RESEARCH

Open Access

Machine learning models for optimization, validation, and prediction of light emitting diodes with kinetin based basal medium for in vitro regeneration of upland cotton (Gossypium hirsutum L.)

ÖZKAT Gözde Yalçın¹, AASIM Muhammad^{2*}, BAKHSH Allah^{3*}, ALI Seyid Amjad⁴ and ÖZCAN Sebahattin⁵

Abstract

Background Plant tissue culture has emerged as a tool for improving cotton propagation and genetics, but recalcitrance nature of cotton makes it difficult to develop in vitro regeneration. Cotton's recalcitrance is influenced by genotype, explant type, and environmental conditions. To overcome these issues, this study uses different machine learning-based predictive models by employing multiple input factors. Cotyledonary node explants of two commercial cotton cultivars (STN-468 and GSN-12) were isolated from 7–8 days old seedlings, preconditioned with 5, 10, and 20 mg·L⁻¹ kinetin (KIN) for 10 days. Thereafter, explants were postconditioned on full Murashige and Skoog (MS), $\frac{1}{2}$ MS, $\frac{1}{4}$ MS, and full MS + 0.05 mg·L⁻¹ KIN, cultured in growth room enlightened with red and blue light-emitting diodes (LED) combination. Statistical analysis (analysis of variance, regression analysis) was employed to assess the impact of different treatments on shoot regeneration, with artificial intelligence (AI) models used for confirming the findings.

Results GSN-12 exhibited superior shoot regeneration potential compared with STN-468, with an average of 4.99 shoots per explant versus 3.97. Optimal results were achieved with 5 mg·L⁻¹ KIN preconditioning, ¼ MS postconditioning, and 80% red LED, with maximum of 7.75 shoot count for GSN-12 under these conditions; while STN-468 reached 6.00 shoots under the conditions of 10 mg·L⁻¹ KIN preconditioning, MS with 0.05 mg·L⁻¹ KIN (postconditioning) and 75.0% red LED. Rooting was successfully achieved with naphthalene acetic acid and activated charcoal. Additionally, three different powerful Al-based models, namely, extreme gradient boost (XGBoost), random forest (RF), and the artificial neural network-based multilayer perceptron (MLP) regression models validated the findings.

Conclusion GSN-12 outperformed STN-468 with optimal results from 5 mg·L⁻¹ KIN + ½ MS + 80% red LED. Application of machine learning-based prediction models to optimize cotton tissue culture protocols for shoot regeneration is helpful to improve cotton regeneration efficiency.

Keywords Machine learning, Cotton, In vitro regeneration, Light emitting diodes, Optimization, Kinetin

*Correspondence: Aasim Muhammad mshazim@gmail.com Bakhsh Allah abthebest@gmail.com Full list of author information is available at the end of the article



© The Author(s) 2025. **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, wish http://creativecommons.org/licenses/by/4.0/.



Introduction

Cotton (*Gossypium hirsutum* L.), a prominent crop used for both oil and fiber, is a member of Malvaceae family. It is regarded as the mainstay of the economies of several nations throughout the world, supplying raw materials largely to the textile and oil sectors (Juturu et al. 2015; Jabran et al. 2019). Cotton is cultivated in around 75 countries or regions, and faces similar types of issues, such as heavy infestation of insect pests, disease outbreaks, and other abiotic stresses (Bakhsh et al. 2015; Zafar et al. 2024). The recent advancements in the field of biotechnology and molecular biology have enabled researchers to improve the germplasm by developing resistant varieties against the aforementioned issues (Nadeem et al. 2023).

Using biotechnology, researchers around the world have developed transgenic cotton against insect pests and herbicides (Brookes et al. 2012). Other traits due to the use of biotechnological tools include enhanced fiber quality (Chen et al. 2015). However, the standardization of genetic transformation protocol highly depends on the optimization of in vitro regeneration protocol as cotton is considered one of the most recalcitrant crops under in vitro conditions due to low response and non-reproducible protocols (Kumari et al. 2017; Rajasekaran 2004; Yavuz et al. 2020). Although several protocols utilizing various explants and plant growth regulators (PGR) have been developed recently, researchers are still having difficulty in finding an effective and repeatable strategy that induces multiple shoots. The in vitro regeneration studies on cotton demonstrated somatic embryogenesis, embryogenic cell cultures, and organogenesis from meristematic regions (Rajasekaran 2004).

Meristem transformation protocols offer the unique ability to change individuals without regard to their genotypes, with the ability to generate chimeras in large quantities. Slightly fewer cells survive in regenerated tissues due to the limited quantities of meristematic cells, which lead to low transgenic selection (Firoozabady et al. 1987; Sunilkumar et al. 2001). However, the procedure is still time-consuming, and the time it takes to find transformants is influenced by a cultivar's capacity to regenerate multiple shoots (Keller et al. 1997). The following generation under this circumstance only demonstrates achievement in transformation (John 1997). On the other hand, with the benefit of a short period of culture, embryogenic lines can be routinely subcultured after establishment and used for multiple transformations. In this way, there is a possibility of converting more transformants into regenerated plantlets (Leelavathi et al. 2004). One of the limiting factors for cotton in vitro regeneration is the explant; previous studies emphasized the need to exploit the full potential of the explant (Sunilkumar et al. 2001). Some explants used for cotton in vitro regeneration include leaf petiole (Zhang et al. 2011; Yang et al. 2014), hypocotyl (Kumar et al. 2013), immature zygotic embryos (Hussain et al. 2009), and cotyledonary nodes (Bakhsh et al. 2016). Selection of proper explants with proper PGR and culture conditions is also highly crucial for establishing a protocol due to the phenolic compound's elucidation in the culture medium inhibiting the in vitro regeneration of cotton.

Precision agriculture is increasingly relying on artificial intelligence (AI)-based algorithms in agricultural and biological sciences (Sharma et al. 2020; Whitmire et al. 2021) with very limited use in plant biotechnology, especially in plant tissue culture (Salehi et al. 2021; Aasim et al. 2024a) as compared with other research areas. In recent times, researchers used different models in plant tissue culture for optimizing in vitro germination, sterilization, organogenesis, and somatic embryogenesis (Jafari et al. 2023; Özcan et al. 2023; Şimşek et al. 2024). The selection of hyperparameters and input and output relationships largely influenced the performance and choice of models used in these investigations. Several studies used random forest (RF), extreme gradient boost (XGBoost), and multilayer perceptron (MLP) algorithms along with various performance metrics to validate their results (Katirci et al. 2021; Yan et al. 2020).

In this study, a novel protocol was established by exposing cotton explants to high cytokinin concentration (preconditioning) followed by postconditioning on a medium with no or low cytokinin. Furthermore, explants were cultivated for in vitro regeneration and rooting under various light emitting diodes (LED) lighting, and the outcomes were predicted using AI models. To precisely estimate shoot counts and validate the experimental findings, integrating cutting-edge AI models, such as artificial neural networks (ANN), RF, and XGBoost, was a major goal. With the combination of AI-driven predictive modeling and growth condition optimization, this method offers a fresh framework for enhancing tissue culture techniques.

