Journal of Plant Growth Regulation
https://doi.org/10.1007/s00344-022-10808-w

Machine Learning and Artificial Neural Networks-Based Approach to Model and Optimize Ethyl Methanesulfonate and Sodium Azide Induced In Vitro Regeneration and Morphogenic Traits of Water Hyssops (*Bacopa monnieri* L.)

Kubra Mirza[3] · Muhammad Aasim[1] · Ramzan Katırcı[2] · Mehmet Karataş[4] · Seyid Amjad Ali[5]

Received: 17 March 2022 / Accepted: 10 September 2022

Abstract
Application of chemical mutagens is used for artificially induced in vitro mutation to develop new cultivars with elite characteristics. However, the optimization of selecting proper mutagen, its concentration, and exposure time is of utmost importance, especially for plants containing noteworthy secondary metabolites. In this study, the effect of sodium azide (NaN_3) and ethyl methanesulfonate (EMS) in different concentrations (0.025, 0.05, 0.1, and 0.2 mg l^{-1}), and treatment time (30, 60, and 120 min) was investigated on *Bacopa monnieri;* an important medicinal plant. The maximum shoot counts (57.0) were achieved from the combination of 0.10 mg l^{-1} EMS × 60 min. Whereas, maximum shoot length (4.07 cm), node numbers (4.97) and leaf numbers (12,23) were achieved from the combination of 0.20 mg l^{-1} EMS × 120 min, respectively. Combination of 0.025 mg l^{-1} NaN_3 × 120 mg/l yielded maximum shoot counts (52.30), shoot length (3.23 cm), node numbers (6.07) and leaf numbers (12.13). The trained model to predict the outputs were designed and calibrated with machine learning (ML) algorithms. Support Vector Classifier (SVC), Gaussian Process (GP), Extreme Gradient Boosting (XGBoost), Random Forest (RF) models, and Multilayer Perceptron (MLP) neural network algorithms were used to discover the best models and their hyperparameters. The RF model gave exceptional results in the prediction of the outputs. F1 scores of the RF were acquired in the range of 0.98–1.00 for different outputs. The other models' F1 scores varied in the range of 0.65 and 0.85. The present work opens the new era of applying ML and artificial neural network (ANN) models in plant tissue culture with the possibility of application for other economic crops.

Keywords Artificial neural network · Bacopa · Machine learning · Mutagens

Introduction

Bacopa monnieri (L.) is an important medicinal plant that has been used extensively in the Ayurvedic system. It is native to both India and Australia (Aguiar and Borowski 2013). The plant is a succulent semi-aquatic creeping herb, used as complementary and alternative medicines (CAM), since ancient times in the traditional Ayurvedic medicinal system (Kean et al. 2017) for curing neurological and neuropsychiatric diseases. The importance of Bacopa plant is mainly due to the presence of Bacoside A (triterpenoid saponin) a major bioactive compound (Sivaramakrishna et al. 2005) in recent years, due to extensive research work on Alzheimer's disease (Chaudhari et al. 2017) and other major illnesses and disorders (Aasim et al. 2019). Due to its economic and scientific significance, biotechnological

Handling Editor: Tariq Aftab.

✉ Muhammad Aasim
mshazim@gmail.com

[1] Department of Molecular Biology and Genetics, Faculty of Science, Necmettin Erbakan University, Konya, Turkey

[2] Department of Plant Protection, Faculty of Agricultural Science and Technologies, Sivas University of Science and Technology, Sivas, Turkey

[3] Department of Metallurgical and Materials Engineering, Faculty of Engineering and Natural Sciences, Sivas University of Science and Technology, Sivas, Turkey

[4] Department of Biotechnology, Faculty of Science, Necmettin Erbakan University, Konya, Turkey

[5] Department of Information Systems and Technologies, Bilkent University, Ankara, Turkey

Published online: 09 October 2022

and molecular biological techniques along with classical breeding have been employed for its improvement with the foremost focus on Bacoside A.

