, Tissue Organ Cult* 138:427–435
- Bello-Bello JJ, Martínez-Estrada E, Caamal-Velázquez JH, Morales-Ramos V (2016) Effect of LED light quality on *in vitro* shoot proliferation and growth of vanilla (*Vanilla planifolia* Andrews). *African J Biotechnol* 15:272–277
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Chang HS, Charkabarty D, Hahn EJ, Paek KY (2003) Micropropagation of calla lily (*Zantedeschia albomaculata*) via *in vitro* shoot tip proliferation. *Vitr Cell Dev Biol* 39:129–134
- Chen T, Guestrin C (2016) XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp 785–794
- Dogan M (2018) In vitro shoot regeneration of *Limnophila aromatica* (Lamk.) Merr. from nodal and internodal explants. *Iğdır Univ J Inst Sci Tech* 8:77–84
- Doğan M (2019) In vitro rapid propagation of an aquatic plant *Pogostemon erectus* (Dalzell) Kuntze. *Anatol J Bot* 3:1–6
- Dogan M (2020) The effectiveness of light emitting diodes on shoot regeneration *in vitro* from shoot tip tissues of *Limnophila aromatica* (Lamk.) Merr. and *Rotala rotundifolia* (Buch-Ham. ex Roxb) Koehne. *Biotech Histochem* 95:225–232
- Dogan M, Karatas M, Aasim M (2016) In vitro shoot regeneration from shoot tip and nodal segment explants of *Pogostemon erectus* (Dalzell) Kuntze, a multipurpose ornamental aquatic plant. *Fresenius Environ Bull* 25:4777–4782
- Farhadi S, Salehi M, Moieni A et al (2020) Modeling of paclitaxel biosynthesis elicitation in *Corylus avellana* cell culture using adaptive neuro-fuzzy inference system-genetic algorithm (ANFIS-GA) and multiple regression methods. *PLoS ONE* 15:1–16. <https://doi.org/10.1371/journal.pone.0237478>
- Gonsalves J (2010) Economic botany and ethnobotany. Mittal Publications
- Hesami M, Jones AMP (2021) Modeling and optimizing callus growth and development in *Cannabis sativa* using random forest and support vector machine in combination with a genetic algorithm. *Appl Microbiol Biotechnol* 1–12
- Hesami M, Naderi R, Tohidfar M (2019) Modeling and optimizing In vitro sterilization of chrysanthemum via multilayer perceptron-non-dominated sorting genetic algorithm-II (MLP-NSGAI). *Front Plant Sci* 10:1–13. <https://doi.org/10.3389/fpls.2019.00282>
- Hesami M, Naderi R, Tohidfar M, Yoosefzadeh-Najafabadi M (2020) Development of support vector machine-based model and comparative analysis with artificial neural network for modeling the plant tissue culture procedures: effect of plant growth regulators on somatic embryogenesis of chrysanthemum, as a case study. *Plant Methods* 16:1–15
- Hesami M, Pepe M, Monthony AS et al (2021) Modeling and optimizing *in vitro* seed germination of industrial hemp (*Cannabis sativa* L.). *Ind Crops Prod* 170:113753. <https://doi.org/10.1016/j.indcrop.2021.113753>
- Hu J, Sun Y, Li G et al (2019) Probability analysis for grasp planning facing the field of medical robotics. *Meas J Int Meas Confed* 141:227–234. <https://doi.org/10.1016/j.measurement.2019.03.010>
- Jafari M, Shahsavari A (2020) The application of artificial neural networks in modeling and predicting the effects of melatonin on morphological responses of citrus to drought stress. *PLoS ONE* 15:e0240427
- Jamshidi S, Yadollahi A, Arab MM et al (2020) High throughput mathematical modeling and multi-objective evolutionary algorithms for plant tissue culture media formulation: case study of pear rootstocks. *PLoS ONE* 15:e0243940
- Karatas M, Aasim M, Dogan M (2014) Multiple shoot regeneration of *Ceratophyllum demersum* L. on agar solidified and liquid mediums. *Fresenius Environ Bull* 23:
- Karataş M, Aasim M, Dazkırılı M (2016) Influence of light-emitting diodes and benzylaminopurine on adventitious shoot regeneration of water hyssop (*Bacopa monnieri* (L.) Pennell) *in vitro*. *Arch Biol Sci* 68:. <https://doi.org/10.2298/ABS150803039K>
- Karataş M, Aasim M, Dazkırılı M (2018) Efficacy of light emitting diodes (LEDs) lighting system for *in vitro* shoot regeneration of medicinal water hyssop (*Bacopa monnieri* L. PENNEL). *Rom Biotechnol Lett* 23:13197