Material and methods

In vitro regeneration

The experiment was conducted at The Biotechnology laboratory of Karamanoglu Mehmetbey University, Faculty of Science, Karaman, Türkiye. Two different commercial upland cotton cultivars named GSN-12 (May Agro Seeds Pvt Ltd, Türkiye) and Stoneville 468 (STN-468, Nazili Cotton Research Institute, Nazili, Türkiye) were procured. The seeds of both cultivars were delinted by using the standard procedure of treating seeds with

concentrated H₂SO₄ followed by washing with water. The surface sterilization of seeds was performed following the protocol as described previously (Bakhsh et al. 2016). The seeds were exposed to 1% HgCl₂ for 10 min, followed by immediate rinsing for 5 min with sterile water for one time only. Following a 10-min treatment with 0.2% sodium dodecyl sulfate (SDS, Sigma-Aldrich) and 0.2% HgCl₂, the seeds were washed 3 times for 5 min each time with water. Surface-sterilized seeds were inoculated on Murashige & Skoog (MS) medium solidified with agar for 7-8 d to obtain cotyledonary node explants. After isolation, explants were preconditioned on phytagelsolidified MS medium (Murashige et al. 1962) augmented with different concentrations (5, 10, and 20 mg·L) of kinetin (KIN) for 10 d cultured under 80.0% R-LEDs (R:B =4:1 LEDs) lights. Subsequently, explants were cultured on a postconditioned medium containing full MS (4.4 g·L^{-1}), $\frac{1}{2}$ MS (2.2 g·L^{-1}), and $\frac{1}{4}$ MS (1.1 g·L^{-1}) without provided with any PGR. Whereas MS +0.05 mg·L KIN was also used as a fourth postconditioned medium. Postconditioned mediums with explants were cultured under different red and blue (R:B) LEDs combinations in the growth room. The R:B LED combinations used in this study were 80.0% R-LEDs, 75.0% R-LEDs (R:B = 3:1), and 66.67% R-LEDs (R:B = 2:1). Whereas light photoperiod (16 h) and growth room temperature (24 $^{\circ}C \pm 1 ^{\circ}C$) were maintained for all culture conditions. Rooting was done by separating the shoots, followed by culture on MS media containing naphthalene acetic acid (NAA, 0.1 $mg \cdot L^{-1}$), and activated charcoal (1 $g \cdot L^{-1}$). Pots filled with organic peat moss were used for the plant establishment.

The culture media employed in this investigation were made according to the standard protocol containing 30 g·L⁻¹ sucrose with MS at different strengths. Preconditioned culture medium was solidified with agar (6.5 g·L⁻¹), and phytagel (2.5 g·L⁻¹) was used as a gelling agent. All culture mediums used in this study were adjusted at an approximate pH of 5.8 by utilizing a 1 mol·L⁻¹ solution of HCl or NaOH. The filter-sterilized (0.22 μ m) KIN was added after autoclaving the culture medium.

Statistical analysis

In this study, six explants per replication were used, and the experiment was repeated twice. The comparison of shoot numbers for different cultivars, preconditioning doses, postconditioning doses, LEDs, and interaction of all individual cultivars was analyzed by ne-way analysis of variance (ANOVA) and factorial regression analysis. Minitab 20.4 program was used for the analysis, and the difference between the means was compared by Tukey's test. Pareto charts, normal plots, and response optimizers were used for optimizing the input variables by using Minitab 20.4. To illustrate the association between the independent and dependent variables, 2-D contour plots and 3-D surface plots were created using the statistical software Design Expert.

Machine learning (ML) modeling

In this study, preconditioning medium, postconditioning medium, and LEDs were used as input variables for two different cotton cultivars, and shoot count was used as the output variable for ML analysis. Three different models named XGBoost, RF, and MLP were used as AI models for data analysis (Chen et al. 2016; Aggarwal 2018; Silva et al. 2019).

In the field of data science, RF algorithm is the most popular advanced decision tree model due to its extraordinary accuracy, speed, stability, and usability, especially for indicating non-linear relationships. Bagging, also known as bootstrap aggregation, is a technique used by RF that trains many decision trees simultaneously and aggregates their predictions to enhance the performance of the model as a whole to predict the output, which is expressed in equation (1) (Pavlov 2019). This process involves creating multiple bootstrap samples from the training data, training individual models on these samples, and then combining their predictions.

$$\widehat{y} = \frac{1}{M} \sum_{i=1}^{M} f_i(x) \tag{1}$$

where, \hat{y} is the predicted value, $f_i(x)$ is the prediction of the *i*th tree for input *x*, and *M* represents the total number of trees.

XGBoost, another decision-tree-based ML algorithm, was employed for regression, classification, and ranking types of supervised learning tasks (Chen et al. 2016). By incorporating trees into the earlier models, the prediction error is decreased. Equations (2) and (3) present the XGBoost objective function and show the model that must be minimized at the *j*th iteration of XGBoost.

$$y_i = F(x_i) = \sum_{d=1}^{D} f_d(x_i), f_d \in F, i = 1, \dots, n$$
 (2)

The above equation shows a tree ensemble model that uses D additive function to predict the output. Here, y_i is the predicted output; $F(x_i)$ is the ensemble of decision trees; each f_d corresponds to an independent tree structure.

$$L_{j} = \sum_{i=1}^{n} l \left[y_{i}, \hat{y}_{i}^{(j-1)} + f_{j}(x_{i}) \right] + \Omega(f_{j})$$
(3)

where L_j is the objective function at iteration *j* that needs to be minimized, *l* is a function of classification or regression trees (CART) learners composed of a sum of the current and previous additive trees, $\hat{y}_i^{(j-1)}$ is the prediction up to (j-1) trees, $f_j(x_i)$ represents the prediction of the *j*th tree for the *i*th data point, and Ω is the regularization term.

With multiple perceptrons and using the deep and feedforward method, MLP is a preferred ANN model. The whole system is dependent on three primary components or layers (input, output, and hidden). Using the backpropagation technique, the weights and biases are modified in relation to the error (Katirci et al. 2021), and data training keeps going till the next equation is minimized [equation (4)].

$$E = \frac{1}{K} \sum_{k=1}^{K} (y_k - \hat{y}_k)^2$$
 (4)

where *E* is the error, y_k is the value of data point *k*, \hat{y}_k is the predicted value of data point *k*, and *K* is the sample size.

The data, which was composed of 216 data points, were divided into two different sets of training and testing with the aid of leave-one-out cross-validation (LOO-CV) methodology (Webb et al. 2011). The optimization of hyperparameters was performed using Grid search for the optimal model. Using the free and open-source Python programming language (Van Rossum et al. 2009) and the sklearn package (Pedregosa et al. 2011), all of the supervised ML algorithms were developed.

The performance metrics used to include coefficient of determination (R^2) , which exhibits the strength of the association between the dependent variables and the model used, and its values vary from 0 to 1. Mean absolute error (MAE) indicates the average magnitude of the deviations between an observation's predicted value and its actual value. Mean squared error (MSE) shows the distance between a regression line and the observed data points. Mean absolute percentage error (MAPE) is a forecasting system for prediction accuracy. The ratio between the actual and anticipated values is calculated using mean squared logarithmic error (MSLE), whereas median absolute error (MedAE) compares the actual observed reaction with the anticipated response $(0, \infty)$. The performance metrics listed above are all represented mathematically in the following equations

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (Y_{i} - \widehat{Y}_{i})^{2}}{\sum_{i=1}^{n} (Y_{i} - \widetilde{Y})^{2}}$$
(5)

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$$
 (6)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| Y_i - \widehat{Y}_i \right|$$
(7)

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \widehat{Y}_i}{Y_i} \right| \times 100$$
(8)

$$MSLE = \frac{1}{n} \sum_{i=1}^{n} \left(\log(Y_i + 1) - \log(\widehat{Y}_i + 1) \right)^2$$
(9)

$$MedAE = median\left(\left|Y_1 - \widehat{Y}_1\right|, \dots, \left|Y_n - \widehat{Y}_n\right|\right) \quad (10)$$

where Y_i is the actual value, \hat{Y}_i is the predicted value, \tilde{Y} is the mean of actual values, $\log(x)$ is the natural logarithm of *x*, *n* is the sample size.

Additionally, prior to testing and training of the models, all numerical inputs were scaled by using the formula (11).

$$X' = \frac{X_i - \mu}{\sigma} \tag{11}$$

where X' is the standardized value, X_i is the actual data, μ is the mean of the feature values, and σ is the standard deviation of the feature values.