In vitro mutation breeding is a simple, reliable, and feasible method that is commonly practiced to induce genetic variation toobtain mutants with desired characteristics. Spontaneous mutations and induced mutations are two common types of mutations that are used for developing mutants of desired characteristics. Exposing different plants to different physical or chemical mutagens under in vitro conditions (Suprasanna et al. 2014; Bado et al. 2015) leads to enhanced spontaneous mutation (Bairu et al. 2011), and that phenomenon is considered as 'somaclonal variation' (GMO et al. 2021). The occurrence of in vitro somaclones using mutagens, known as mutation breeding (Kharkwal 2012) is the major source of genetic diversity and can be employed for breeding programs (Bairu et al. 2011). Application of in vitro random mutagenesis offers certain advantages like the uniformity of the treatment, application of selective agent, screening large population in less time and space, and easy handling (Suprasanna et al. 2012) with a relatively high frequency of induced mutants (Suprasanna et al. 2014) as compared to in vivo mutagenesis. The development of chimerism due to random mutagenesis is another major issue, and it can be eliminated or reduced using plant tissue culture techniques (Jankowicz-Cieslak and Till 2017). The final step is the screening of mutants and can be estimated using in vitro regeneration parameters, morphogenic characteristics of in vitro induced shoots (Rayan et al. 2014), or acclimatized plantlets and estimation of genetic diversity using different molecular biology techniques like employing molecular markers (Rayan et al. 2014; Aasim et al. 2019).

Recent advances in the field of machine learning (ML) and artificial neural networks (ANN) allow researchers to apply computer-based technologies for precision agriculture (Sharma et al. 2020). The use of computer-based AI technologies under field conditions for precision agriculture has been documented for soil characteristics to estimate or predict crop yield, predicting climate, insect infestation forecasting, and other areas of commercial farming. Recently, the application of ML modeling in plant biotechnology is also prevailing but with limited reports, especially in the field of plant tissue culture, where experiments are performed under in vitro conditions. To date, in vitro sterilization (Hesami et al. 2019), in vitro germination (Hesami et al. 2021), in vitro elicitation (Salehi et al. 2020), in vitro cell culture (Farhadi et al. 2020; Kirtis et al. 2022), and in vitro somatic embryogenesis (Hesami et al. 2020b) have been optimized using different ML algorithms. In these studies, different models like Multilayer Perceptron (MLP), Generalized Regression Neural Network (GRNN), Radial Basis Function (RBF), Random Forest (RF) have been favored by the researchers (Gago et al. 2010; Hesami and Jones 2020; Zhang et al. 2020). The selection of a proper model depends on the association between input and output variables (Hesami et al. 2021), and the optimization of hyperparameters. The objective of this study is to generate the best model predicting the in vitro organogenesis and morphological indices of regenerated shoots induced by exposing explants to two different mutagens at different concentrations and exposure times. Thus, support vector classifier (Metlek and Kayaalp 2020), gaussian process (S Ad et al., 2018; Hu et al. 2019), XGBoost (Katirci et al. 2021), and random forest (Yan et al. 2020) ML algorithms and multilayer perceptron neural network (Katırcı et al. 2021) algorithm were employed. The treatment of Sodium azide (NaN_3) and Ethyl methanesulfonate (EMS) concentration and treatment time were used as input variables. Whereas different in vitro regeneration parameters (regeneration frequency, shoots per explant, shoot length), and morphological indices (number of nodes, internode length, number of leaves) were specified as outputs.

Methods and Materials

In Vitro Mutagen Experiment

Bacopa monnieri L. used as plant material in this study was procured from the Biotechnology laboratory, Faculty of Science, Necmettin Erbakan University, Turkey. Sodium azide (NaN_3) and Ethyl methanesulfonate (EMS) was used as a chemical mutagen. The stock solution of both mutagens was prepared at the rate of 1.0 mg/ml. The leaf explants were isolated aseptically from in vitro stock material and exposed to the solution including (0.025, 0.05, 0.10, and 0.20 mg l^{-1}) NaN_3.

(Sigma Aldrich, CAS: 26628-22-8) and EMS (Sigma Aldrich, CAS: 62-50-0) mutagens for 30, 60, and 120 min. After subsequent treatment, explants were cultured in the MS medium enriched with 1.0 mg l^{-1} BAP (Karataş and Aasim 2014). The standard basal medium for in vitro regeneration was prepared using 0.44% MS and gelled with 0.65% agar. The pH of the medium was adjusted to 5.8 using 1 N NaOH or 1 N HCl. The basal medium was sterilized at 1.2-atmosphere pressure and 121 °C. The explants were cultured in a place equipped with White Light-emitting Diodes at 24 ± 1 °C and 16 h light photoperiod. Explants were cultured for 14 weeks on basal medium and thereafter, in vitro organogenesis and morphological traits of regenerated shoots were specified as output. The phenotype data (number of leaves, number of nodes, internode length) for somaclonal variation were also recorded as outputs. For phenotype data (node number, internode length, and leaf numbers), randomly selected 10 shoots per replication were used.