- Katirci R, Aktas H, Zontul M (2021) The prediction of the ZnNi thickness and Ni % of ZnNi alloy electroplating using a machine learning method. *Trans Inst Met Finish* 99:162–168. <https://doi.org/10.1080/00202967.2021.1898183>
- Kirtis A, Aasim M, Katurci R (2022) Application of artificial neural network and machine learning algorithms for modeling the *in vitro* regeneration of chickpea (*Cicer arietinum* L.). *Plant Cell, Tissue Organ Cult* 1–12
- Kul M, Oskay KO, Erden F et al (2020) Effect of process parameters on the electrodeposition of zinc on 1010 steel: central composite design optimization. *Int J Electrochem Sci* 15:9779–9795
- Lewis CD (1982) Industrial and business forecasting methods: a practical guide to exponential smoothing and curve fitting. Butterworth-Heinemann
- Li H, Xu Z, Tang C (2010) Effect of light-emitting diodes on growth and morphogenesis of upland cotton (*Gossypium hirsutum* L.) plantlets *in vitro*. *Plant Cell, Tissue Organ Cult* 103:155–163
- Lian M-L, Murthy HN, Paek K-Y (2002) Effects of light emitting diodes (LEDs) on the *in vitro* induction and growth of bulblets of *Lilium* oriental hybrid ‘Pesaro.’ *Sci Hortic (amsterdam)* 94:365–370
- Pavlov YL (2019) Random forests. *Random For* 1–122 <https://doi.org/10.1201/9780429469275-8>
- Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in {P}ython. *J Mach Learn Res* 12:2825–2830
- Prasad VSS, Gupta SD (2008) Applications and potentials of artificial neural networks in plant tissue culture. In: *Plant tissue culture engineering*. Springer, pp 47–67
- Ramazan K, Esma S, Belkis U (2015) Statistical optimisation of organic additives for maximum brightness and brightener analysis in a nickel electroplating bath. *Trans IMF* 93:89–96
- Rocha PSG, Oliveira RP, Scivittaro WB, Saints UL (2010) Diodes emitting light and BAP concentrations in the multiplication *in vitro* of strawberry. *Cienc Rural St Maria* 40:1922–1928
- Saini DC, Singh SK, Rai K, Singh SK (2010) Biodiversity of aquatic and semi-aquatic plants of Uttar Pradesh: (with special reference to eastern Uttar Pradesh). Uttar Pradesh State Biodiversity Board
- Salehi M, Farhadi S, Moieni A et al (2020) Mathematical modeling of growth and paclitaxel biosynthesis in *Corylus avellana* cell culture responding to fungal elicitors using multilayer perceptron-genetic algorithm. *Front Plant Sci* 11:1–12. <https://doi.org/10.3389/fpls.2020.01148>

- Salehi M, Farhadi S, Moieni A et al (2021) A hybrid model based on general regression neural network and fruit fly optimization algorithm for forecasting and optimizing paclitaxel biosynthesis in *Corylus avellana* cell culture. *Plant Methods* 17:1–13
- Sharma A, Jain A, Gupta P, Chowdary V (2020) Machine learning applications for precision agriculture: a comprehensive review. *IEEE Access* 9:4843–4873
- Sotthikul C, Kaewpoowat C, Saimoon N (2015) In vitro propagation of *Habenaria* hybrids. In: VI International Symposium on Production and Establishment of Micropropagated Plants 1155. pp 293–300
- Stirk WA, van Staden J, Novák O et al (2011) Changes in endogenous cytokinin concentrations in *Chlorella* (Chlorophyceae) in relation to light and the cell cycle 1. *J Phycol* 47:291–301
- Sugimoto K, Temman H, Kadokura S, Matsunaga S (2019) To regenerate or not to regenerate: factors that drive plant regeneration. *Curr Opin Plant Biol* 47:138–150
- Van Rossum G, Drake FL (2009) Python 3 reference manual. CreateSpace, Scotts Valley, CA
- Webb GI, Sammut C, Perlich C et al (2011) Leave-one-out cross-validation. *Encyclopedia of machine learning*. Springer, US, Boston, MA, pp 600–601
- Yoosefzadeh-Najafabadi M, Earl HJ, Tulpan D et al (2021a) Application of machine learning algorithms in plant breeding: predicting yield from hyperspectral reflectance in soybean. *Front Plant Sci* 11:2169
- Yoosefzadeh-Najafabadi M, Tulpan D, Eskandari M (2021b) Application of machine learning and genetic optimization algorithms for modeling and optimizing soybean yield using its component traits. *PLoS ONE* 16:e0250665

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.