Results

In vitro regeneration

A protocol for in vitro propagation of two commercially grown upland cotton cultivars from Türkiye has been developed, resulting in enhanced shoot regeneration. In this study, preconditioning explants placed oriented at 30-60° showed high shoot regeneration frequencies, with single shoot induction occurring within 2-3 weeks and multiple shoots forming within 4–5 weeks, achieving 100% shoot regeneration for both cultivars. The use of phytagel as a gelling agent further improved the outcomes, reducing phenolic compound leakage and promoting a prolonged culture period without subculturing. The kind of cultivar has a significant impact on the in vitro regeneration of recalcitrant crops along with other factors. Cotton's in vitro regeneration is heavily genotype-dependent, with several cultivars showing comparatively low shoot numbers. The two different cultivars used in this study responded in a variable way (P =0.000), and an average of 4.99 shoots were attained from GSN-12 compared with STN-468, which produced 3.97 shoots per explant (Table S1).

The approach of preconditioning explants with higher cytokinin levels and culturing on a basal medium with varied MS concentrations and KIN was crucial in inducing multiple shoots without callus formation. Exposure of explants to a high KIN-containing medium exerted a positive but statistically insignificant impact on shoot regeneration frequency (100%) and mean shoot counts (P = 0.095). The highest mean shoot count (4.68) were attained from media supplemented with 10 mg·L⁻¹ KIN followed by 5 mg·L⁻¹ (4.54 shoots) and 20 mg·L⁻¹ (4.22 shoots). Results revealed the insignificant (P = 0.077) but clear impact of postconditioning with different MS medium concentrations on shoot induction, with the mean shoot count in order of $\frac{1}{4}$ MS (4.68) $\geq \frac{1}{2}$ MS $(4.57) \ge MS$ (4.09). Optimized culture conditions, including red-to-blue LED combinations, further supported shoot proliferation. The three distinct R:B LED combinations employed in this investigation had a discernible effect on in vitro propagation of cotton. Results revealed the need for the relatively low level of B-LEDs in combination with R-LEDs to induce a maximum number of shoots. The mean shoot counts in response to R:B LEDs were 80.0% R-LED (5.11) >75.0% R-LED (4.38) >66.67% R-LED (3.95). The comparison of individual parameters like cultivar, KIN dosage for preconditioning, culture medium, and LED revealed that GSN-12 cultivar, 10 mg·L⁻¹ KIN, ¹/₄ MS, and 80.0% R-LED were superior and yielded more mean shoot counts compared with their respective treatments (Table S1).

The results of individual input variables were also evaluated by constructing the boxplot. In Fig. 1a, the boxplot shows a higher median value for the shoot counts of GSN-12 compared with STN-468, whereas the spread between the two cultivars is similar. Figure 1b illustrates the highest median value achieved by the preconditioning of 10 mg·L⁻¹ KIN and has the highest spread for the shoot counts. Figure 1c shows ½ MS has the highest median value, the presence of a relatively large number of extreme values in the upper whisker of ¼ MS, and low spread makes the mean shoot count of ¼ MS larger than others for postconditioning. Finally, the 80% R-LED arrangement demonstrates the highest median values together with the spread on the shoot counts (Fig. 1d).

Table 1 presents the combined impact of all the parameters (cultivar, KIN preconditioning dosage, culture medium, and LED) on cotton shoot count, and the results were statistically significant (P = 0.000). For GSN-12, the maximum shoot count (7.75) was attributed



Fig. 1 Boxplot depicting the influence of individual input variables on shoot count in cotton (a) cultivar; (b) preconditioning; (c) postconditioning, 0.05-KIN indicate MS + 0.05 mg·L⁻¹ KIN; 1.10-MS, 2.20-MS, and 4.40-MS indicate 1.10, 2.20, and 4.40 g·L⁻¹ MS, respectively; (d) LED

| PrC | PoC | LED (R:B) | GSN- 12 | | | STN- 468 | | |
|-----|-----------------------------|-----------|---------------|-------|----------------|----------------|-------|----------------|
| | | | Mean | StDev | 95% CI | Mean | StDev | 95% CI |
| 5 | MS | 4:1 | 5.75 ABCDEF | 0.050 | (5.087; 6.413) | 5.00 CDEFGHI | 0.100 | (4.337; 5.663) |
| 5 | MS | 3:1 | 4.75 CDEFGHIJ | 0.100 | (4.087; 5.413) | 4.25 EFGHIJKL | 0.250 | (3.587; 4.913) |
| 5 | MS | 2:1 | 5.25 CDEFGH | 0.010 | (4.587; 5.912) | 3.00 IJKL | 0.050 | (2.337; 3.663) |
| 5 | 1⁄2 MS | 4:1 | 6.75 ABC | 0.750 | (6.087; 7.413) | 2.75 JKL | 0.500 | (2.087; 3.413) |
| 5 | 1⁄2 MS | 3:1 | 5.00 CDEFGHI | 1.000 | (4.337; 5.663) | 5.00 CDEFGHI | 0.000 | (4.337; 5.663) |
| 5 | 1⁄2 MS | 2:1 | 5.00 CDEFGHI | 0.000 | (4.337; 5.663) | 2.75 IJKL | 0.750 | (2.087; 3.413) |
| 5 | 1⁄4 MS | 4:1 | 7.75 A | 0.250 | (7.087; 8.413) | 5.00 CDEFGHI | 1.000 | (4.337; 5.663) |
| 5 | 1⁄4 MS | 3:1 | 4.25 EFGHIJKL | 0.250 | (3.587; 4.913) | 4.50 DEFGHIJK | 0.300 | (3.837; 5.163) |
| 5 | 1⁄4 MS | 2:1 | 4.25 EFGHIJKL | 0.050 | (3.587; 4.913) | 4.00 EFGHIJKL | 0.250 | (3.337; 4.663) |
| 5 | 0.05 mg·L ⁻¹ KIN | 4:1 | 5.75 ABCDEF | 0.250 | (5.087; 6.413) | 3.75 FGHIJKL | 1.000 | (3.087; 4.413) |
| 5 | 0.05 mg·L ⁻¹ KIN | 3:1 | 3.75 FGHIJKL | 0.050 | (3.087; 4.413) | 3.50 GHIJKL | 0.200 | (2.837; 4.163) |
| 5 | 0.05 mg·L ⁻¹ KIN | 2:1 | 4.25 EFGHIJKL | 0.100 | (3.537; 4.863) | 3.00 IJKL | 2.000 | (2.340; 3.660) |
| 10 | MS | 4:1 | 6.50 ABCD | 0.100 | (5.837; 7.163) | 2.50 KL | 0.500 | (1.837; 3.163) |
| 10 | MS | 3:1 | 4.50 DEFGHIJK | 0.010 | (3.837; 5.163) | 2.25 L | 0.250 | (1.587; 2.913) |
| 10 | MS | 2:1 | 4.75 CDEFGHIJ | 0.090 | (4.087; 5.413) | 3.00 IJKL | 0.300 | (2.337; 3.663) |
| 10 | 1⁄2 MS | 4:1 | 6.50 ABCD | 0.500 | (5.837; 7.163) | 5.50 BCDEFG | 0.100 | (4.837; 6.163) |
| 10 | 1⁄2 MS | 3:1 | 5.75 ABCDEF | 0.050 | (5.087; 6.413) | 4.25 EFGHIJKL | 0.010 | (3.587; 4.913) |
| 10 | 1⁄2 MS | 2:1 | 5.00 CDEFGHI | 0.010 | (4.337; 5.663) | 2.50 KL | 1.000 | (1.837; 3.163) |
| 10 | 1⁄4 MS | 4:1 | 5.75 ABCDEF | 0.750 | (5.087; 6.413) | 4.75 CDEFGHIJ | 0.750 | (4.087; 5.413) |
| 10 | 1⁄4 MS | 3:1 | 5.25 CDEFGH | 0.250 | (4.587; 5.913) | 4.750 CDEFGHIJ | 0.050 | (4.087; 5.413) |
| 10 | 1⁄4 MS | 2:1 | 4.00 EFGHIJKL | 0.000 | (3.337; 4.663) | 4.00 EFGHIJKL | 0.000 | (3.337; 4.663) |
| 10 | 0.05 mg·L ⁻¹ KIN | 4:1 | 7.50 AB | 0.100 | (6.837; 8.163) | 2.75 JKL | 0.030 | (2.087; 3.413) |
| 10 | 0.05 mg·L ⁻¹ KIN | 3:1 | 4.75 CDEFGHIJ | 0.500 | (4.087; 5.413) | 6.00 ABCDE | 2.000 | (5.340; 6.660) |
| 10 | 0.05 mg·L ⁻¹ KIN | 2:1 | 6.75 ABC | 0.750 | (6.087; 7.413) | 3.00 IJKL | 0.500 | (2.337; 3.663) |
| 20 | MS | 4:1 | 4.75 CDEFGHIJ | 0.250 | (4.087; 5.413) | 3.50 GHIJKL | 0.200 | (2.837; 4.163) |
| 20 | MS | 3:1 | 4.33 EFGHIJK | 0.080 | (3.667; 4.993) | 3.00 IJKL | 0.000 | (2.337; 3.663) |
| 20 | MS | 2:1 | 4.00 EFGHIJKL | 1.000 | (3.337; 4.663) | 2.50 KL | 0.500 | (1.837; 3.163) |
| 20 | 1⁄2 MS | 4:1 | 5.00 CDEFGHI | 0.100 | (4.337; 5.663) | 4.50 DEFGHIJK | 0.500 | (3.837; 5.163) |
| 20 | 1⁄2 MS | 3:1 | 3.50 GHIJKL | 0.020 | (2.837; 4.163) | 4.00 EFGHIJKL | 0.250 | (3.337; 4.663) |
| 20 | 1⁄2 MS | 2:1 | 4.25 EFGHIJKL | 0.500 | (3.587; 4.913) | 4.25 EFGHIJKL | 0.500 | (3.587; 4.913) |
| 20 | 1⁄4 MS | 4:1 | 4.00 EFGHIJKL | 0.100 | (3.337; 4.663) | 5.67 BCDEF | 1.258 | (5.004; 6.329) |
| 20 | 1⁄4 MS | 3:1 | 4.25 EFGHIJKL | 0.050 | (3.587; 4.913) | 4.75 CDEFGHIJ | 0.050 | (4.087; 5.413) |
| 20 | 1⁄4 MS | 2:1 | 4.00 EFGHIJKL | 0.000 | (3.337; 4.663) | 3.25 HIJKL | 0.150 | (2.587; 3.913) |
| 20 | 0.05 mg·L ⁻¹ KIN | 4:1 | 5.50 BCDEFG | 1.500 | (4.837; 6.163) | 5.50 BCDEFG | 0.500 | (4.837; 6.163) |
| 20 | 0.05 mg·L ⁻¹ KIN | 3:1 | 3.50 GHIJKL | 0.250 | (2.837; 4.163) | 5.33 CDEFG | 0.170 | (4.667; 5.993) |
| 20 | 0.05 mg·L ⁻¹ KIN | 2:1 | 3.00 IJKL | 0.300 | (2.337; 3.663) | 5.00 CDEFGHI | 1.000 | (4.337; 5.663) |