Machine Learning Models

The experimental data were assessed using SVC, GP, XGB, RF machine learning models, and MLP neural networks. The model was generated by training the data with ML algorithms to predict the output variables. Different levels (concentration and treatment time) of sodium azide (NaN_3) and Ethyl methanesulfonate (EMS) were used as inputs, and different in vitro regeneration and growth parameters presented in Table 1 were used as the outputs. The leave-one-out cross-validation (LOO-CV) method (Qi et al. 2019), was used to validate and measure the performance of different models via test data. This method of training and cross-validation is generally preferred for small datasets as it becomes computationally expensive with large datasets. The confusion matrix was used to summarize the performance of the employed classification algorithms. The optimum value for the hyperparameters was obtained by performing a grid search technique. The coding was performed with Python language. Sklearn library was used for model training, fitting, and testing. Before training the model for the dataset, the threshold which reflects the superiority of the outputs in response to the treatments was specified to classify the target outputs (Table 1). The outputs above the given threshold levels were coded as 1, whereas values below the threshold levels were coded as 0.

Support Vector Classifier (SVC)

The SVC is a supervised machine learning algorithm that is inspired by statistical learning theory. The major benefit of using this algorithm is due to its ability to run with small datasets with high performance. In general, the application of traditional artificial intelligence algorithms needs a lot of training data, getting stuck in a local minimum, relatively low convergence rate, and overfitting may be encountered in other algorithms. The SVC algorithm can easily overcome these issues. The function of the SVC algorithm is presented in Eq. 1, which tries to detect the farthest separator line among the classes (Metlek and Kayaalp 2020).

$$f(x) = (w, x) + b \tag{1}$$

Table 1 Targeted thresholds of the in vitro organogenesis and morphological traits of regenerated shoots for data training

The outputs	Thresholds
In vitro organogenesis	
Regeneration frequency (%)	50
Shoots per explant (numbers)	10
Mean shoot length (cm)	1
Morphological traits of regenerated shoots	
Mean node numbers (numbers)	2
Mean internodal length (cm)	0.5
Mean leaf numbers(numbers)	5

Gaussian Process (GP)

The GP is another powerful supervised learning algorithm that can solve both classification and regression problems. The Gaussian probability density function is used to describe the random variable distribution. The GP classifier (non-parametric algorithm), is applied to a binary dataset, and it calculates the probability of any input sample belonging to a certain class. It runs with small datasets, provides accuracy, ease of calculation, and consistency altogether (Hu et al. 2019). The function used to calculate the relation between an input x and output y is illustrated in Eq. 2.

$$y_i = f(x_i) + \varepsilon \tag{2}$$

Extreme Gradient Boosting (XGBoost)

XGBoost is another popular algorithm that is extensively used for regression and classification problems (Chen and Guestrin 2016). The XGBoost model is an algorithm of Gradient Boosting Decision Tree, designed for performance and speed, and it utilizes a gradient boosting framework. The principal advantage of the XGBoost is that it learns from errors and lessens the error rate via iteration. The XGBoost objective function and iterative model are represented in Eqs. 3 and 4, respectively (Katirci et al. 2021).

$$y_i = F(x_i) = \sum_{(d=1)}^{D} f_d(x_i), f_d \in F, i = 1, \ldots, n \tag{3}$$

$$L_j = \sum_{(i=1)}^{n} l\left(y_i, \hat{y}_i^{(j-1)} + f_j(x_i)\right) + \Omega(f_j) \tag{4}$$

Random Forest (RF)

RF is a decision tree-based supervised machine learning algorithm that can be utilized for both regression and classification problems. Decision trees produce the forest, and each tree owns the same distribution. A random sampling of data is used for calculation and feature selection for each decision tree during training. The RF algorithm prevents overfitting and provides lower noise performance (Yan et al. 2020).

Multilayer Perceptron (MLP)

MLP is a feedforward artificial neural network that consists of more than one perceptron, which utilizes a nonlinear

activation function besides the input nodes. It is composed of fully interconnected three layers namely (i) input layer, (ii) one or more hidden layers, and (iii) an output layer. The input layer occurs from the input of the dataset, while the output layer is composed of a single or more neurons representing the class number of outputs. MLP is often executed in supervised learning tasks. The backpropagation method is utilized to modify the weights and biases to reduce the error (Katırcı et al. 2021).