PrC preconditioning medium; 5, 10, 20 indicate the concentration (mg·L⁻¹) of KIN; *PoC* postconditioning medium, *LED* light emitting diodes, *StDev* standard deviation, *CI* confidence interval, *R* red, *B* blue

to the combination of 5 mg·L⁻¹ KIN ×¼ MS ×80.0% R-LED, more shoots (5.75) were attributed to 10 mg·L⁻¹ KIN ×¼ MS ×80.0% R-LED combination. On the other hand, STN-468 responded differently, and the maximum shoot count was attributed to the combination of 10 mg·L⁻¹ KIN ×MS with 0.05 mg·L⁻¹ KIN ×75.0% R-LED. Similarly, both cultivars induced minimum shoots under different combinations, and recorded as 3.00 shoots for GSN-12 under 20 mg·L⁻¹ KIN ×MS with 0.05 mg·L⁻¹

KIN ×66.67% R-LED. Whereas, minimum shoot count (2.25) of STN-468 were linked with a combination of 10 mg·L⁻¹ KIN ×MS ×75.0% R-LED. These results enlighten the significance of cultivar, culture medium, and culture condition on in vitro shoot induction of cotton.

Factorial regression analysis

Factorial regression analysis was used for investigating the impact of input variables on shoot counts by

considering the significant factors and their level by constructing Pareto charts and normal plots. Results of the Pareto chart revealed the fitted line score of 1.984 for both cultivars (Fig. 2a, b). Results of cotton STN-468 revealed the significant impact of postconditioning (B), LED (C), and preconditioning × post-conditioning (AB). Whereas, preconditioning (A), ABC, AC, and BC were non-significant and showed scores of less than 1.984. Considering the significance order, it was registered as B > C > AB > A > ABC > AC > BC (Fig. 2a). On the other hand, a totally different pattern was registered for GSN-12; the order of C > A > AC > BC > B > AB > ABC was registered, and only C and A input variables were significant and exhibited similar impact on shoot counts (Fig. 2b). The significance shown by Pareto chart was further investigated by normal plots (Fig. 2c, d), which reflects the significance in terms of the relationship (direct proportional or inverse proportional impact) between input variables and respective output variables. Results of STN-468 illustrated the positioning of B and AB variables on the left side of the fitted line (Fig. 2c). Whereas factor C was positioned on the right top of the line with a significant level adjusted at around 90.0%. The significance level in terms of percentage was 20.0% for AB and 100% for the B variable, reflecting the weightage of given input parameters on the shoot counts. On the contrary, *C* and A factors were significant and placed on the right and left side of the standard line for GSN-12, respectively (Fig. 2d).

Results of contour plot and surface plot of interaction of two input factors for STN-468 revealed 5.0-5.2 shoot counts from 17.6-20.0 mg·L⁻¹ preconditioning KIN $\times 0.05-0.40$ mg·L⁻¹ postconditioning KIN (Fig. 3a, d). The combination of AC (preconditioning KIN \times LED) optimized the maximum shoot count of 4.4-4.5 from 14.5–20.0 mg·L⁻¹ preconditioning KIN \times 78–80% R-LED (Fig. 3b, e), whereas BC combination was optimized as 4.5-5.0 shoots from 0.05-1.72 mg·L⁻¹ postconditioning KIN ×73.0-80.0% R-LED (Fig. 3c, f). Investigation of GSN-12 optimized the shoot count range of 4.5-4.8 from 6.25–14.0 mg·L⁻¹ preconditioning KIN \times 0.05–4.44 $mg \cdot L^{-1}$ postconditioning KIN (Fig. 4a, d). The combination of AC was registered as 5.0-16.0 mg·L⁻¹ preconditioning KIN \times 79–80% R-LED for maximum shoot count of 6.0-6.5 (Fig. 4b, e). A similar number of shoot counts can also be attained from $0.05-4.44 \text{ mg} \cdot \text{L}^{-1}$ postconditioning KIN ×79-80% R-LED (Fig. 4c, f). It is evident from the results that exposing cotton to 80% red +20%blue lighting leads to maximum shoot counts.



Fig. 2 Pareto chart (a-b) and normal plot (c-d) analysis of shoot counts of in vitro regenerated cotton



Fig. 3 Contour plots (a-c) and surface plots (d-f) for in vitro regenerated shoot counts of STN-468



Fig. 4 Contour plots (a-c) and surface plots (d-f) for in vitro regenerated shoot counts of GSN-12

Results were further optimized by using a response optimizer statistical tool for optimizing the best input condition for achieving maximum shoot count. All three input parameters were optimized. Results revealed a similar requirement of postconditioning of MS with 0.05 mg·L KIN and 80.00% R-LED to induce a maximum shoot count of 5.83 for STN-468, and 6.35 for GSN-12, respectively (Table 2). However, the preconditioning dose requirements were different for the two cultivars (20 $mg \cdot L^{-1}$ KIN for STN-468, and 5.0 $mg \cdot L^{-1}$ KIN for GSN-12). The results showed clearly that all statistical tools optimized the better performance of 80.0% R-LED for yielding maximum shoot counts for both cultivars.