Confusion Matrix

The confusion matrix map was used to evaluate the employed models using values of TP, TN, FP, and FN metrics. (Table 2). Accuracy, recall, precision, and F1 scores were computed in accordance with Eqs. 5–8. F1 score is a powerful metric to select the best model as it considers precision and recall metrics together.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

Data Analysis

The analysis of data and graphs regarding in vitro regeneration and morphological traits was performed using SPSS 20.0 program via one-way ANOVA (analysis of variance). DMRT test was used for comparing the means (post hoc test) and data with standard error are presented as supplementary material. The arcsine square root transformation was used before performing the statistical analysis of the data. All contour plots were generated with the aid of Minitab 15.0 program.

Table 2 The confusion matrix table

Confusion Matrix		Y-predicted	
		0	1
Y-true	0	True negative (TN)	False positive (FP)
	1	False negative (FN)	True positive (TP)

Results

Impact of Mutagens Type, Concentration, and Exposure Time

Two different chemical mutagens used in this study revealed a significant impact on shoot regeneration frequency (*p0.01*), shoot counts (*p0.05*), and internode length (*p0.05*). Shoot length, node numbers, and leaf numbers were found to be statistically insignificant. Comparing mutagen type, EMS induced more shoot regeneration frequency (76.11), shoot counts (30.15), shoot length (2.10 cm), internode length (0.83 cm), and leaf numbers (6.35) when compared to NAN_3. Whereas EMS induced more node numbers (2.74) compared to EMS (2.33) (Fig. 1, Table S1).

Results on type of mutagens, concentration, and treatment time revealed a significant impact on shoot regeneration frequency ($p < 0.01$), shoot length ($p < 0.05$), node numbers ($p < 0.05$), and leaf numbers ($p < 0.05$). On the other hand, statistically insignificant impact was observed on shoot counts and internode length. Results on shoot regeneration frequency revealed the variable response of both mutagens, and increased with increased concentration and exposure time of EMS. On the contrary, elevated NaN_3 concentration exerted a negative impact on shoot regeneration frequency. Shoot regeneration frequency of EMS and NaN_3 treated explants ranged 33.33–100%, and 0–93.33%, respectively (Fig. 2, Table S2). Results on shoot counts and shoot length exhibited a similar response to both mutagens treatments. In general, elevated mutagen concentration and treatment time resulted in enhanced shoot counts and shoot length. Shoot counts ranged 1.05–57 and 0.00–52.30 for EMS and NaN_3, respectively (Fig. 2, Table S2). Whereas shoot length was recorded as 0.34–4.07 cm (EMS) and 0.00–3.23 cm (NaN_3). Maximum shoot counts and shoot length were recorded on same treatments of highest concentration and treatment time (0.025 mg l^{-1} + 120 min, 0.05 mg l^{-1} + 120 min, 0.20 mg l^{-1} + 120 min) for both mutagens. However, the application of 0.10 mg l^{-1} of EMS and NaN_3 generated maximum shoot counts and shoot length with 60 min treatment (Fig. 2, Table S2).

Results on node number revealed the effect of mutagen type, concentration, and treatment time. Treating leaf explants to EMS required high concentration and treatment time for inducing more node numbers. On the contrary, NaN_3 induced relatively high node numbers when leaf explants were treated with low NaN_3 concentration for a prolonged treatment time of 60 to 120 min. Maximum node number of 4.97 from EMS treated explants was attained from 0.20 mg l^{-1} EMS + 120 min treatment (Fig. 3, Table S3). On the other hand, a maximum

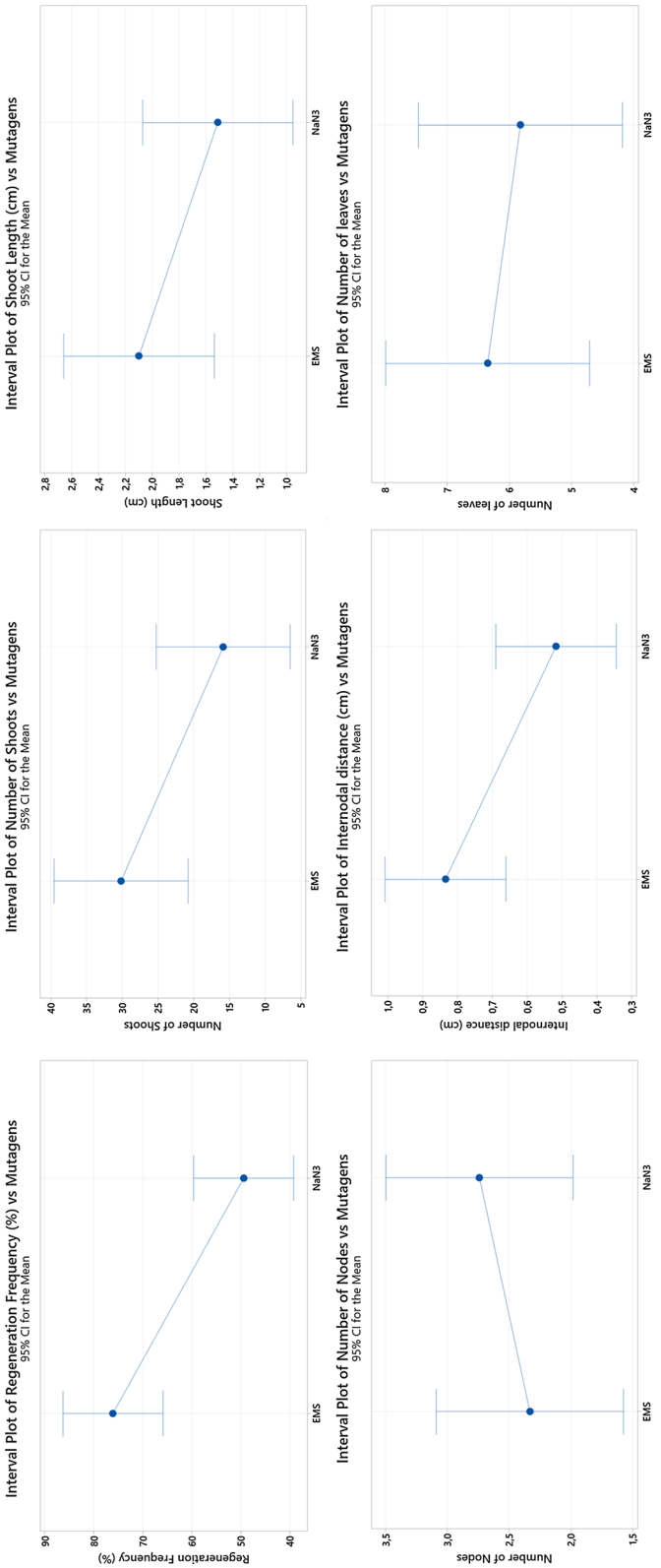

Fig. 1 Impact of mutagens type on in vitro shoot regeneration traits of *B. monnieri*