Application of ML modeling

The data generated were thereafter validated and predicted by ML models. R^2 for the tested models were very

 Table 2
 Optimizing input variables for individual cultivars using Response optimizer

| Optimization | STN-468 | GSN-12 |
|---|----------------|----------------|
| Preconditioning KIN concen- tration/(mg·L ⁻¹) | 20.00 | 5.00 |
| Postconditioning KIN con- centration/(mg·L ⁻¹) | 0.05 | 0.05 |
| R-LED /% | 80.00 | 80.00 |
| Shoot count | 5.83 | 6.35 |
| SE fit | 0.37 | 0.31 |
| 95% CI | (5.099; 6.561) | (5.735; 6.964) |
| | | |

SE standard error, CI confidence interval

close and ranged from 0.69 to 0.71, with the maximum R^2 (0.71) recorded from the MLP model, followed by RF (0.70) and XGBoost (0.69). The results of actual and predicted scores of shoot counts are presented in Fig. 5. A 1:1 line in ML is used in regression analysis to compare the predicted values with actual values. It represents a perfect agreement between the predicted and actual outcomes. Deviations from this line indicate prediction

errors. The line is also used in residual analysis to visualize how well a model fits the data. The performance metrics of different models ranged from 0.477 to 0.515 for MSE, 0.372 to 0.414 for MAE, 10.116% to 10.976% for MAPE, 0.0194 to 0.0204 for MSLE, and 0.078 to 0.146 for MedAE (Table 3). The minimum scores for MSE and MSLE were attributed to the MLP model, whereas, minimum MAE, MAPE, and MedAE were associated with RF model. Overall, MLP exhibited better performance, and all three models exhibited very similar performance. The RF model outperformed the MLP and XGBoost models in terms of data prediction and validation.

Discussion

The establishment of in vitro propagation protocol is regulated by the combinations of variable physical, chemical, and biological factors. The selection of proper cultivar, culture medium, and culture conditions are direly decisive for establishing successful and repeatable in vitro regeneration protocol of recalcitrant plants (Wang et al. 2011; Parris et al. 2012). Cotton is believed to be one of the most recalcitrant plants to manipulate, as multiple shoot induction in cotton is challenging (Pathi et al.



Fig. 5 Actual and predicted values of shoot count of different ML models (a) MLP, (b) XGBoost, and (c) RF model

| ML models | R ² | MSE | MAE | МАРЕ | MSLE | MedAE |
|-----------|----------------|-------|-------|--------|--------|-------|
| RF | 0.70 | 0.492 | 0.372 | 10.116 | 0.0201 | 0.078 |
| XGBoost | 0.69 | 0.515 | 0.414 | 10.976 | 0.0204 | 0.146 |
| MLP | 0.71 | 0.477 | 0.402 | 10.888 | 0.0194 | 0.137 |

 Table 3
 Performance metrics for the validation of ML models

RF random forest, *XGBoost* extreme gradient boosting, *MLP* multilayer perceptron, *R*² coefficient of determination, *MSE* mean squared error, *MAE* mean absolute error, *MAPE* mean absolute percentage error, *MSE* mean square log error, *MedAE* median absolute error

2013). Therefore, it is always a priority to induce reproducible and efficient multiple shoots for cotton breeding programs (Khan et al. 2023). Optimization of protocol can be achieved by using traditional statistical tools or employing modern optimizing tools like AI-based models. In this study, a novel protocol with multiple shoot induction in cotton followed by successful rooting and acclimatization of two commercially grown cotton cultivars in Türkiye was established by exposing explants to high KIN concentrations. Thereafter, different culture conditions were optimized by using various optimizing tools like contour plots, surface plots, and response optimizers. In the end, the attained data were validated and predicted through ANN and ML models.

Explant orientation (the placement of explant and contact with a culture medium) is highly significant, but neglected factors that regulate in vitro regeneration (Bhatia et al. 2005; García-Luis et al. 2006). The preconditioned explants with KIN were placed on the culture medium in a special position (angelized position of 30-60°) rather than placing explants horizontally or vertically. The reason behind placing it in a certain position was the initial observations taken in other experiments (data not provided), which resulted in minimum necrosis. For a total of eight weeks, the explants were continually grown on the same medium. There is the possibility of hindering metabolite movements and minimizing the leakage of phenolic compounds in the culture medium due to explant orientation, which in turn leads to high shoot regeneration frequency. A study by García-Luis et al. (2006) revealed the significant impact of explant orientation on callus growth and shoot induction of Troyer citrange. Similarly, the positive impact of explant orientation for the inhibition of phenolic compounds has been documented for pistachio by placing explants at a 60° angle (Nezami et al. 2015).

The incorporation of gelling agents in the culture media is important for in vitro regeneration and multiple shoot induction. Agar is the most preferred gelling agent; however, certain issues like impurities, growth inhibitory compounds, and vitrification are associated with it in plant tissue culture (Nairn et al. 1995). The incorporation of phytagel as an alternative gelling agent can address these concerns owing to its high ash content and low impurities (Huang et al. 1995). Relatively low leakage of phenolic compounds was observed in the culture medium, which in turn resulted in prolonged culture with avoidance of the subculture. The positive impact of phytagel might be due to the provision of a more suitable environment (hydration and nutrition) to explants and controlling the leakage of phenolic compounds in the regeneration medium (Kumar et al. 2003).

The physical state of the explants prior to culture controls the entire in vitro regeneration process. In this research, explants were initially exposed to higher KIN concentrations, followed by culturing on the basal medium (different concentrations of MS and KIN) under different R:B LED combinations. Preconditioning (pretreatment) is exposing explants to higher cytokinin levels for a certain period. Using this approach will cause recalcitrant crops to regenerate more quickly and induce multiple shoots (Kumari et al. 2017). The effectiveness of the technique is the rapid and more cell division of explants at the initial stage, followed by shoot induction on media supplemented without any PGR or containing low concentrations of cytokinin or cytokinin-auxin combination. There was no sign of callus induction on the preconditioned with a high concentration of KIN. Contrarily, callus induction from the basal end of the explants in response to preconditioning has been documented in other crops like chickpeas (Aasim et al. 2013). However, other factors also regulate the whole morphogenesis. In this study, single shoot induction from explants in both cultivars was initiated simultaneously after approximately 2-3 weeks and it took 4-5 weeks for multiple shoots, resulting in 100% shoot regeneration frequency. Previous studies also revealed that preconditioning had no negative impact on regeneration frequency (Kumari et al. 2017). The findings demonstrate the usefulness of preconditioning doses in producing multiple shoots and corroborate with the earlier research (Tang et al. 2012; Kumari et al. 2017).

Results further illustrated the significance of genotypes, as the in vitro regeneration performance of GSN-12 was better than STN-468. The results confirmed previous studies that emphasized the role of cotton genotype or cultivar in regulating in vitro regeneration (Sakhanokho et al. 2004; Khan et al. 2010; Pushpa et al. 2010). However, proper selection of PGR with relative dose and exposure time along with explants also plays a role. KIN is a naturally found cytokinin used for inducing in vitro regeneration at relatively low concentrations. The type and concentration of the basal medium along with PGR are the prerequisites for inducing multiple shoots, especially for recalcitrant crops like cotton. The culture medium (postconditioned medium) for preconditioned explants is highly significant for inducing multiple shoots with high regeneration frequency. Therefore, postconditioned mediums are generally enriched with low cytokinin or cytokinin-auxin combinations (Aasim et al. 2013). In this research, preconditioned explants

were kept on media supplemented with variable MS concentrations without any PGR, and MS medium with a minute amount of KIN (0.05 mg \cdot L⁻¹). Concentration of the MS culture medium also regulated the in vitro regeneration behaviors of cotton, and a significant impact was noted. On the contrary, a negative impact of low MS concentration on shoot induction has been documented for Phytolacca dioica (El-Afry et al. 2017), Ophiorrhiza prostrata (Gopalakrishnan et al. 2018), and Eryngium viviparum (Ayuso et al. 2019). The difference is supposed to be related to the specific demand for macro- and micronutrients of each genotype. The results also revealed that full MS was more detrimental for inducing shoots, and this restriction can be overcome by the provision of KIN. It has been established that reduced MS concentration in the culture media promotes somatic embryogenesis in cotton (Kumria et al. 2003). The usage of LED light, either alone or in conjunction with R:B LED light at various ratios, has been shown to have a significant impact on in vitro germination and regeneration of different plants (Özcan et al. 2023). The study revealed the positive impact of R-LED on somatic embryogenesis in cotton. Whereas, the significance of the balanced use of LED (R:B = 1:1) for gaining the highest growth and morphogenesis in cotton has already been documented (Li et al. 2010). Within the scope of our study, shoot regeneration increased as the ratio of red LED light increased. The findings underscore the synergistic role of explant orientation, preconditioning treatments, and culture medium composition in overcoming recalcitrance in cotton for in vitro regeneration, paving a way for more effective cotton breeding and genetic improvement programs.

Pareto charts and normal plots are powerful statistical tools to estimate the significant impact in terms of ranking or placing the input variables in order. Whereas normal plots illustrate the relationship between the input and output variable in terms of direct proportional or inverse proportional impact with efficiency expressed in percentage (Katirci 2015). The use of both tools has been increasing in plant sciences and registered for in vitro regeneration (Aasim et al. 2024b), nanoparticle biosynthesis (Keijok et al. 2019), and indole acetic acid generation (Myo et al. 2019). Results of both cultivars exhibiting different responses to input variables were confirmed with contour and surface plots in this study. Both plots are highly significant and powerful tools for optimizing two input variables for a desired output target by splitting the data and expressing it with different colors (Kasman et al. 2019; Younis et al. 2023). The results of contour plots optimized the two input variables and their impact on generating the final shoot counts. It has been proven that both charts can be used for phytoremediation investigations (Jaskulak et al. 2020; Mohamad Thani et al. 2020) and in vitro propagation studies (Özcan et al. 2023).

The outcomes of the one-way ANOVA showed how input variables significantly affected the shoot number of two distinct cotton cultivars. Nevertheless, the application of traditional methods is unable to capture the apparent impact of input factors on the final output variables. Application of AI-based ML/ANN models for validating, predicting, and optimizing the data set through inferring the connection between the variables (input and output) has been documented (Balasubramani et al. 2020; Razzaghi et al. 2018). The data were exposed to several ML (RF, XGBoost) and ANN models (MLP) in order to predict and verify the accuracy. By considering the six measured performance criteria, all tested models predicted the results equally and precisely. These performance metrics are highly significant for evaluating the performance of the models. Among the performance metrics, R^2 is the most widely used and well-established metric for data prediction. However, the use of multiple metrics is generally recommended and used for better modeling performance (Özcan et al. 2023; Wu et al. 2023). A high R^2 value of 1 or nearly 1 combined with low values of other metrics indicates that the model is doing better (Arab et al. 2016). Comparably, RF model predicts outcomes better than alternative models for maximizing in vitro hemp callus formation and development (Hesami et al. 2021). On the other hand, in vitro regeneration of common bean has shown that MLP model performs better than RF and XGBoost models. Due to their widespread use in plant sciences and biotechnology, these models have all become quite prominent recently. The studies conducted for GSN-12 and STN-468 cultivars are very limited in the literature, and studies on in vitro germination and regeneration used for genetic transformations have been documented by various researchers (Bakhsh et al. 2016). Our results are highly significant and can be used for future biotechnological applications in cotton (Khan et al. 2023).

Conclusion

In this study, an effective in vitro regeneration strategy involving shoot induction, roots, and acclimation was established. The findings demonstrated the influence of individuals and combinations of various external factors on shoot counts of two distinct cotton cultivars grown in Türkiye. The response of GSN-12 was far better than STN-468 and yielded more shoots. Application of 10 mg·L⁻¹ KIN (preconditioning) for 10 days was superior to other doses. Provision of 80.0% R-LED was beneficial for both cultivars. The impact of combinations of factors revealed that both cultivars require MS with 0.05 mg·L⁻¹ KIN and 80.0% R-LED but different preconditioning KIN doses to generate more shoots. The use of AI/ML models accurately predicted and validated the outcomes. Results illustrate the possible use of other recalcitrant cotton cultivars and may be successfully used for the application of biotechnological tools to improve cotton genotypes. This study suggests that integrating advanced AI and ML techniques could improve cotton tissue culture protocols. Deep learning algorithms and complex predictive models could optimize regeneration strategies for recalcitrant cotton cultivars. These models could predict and enhance external factors, improving shoot induction and acclimatization processes.

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s42397-025-00222-4.

Supplementary Material 1

Acknowledgements

The authors are thankful to the Director, Cotton Research Institute Nazili, and May Seeds Company for providing seeds of cultivars used for this study.

Authors' contributions

Özcan S and Aasim M conceptualized the study, Özkat GY and Bakhsh A executed the projects. Aasim M and Bakhsh A drafted the article and edited it. Ali SA and Aasim M analyzed the data. Özcan S supervised the overall study. All authors read and approved the final manuscript.

Funding

Not applicable.

Data availability

The data generated and materials used in this 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 Bakhsh A is a member of the Editorial Board of Journal of Cotton Research. Author Bakhsh A was not involved in the journal's review of, or decision related to this manuscript.

Author details

¹Department of Bioengineering, Faculty of Engineering, Recep Tayyip Erdogan University, Rize 53020, Türkiye. ²Department of Plant Protection, Faculty of Agricultural Sciences and Technology, Sivas University of Science and Technology, Sivas 58000, Türkiye. ³Centre of Excellence in Molecular Biology, University of the Punjab, Lahore 53700, Pakistan. ⁴Department of Information Systems and Technologies, Bilkent University, Ankara 06800, Türkiye. ⁵Department of Field Crops, Faculty of Agriculture, Ankara University, Ankara 06110, Türkiye.