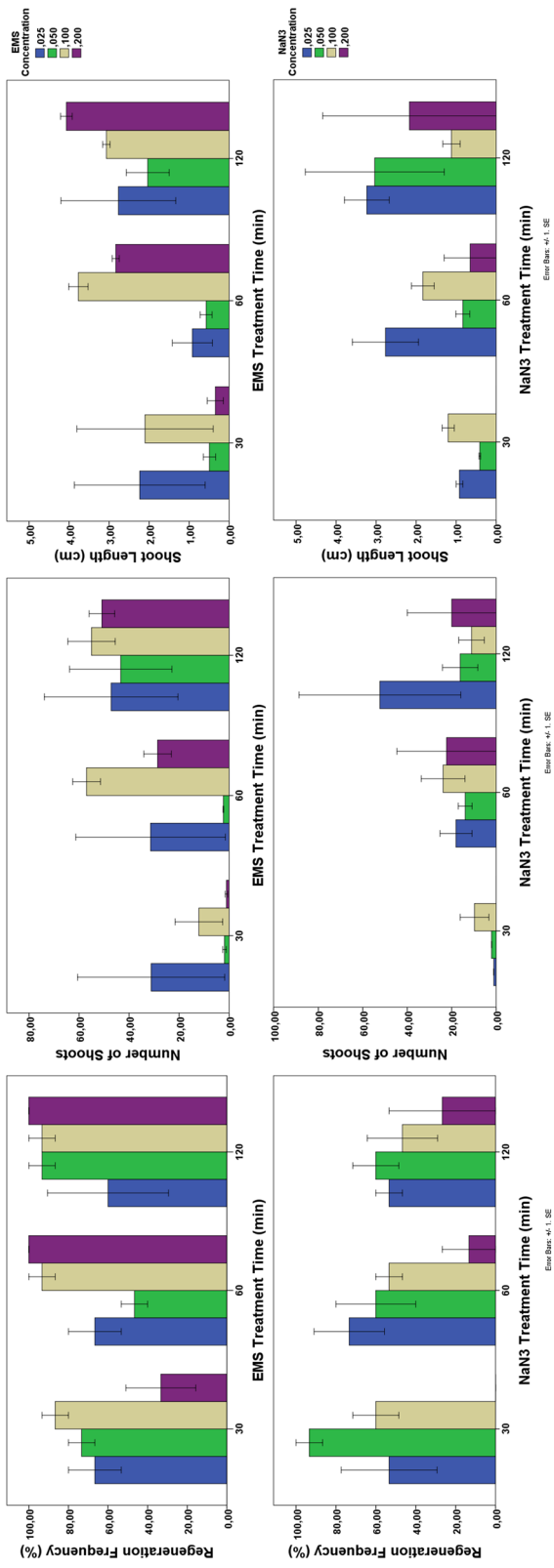

Fig. 2 Impact of different mutagens, concentration and treatment time on in vitro shoot regeneration traits of *B. monnieri*

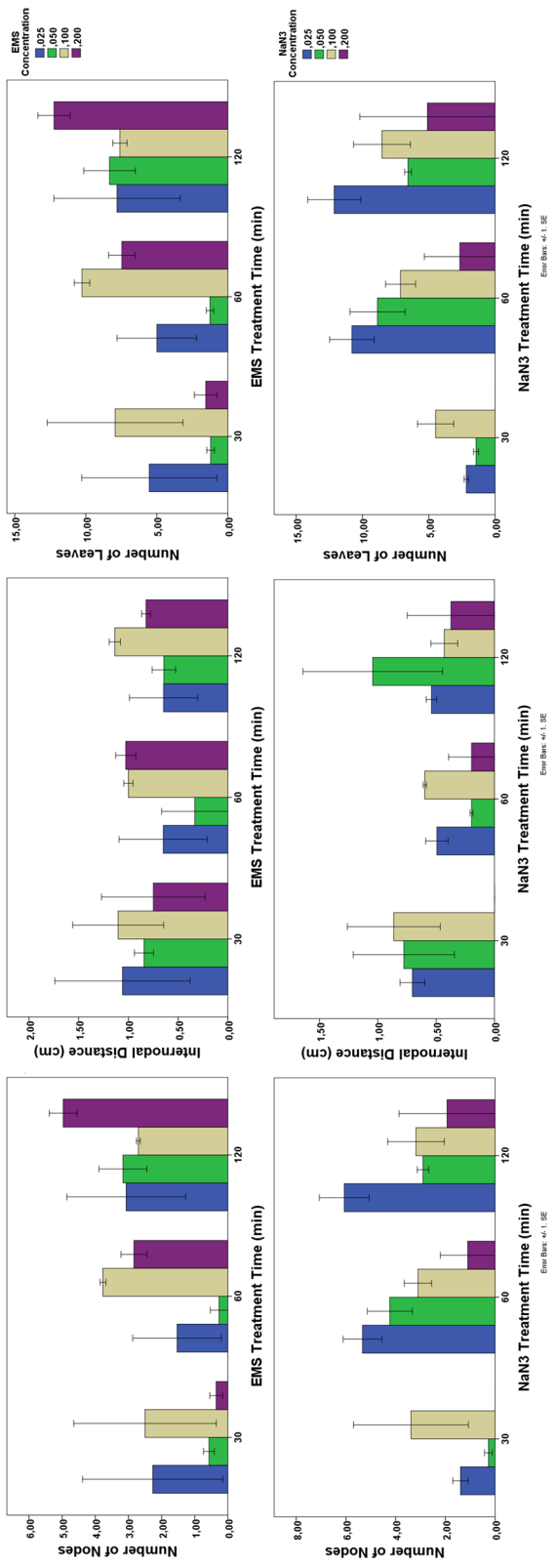

Fig. 3 Impact of different mutagens, concentration and treatment time on morphogenic traits of in vitro induced shoots of *B. monnieri*

of 6.07 node numbers were attributed to 0.025 mg l^{-1} NaN_3 + 120 min followed by 5.33 shoots (0.025 mg l^{-1} NaN_3 + 60 min). The internode length was statistically insignificant and in general, treating explants with EMS induced more internode length compared to NaN_3 and recorded as 0.33–1.14 cm (EMS) and 0.00–1.05 cm (NaN_3). Maximum internode length was recorded as 1.14 cm (0.10 mg l^{-1} EMS + 120 min) and 1.05 cm (0.05 mg l^{-1} EMS + 120 min). Results on leaf numbers were linked with node numbers and followed the same pattern. Leaf numbers ranged 1.21–12.23 and 0.00–12.13, respectively, for EMS and NaN_3. Maximum leaf numbers for EMS-treated explants were recorded from 0.20 mg l^{-1} EMS + 120 min treatment. Whereas maximum leaf numbers (12.13) were recorded as from 0.025 mg l^{-1} NaN_3 + 120 min treated explants (Fig. 3, Table S3).