Received: 3 April 2024 Accepted: 21 March 2025 Published online: 14 May 2025

References

- Aasim M, Katırcı R, Acar AŞ, et al. A comparative and practical approach using quantum machine learning (QML) and support vector classifier (SVC) for light emitting diodes mediated in vitro micropropagation of black mulberry (*Morus nigra* L.). Ind Crop Prod. 2024;213:118397. https://doi.org/10. 1016/j.indcrop.2024.118397.
- Aasim M, Yıldırım B, Say A, et al. Artificial intelligence models for validating and predicting the impact of chemical priming of hydrogen peroxide (H₂O₂) and light emitting diodes on *in vitro* grown industrial hemp (*Cannabis sativa* L.). Plant Mol Biol. 2024b;114:33. https://doi.org/10.1007/ s11103-024-01427-y.
- Aasim M, Day S, Rezaei F, et al. Multiple shoot regeneration of plumular apices of chickpea. Turk J Agric for. 2013;37:33–9. https://doi.org/10.3906/ tar-1204-38.
- Aggarwal CC. Training deep neural networks. In: Aggarwal CC, editor. Neural networks and deep learning. Cham, Switzerland: Springer; 2018. p. 105–67. https://doi.org/10.1007/978-3-319-94463-0_3.
- Arab MM, Yadollahi A, Shojaeiyan A, et al. Artificial neural network genetic algorithm as powerful tool to predict and optimize *in vitro* proliferation mineral medium for G × N15 rootstock. Front Plant Sci. 2016;7:1526. https://doi.org/10.3389/fpls.2016.01526.
- Ayuso M, García-Pérez P, Ramil-Rego P, et al. *In vitro* culture of the endangered plant *Eryngium viviparum* as dual strategy for its *ex situ* conservation and source of bioactive compounds. Plant Cell Tiss Organ Cult. 2019;138:427– 35. https://doi.org/10.1007/s11240-019-01638-y.
- Bakhsh A, Khabbazi SD, Baloch FS, et al. Insect-resistant transgenic crops: retrospect and challenges. Turkish J Agric For. 2015;39:531–48. https://doi. org/10.3906/tar-1408-69.
- Bakhsh A, Anayol E, Khabbazi SD, et al. Development of insect-resistant cotton lines with targeted expression of insecticidal gene. Arch Biol Sci. 2016;68:773–80. https://doi.org/10.2298/abs151012063b.
- Balasubramani K, Sivarajasekar N, Muthusaravanan S, et al. Efficient removal of antidepressant Flupentixol using graphene oxide/cellulose nanogel composite: particle swarm algorithm based artificial neural network modelling and optimization. J Mol Liq. 2020;319:114371. https://doi.org/ 10.1016/j.mollig.2020.114371.
- Bhatia P, Ashwath N, Midmore DJ. Effects of genotype, explant orientation, and wounding on shoot regeneration in tomato. Vitr Cell Dev Biol. 2005;41:457–64. https://doi.org/10.1079/ivp2005649.
- Brookes G, Barfoot P. The income and production effects of biotech crops globally 1996–2010. GM Crops Food. 2012;3:265–72. https://doi.org/10. 4161/gmcr.20097.
- Chen JP, Burke JJ. Developing fiber specific promoter-reporter transgenic lines to study the effect of abiotic stresses on fiber development in cotton. PLoS One. 2015;10(6):e0129870. https://doi.org/10.1371/journal.pone. 0129870.
- Chen TQ, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on knowledge discovery and data mining. New York, USA: ACM; 2016. p. 785–94. https:// doi.org/10.1145/2939672.2939785.
- El-Afry MM, Nofal ES, Saadawy FM, et al. Tissue culture studies on propagation of some ornamental plants 1. Starting and shooting stages of *Phytolacca dioica* L. J Plant Prod. 2017;8(1):27–32. https://doi.org/10.21608/jpp.2017. 37464.
- Firoozabady E, DeBoer DL, Merlo DJ, et al. Transformation of cotton (Gossypium hirsutum L) by Agrobacterium tumefaciens and regeneration of transgenic plants. Plant Mol Biol. 1987;10:105–16. https://doi.org/10.1007/bf000 16148.
- García-Luis A, Molina RV, Varona V, et al. The influence of explant orientation and contact with the medium on the pathway of shoot regeneration *in vitro* in epicotyl cuttings of *Troyer citrange*. Plant Cell Tiss Organ Cult. 2006;85:137–44. https://doi.org/10.1007/s11240-005-9060-4.
- Gopalakrishnan K, Krishnan S, Peringattulli NK. Tissue culture studies and estimation of camptothecin from *Ophiorrhiza prostrata* D. Don. Ind J Plant Physiol. 2018;23:582–92. https://doi.org/10.1007/s40502-018-0391-7.
- Hesami M, Jones AMP. 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. 2021;105:5201–12. https://doi.org/10.1007/s00253-021-11375-y

Huang LC, Kohashi C, Vangundy R, et al. Effects of common components on hardness of culture media prepared with gelrite[™]. In Vitro Cell Dev Biol-Plant. 1995;31:84–9. https://doi.org/10.1007/bf02632242.

Hussain SS, Rao AQ, Husnain T, et al. Cotton somatic embryo morphology affects its conversion to plant. Biol Plant. 2009;53:307–11. https://doi.org/ 10.1007/s10535-009-0055-6.

Jabran K, Ul-Allah S, Chauhan BS, et al. Chapter 1: an introduction to global production trends and uses, history and evolution, and genetic and biotechnological improvements in cotton. In: Jabran S, Chauhan BS, editors. Cotton production. Chichester, UK: John Wiley & Sons Ltd; 2019. p. 1–22. https://doi.org/10.1002/9781119385523.ch1.

Jafari M, Daneshvar MH. Prediction and optimization of indirect shoot regeneration of *Passiflora caerulea* using machine learning and optimization algorithms. BMC Biotechnol. 2023;23:27. https://doi.org/10.1186/ s12896-023-00796-4.

Jaskulak M, Grobelak A, Vandenbulcke F. Modeling and optimizing the removal of cadmium by *Sinapis alba* L. from contaminated soil via response surface methodology and artificial neural networks during assisted phytoremediation with sewage sludge. Int J Phytorem. 2020;22:1321–30. https://doi.org/10.1080/15226514.2020.1768513.

John ME. Cotton crop improvement through genetic engineering. Crit Rev Biotechnol. 1997;17:185–208. https://doi.org/10.3109/073885597091466 13.

Juturu VN, Mekala GK, Kirti PB. Current status of tissue culture and genetic transformation research in cotton (*Gossypium* spp.). Plant Cell Tiss Organ Cult. 2015;120:813–39. https://doi.org/10.1007/s11240-014-0640-z.

Kasman M, Riyanti A, Salmariza S, et al. Response surface methodology approach for analysis of phytoremediation process of Pb (II) from aqueous solution using *Echinodorus palaefolius*. In: IOP conference series: materials science and engineering, vol. 546. Bristol, UK: IOP Publishing; 2019. p. 22009. https://doi.org/10.1088/1757-899x/546/2/022009.

Katirci R. Statistical approach to optimizing a Zn–Ni bath containing ED and TEA. Surf Rev Lett. 2015;22(1):1550015. https://doi.org/10.1142/s0218 625x15500158.

Katirci R, Aktas H, Zontul M. The prediction of the ZnNi thickness and Ni % of ZnNi alloy electroplating using a machine learning method. Trans IMF. 2021;99(3):162–8. https://doi.org/10.1080/00202967.2021.1898183.

Keijok WJ, Pereira RHA, Alvarez LAC, et al. Controlled biosynthesis of gold nanoparticles with *Coffea arabica* using factorial design. Sci Rep. 2019;9:1–10. https://doi.org/10.1038/s41598-019-52496-9.

Keller G, Spatola L, McCABE D, et al. Transgenic cotton resistant to herbicide bialaphos. Transgenic Res. 1997;6:385–92. https://doi.org/10.1023/a: 1018483300902.

Khan RA, Ghayas M, Khalid MN, et al. Transgenic strategies for enhancing cotton disease resistance: current status and future directions. Agrobiol Rec. 2023;13:82–91. https://doi.org/10.47278/journal.abr/2023.028.

Khan T, Reddy VS, Leelavathi S, et al. High-frequency regeneration via somatic embryogenesis of an elite recalcitrant cotton genotype (*Gossypium hirsutum* L.) and efficient *Agrobacterium*-mediated transformation. Plant Cell Tiss Organ Cult. 2010;101:323–30. https://doi.org/10.1007/ s11240-010-9691-y.

Kumar A, Palni LMS. The effect of light source and gelling agent on micropropagation of *Rosa damascena* Mill. and *Rhynchostylis retusa* (L) Bl. J Hortic Sci Biotechnol. 2003;78(6):786–92. https://doi.org/10.1080/14620316. 2003.11511700.

Kumar M, Singh H, Shukla AK, et al. Induction and establishment of somatic embryogenesis in elite Indian cotton cultivar (*Gossypium hirsutum* L. cv Khandwa-2). Plant Signal Behav. 2013;8:e26762. https://doi.org/10.4161/ psb.26762.

Kumari K, Lal M, Saxena S. Enhanced micropropagation and tiller formation in sugarcane through pretreatment of explants with thidiazuron (TDZ). 3 Biotech. 2017;7:282. https://doi.org/10.1007/s13205-017-0910-7.

Kumria R, Sunnichan VG, Das DK, et al. High-frequency somatic embryo production and maturation into normal plants in cotton (*Gossypium hirsutum*) through metabolic stress. Plant Cell Rep. 2003;21:635–9. https:// doi.org/10.1007/s00299-002-0554-9.

Leelavathi S, Sunnichan VG, Kumria R, et al. A simple and rapid Agrobacteriummediated transformation protocol for cotton (*Gossypium hirsutum* L.): embryogenic calli as a source to generate large numbers of transgenic plants. Plant Cell Rep. 2004;22:465–70. https://doi.org/10.1007/ s00299-003-0710-x. Li HM, Xu ZG, Tang CM. Effect of light-emitting diodes on growth and morphogenesis of upland cotton (*Gossypium hirsutum* L) plantlets *in vitro*. Plant Cell Tiss Organ Cult. 2010;103:155–63. https://doi.org/10.1007/ s11240-010-9763-z.

Mohamad Thani NS, Mohd Ghazi R, Abdul Wahab IR, et al. Optimization of phytoremediation of nickel by *Alocasia puber* using response surface methodology. Water. 2020;12(10):2707. https://doi.org/10.3390/w12102707.

Murashige T, Skoog F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant. 1962;15(3):473–97. https://doi.org/ 10.1111/j.1399-3054.1962.tb08052.x.

Myo EM, Ge BB, Ma JJ, et al. Indole-3-acetic acid production by *Streptomyces fradiae* NKZ-259 and its formulation to enhance plant growth. BMC Microbiol. 2019;19:155. https://doi.org/10.1186/s12866-019-1528-1.

Nadeem M, Hasnain M, Shahzad N, et al. Expression analysis of fiber related genes in different staple length genotypes in cotton (*G. hirsutum* L.). Int J Agric Biosci. 2023;12(3):180–87. https://doi.org/10.47278/journal.ijab/ 2023.062

Nairn BJ, Furneaux RH, Stevenson TT. Identification of an agar constituent responsible for hydric control in micropropagation of radiata pine. Plant Cell Tiss Organ Cult. 1995;43:1–11. https://doi.org/10.1007/bf00042665.

Nezami SR, Yadollahi A, Hokmabadi H, et al. Control of shoot tip necrosis and plant death during *in vitro* multiplication of pistachio rootstock UCB1 (*Pistacia integrima* × *P. atlantica*). J Nuts. 2015;6(1):27–35. https://doi.org/ 10.22034/JON.2015.515646.

Özcan E, Atar HH, Ali SA, et al. Artificial neural network and decision tree– based models for prediction and validation of *in vitro* organogenesis of two hydrophytes—*Hemianthus callitrichoides* and *Riccia fluitans*. Vitr Cell Dev Biol. 2023;59:547–62. https://doi.org/10.1007/s11627-023-10367-z.

Parris JK, Touchell DH, Ranney TG, et al. Basal salt composition, cytokinins, and phenolic binding agents influence *in vitro* growth and *ex vitro* establishment of Magnolia 'Ann.' Hort Sci. 2012;47(11):1625–9. https://doi.org/10. 21273/hortsci.47.11.1625.

Pathi KM, Tuteja N. High-frequency regeneration via multiple shoot induction of an elite recalcitrant cotton (*Gossypium hirsutum* L. cv Narashima) by using embryo apex. Plant Signal Behav. 2013;8:e22763. https://doi.org/10. 4161/psb.22763.

Pavlov YL. Random forests. Boston, USA: De Gruyter. 2019. https://doi.org/10. 1515/9783110941975.

Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.

Pushpa R, Raveendran TS. Optimization of shoot apex based cotton regeneration system. Electron J Plant Breed. 2010;1:1235–40.

Rajasekaran K. Agrobacterium-mediated genetic transformation of cotton. In: Curtis IS, editor.Transgenic crops of the world: essential protocols. Dordrecht, Netherlands: Springer; 2004. p. 243–54. https://doi.org/10. 1007/978-1-4020-2333-0_18.

Razzaghi M, Karimi A, Ansari Z, et al. Phenol removal by HRP/GOx/ZSM-5 from aqueous solution: artificial neural network simulation and genetic algorithms optimization. J Taiwan Inst Chem Eng. 2018;89:1–14. https:// doi.org/10.1016/j.jtice.2018.03.040.

Sakhanokho HF, Zipf A, Rajasekaran K, et al. Somatic embryo initiation and germination in diploid cotton (*Gossypium arboreum* L.). Vitr Cell Dev Biol. 2004;40:177–81. https://doi.org/10.1079/ivp2003497.

Salehi M, Farhadi S, Moieni A, et al. 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. 2021;17:1–13. https://doi.org/10.1186/s13007-021-00714-9.

Sharma A, Jain A, Gupta P, et al. Machine learning applications for precision agriculture: a comprehensive review. IEEE Access. 2020;9:4843–73. https://doi.org/10.1109/access.2020.3048415.

Silva JCF, Teixeira RM, Silva FF, et al. Machine learning approaches and their current application in plant molecular biology: a systematic review. Plant Sci. 2019;284:37–47. https://doi.org/10.1016/j.plantsci.2019.03.020.

Şimşek Ö, Dalda Şekerci A, Isak MA, et al. Optimizing micropropagation and rooting protocols for diverse lavender genotypes: a synergistic approach integrating machine learning techniques. Hortic. 2024;10:52. https://doi. org/10.3390/horticulturae10010052.

Sunilkumar G, Rathore KS. Transgenic cotton: factors influencing Agrobacterium-mediated transformation and regeneration. Mol Breed. 2001;8:37– 52. https://doi.org/10.1023/a:1011906701925.

- Tang Y, Chen L, Li XM, et al. Effect of culture conditions on the plant regeneration via organogenesis from cotyledonary node of cowpea (*Vigna unguiculata* L. Walp). Afr J Biotechnol. 2012;11:3270–5. https://doi.org/10. 5897/ajb11.3214.
- Van Rossum G, Drake FL. Python 3 reference manual (python documentation manual part 2). Scotts Valley, USA: CreateSpace Independent Publishing Platform; 2009.
- Wang XJ, Liu ST, Xia H, et al. Peanut (*Arachis hypogaea* L.) omics and biotechnology in China. Plant Omics. 2011;4:339–49.
- Webb GI, Sammut C, Perlich C, et al. Leave-one-out cross-validation. In: Sammut C, Webb GI, editors. Encyclopedia of machine learning. Boston, MA, USA: Springer; 2011. p. 600–1. https://doi.org/10.1007/978-0-387-30164-8_469.
- Whitmire CD, Vance JM, Rasheed HK, et al. Using machine learning and feature selection for alfalfa yield prediction. Al. 2021;2(1):71–88. https://doi.org/ 10.3390/ai2010006.
- Wu JA, Zhao FZ. Machine learning: an effective technical method for future use in assessing the effectiveness of phosphorus-dissolving microbial agroremediation. Front Bioeng Biotechnol. 2023;11:1189166. https://doi. org/10.3389/fbioe.2023.1189166.
- Yan SS, Ye L, Han S, et al. Speech interactive emotion recognition system based on random forest. International wireless communication and mobile computing conference (IWCMC). Limassol, Cyprus: IEEE; 2020. p. 1458–62. https://doi.org/10.1109/IWCMC48107.2020.9148117.
- Yang ZR, Li CF, Wang Y, et al. GhAGL15s, preferentially expressed during somatic embryogenesis, promote embryogenic callus formation in cotton (Gossypium hirsutum L.). Mol Genet Genom. 2014;289:873–83. https://doi.org/ 10.1007/s00438-014-0856-y.
- Yavuz C, Tillaboeva S, Bakhsh A. Apprehending the potential of BABY BOOM transcription factors to mitigate cotton regeneration and transformation. J Cotton Res. 2020;3:29. https://doi.org/10.1186/s42397-020-00071-3.
- Younis M, Ahmed IAM, Ahmed KA, et al. Pulsed electric field as a novel technology for fresh barhi date shelf-life extension: process optimization using response surface methodology. Hortic. 2023;9:155. https://doi.org/ 10.3390/horticulturae9020155.
- Zafar MM, Iqbal MS, Kamal H, et al. Advancing heat tolerance in cotton through integration of multiple stress tolerance indices and multivariate analyses. Int J Agric Biosci. 2024;13(1):65–75. https://doi.org/10.47278/ journal.ijab/2024.088.
- Zhang CJ, Yu SX, Fan SL, et al. Inheritance of somatic embryogenesis using leaf petioles as explants in upland cotton. Euphytica. 2011;181:55–63. https:// doi.org/10.1007/s10681-011-0380-7.