Contour plots of in vitro regeneration and morphogenic traits of in vitro induced shoots are given in Figs. 4 and 5, respectively. Results clearly showed the impact of mutagen type, concentration, and treatment time and contour plots distributed the data regarding regeneration frequency (%), number of shoots, shoot length (Fig. 4), number of nodes, internode length, and number of leaves (Fig. 5) into their respective sub-groups and highlighted with a different color. Comparing mutagen types, EMS was most responsive and contour plots revealed the maximum output variable optimization compared to NaN_3. Supplementation of high EMS concentration of approximately 0.17 mg/l or above with exposure time of 110 min or above may be required for the maximum output parameters. On the other hand, supplementation of low concentration of around or below 0.05 mg/l NaN_3 with exposure time of 110 min or above may lead to maximum output variables. It is evident from the results that prolonged exposure time for both mutagens with all concentration responded better, and yielded better regeneration and with relative better response on morphological characteristics of in vitro induced shoots. Figure 6a–h presents the impact of different mutagens types, concentration and exposure time of leaf explants.

Machine Learning Algorithms

The confusion matrix is a powerful metrics tool used for measuring the accuracy of classification when training ML models. As a result, true negatives (TN), false positives (FP), false negatives (FN), and true positives (TP) were obtained. These values were utilized to compute the parameters of accuracy, F1 score, precision, and recall. Table 3 indicates the confusion matrix of the models for in vitro organogenesis (regeneration frequency, shoots per explant, mean shoot length). Results for morphological traits (mean internode length, mean node numbers, and mean leaf numbers) of in vitro regenerated shoots are presented in Table 4. Among the models tested, the RF algorithm gave the best results and estimated all outputs of morphological traits of in vitro regenerated shoots correctly. In the RF model, the values of FN and FP are 0 for shoots per explant, mean shoot length, mean internode length, node numbers, and leaf numbers. However, the RF model predicted 1 FN for regeneration frequency (Table 3). RF model estimated all outputs (accuracy, F1 score, precision and recall) without error, while regeneration frequency was computed as 0.99 (accuracy), 0.99 (F1), 1 (precision), and 0.98 (recall) values. MLP model ranked second among tested models for all output variables.

Comparison of individual output variables responded in variable way using accuracy, F1 scores, precision and recall values of all tested models. The RF model ranked first for all outputs followed by MLP model. The position of remaining three models changed with output variables. Accuracy values of tested models were registered in similar fashion, and RF model yielded maximum accuracy values followed by MLP model. The RF model yielded all accuracy values as 1, except shoot regeneration frequency which was recorded 0.99. Accuracy values of individual parameters were registered in the order of RF (0.99) > MLP (0.78) > SVC (0.75) > XGBoost (0.71) > GP (0.69) for shoot regeneration frequency (Table 3); RF (1) > MLP (0.81) > SVC (0.75) > GP (0.74) > XGBoost (0.69) for shoot counts (Table 3); RF (1) > MLP (0.79) > GP (0.76) > SVC (0.74) > XGBoost (0.65) for shoot length (Table 3); RF (1) > MLP (0.82) > GP (0.81) > SVC (0.79) > XGBoost (0.69) for numbers of nodes (Table 4); RF (1) > MLP (0.78) > SVC (0.76) > GP (0.72) > XGBoost (0.69) for internode length (Table 4); and RF (1) > MLP (0.82) > GP (0.76) > SVC (0.74) > XGBoost (0.69) for leaf numbers (Table 4).

The F1 score of shoots per explant, shoot numbers, mean internode length, mean node numbers and mean leaf numbers were recorded 1, which reflects that RF model estimated all the test samples without error. The performance scores of the models for regeneration frequency (Table 3) followed the order of RF (0.99) > SVC (0.84) > MLP (0.83) > XGBoost (0.80) > GP (0.79). The RF model predicted all 0's correctly, but it predicted 1 as 0 in only single sample. The F1 scores for shoots per explant (Table 3) were acquired in the order of RF (1) > MLP (0.81) > SVC (0.72) = GP (0.72) > XGBoost (0.69). The F1 scores of ML algorithms for mean shoot length (Table 3) were recorded as RF (1) > MLP (0.81) > GP (0.77) > SVC (0.75) > XGBoost (0.68). The data on mean node numbers (Table 4) exhibited that RF model predicted the output correctly and, overall acquired in the order of RF (1) > MLP (0.84) > GP (0.83) > XGBoost (0.82) > SVC (0.81). The F1 scores of internode length (Table 4) were in the order of RF (1) > MLP (0.85) > SVC (0.83) > GP (0.82) = XGBoost (0.82). The F1 scores of the models for

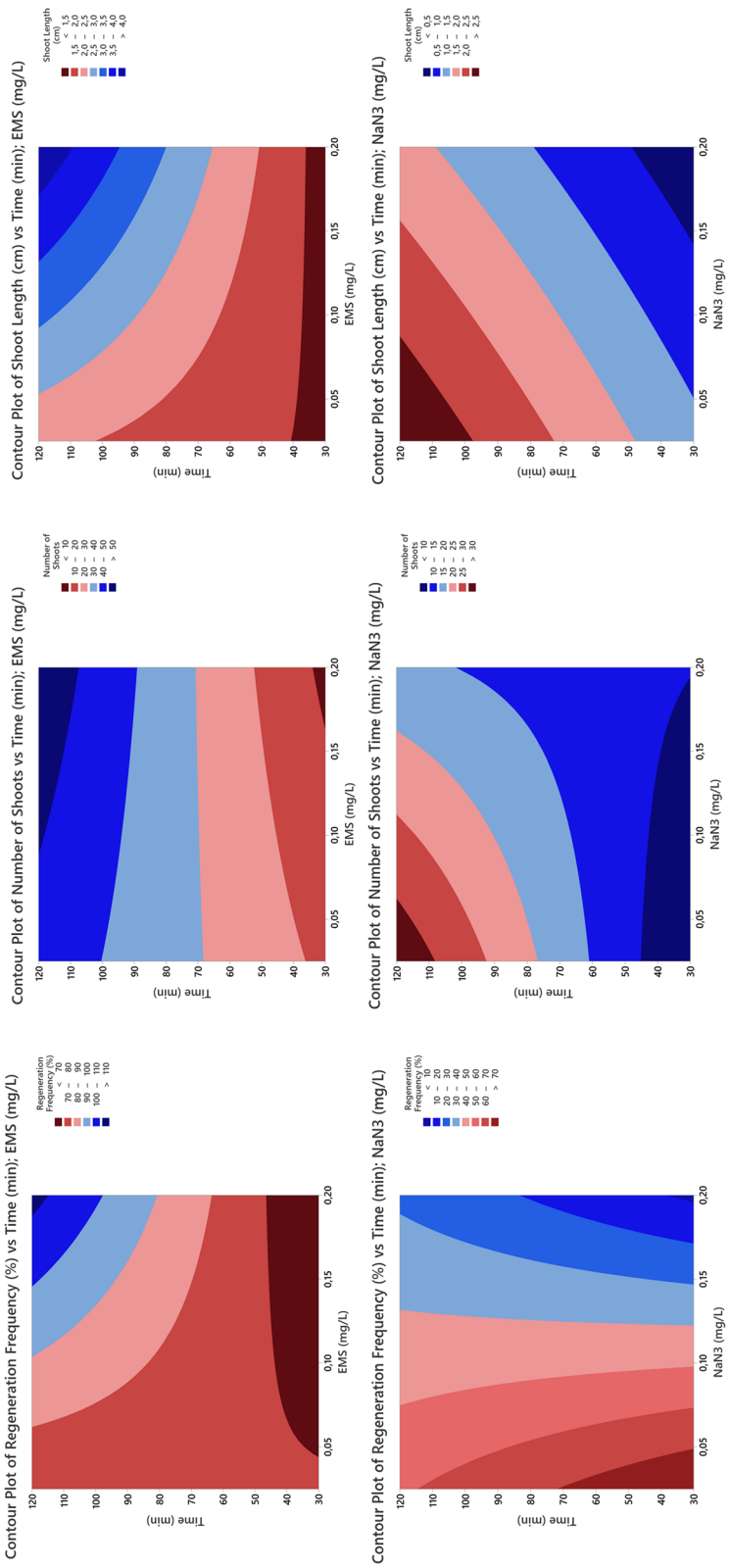

Fig. 4 Contour plots of in vitro shoot regeneration traits of *B. monnieri*

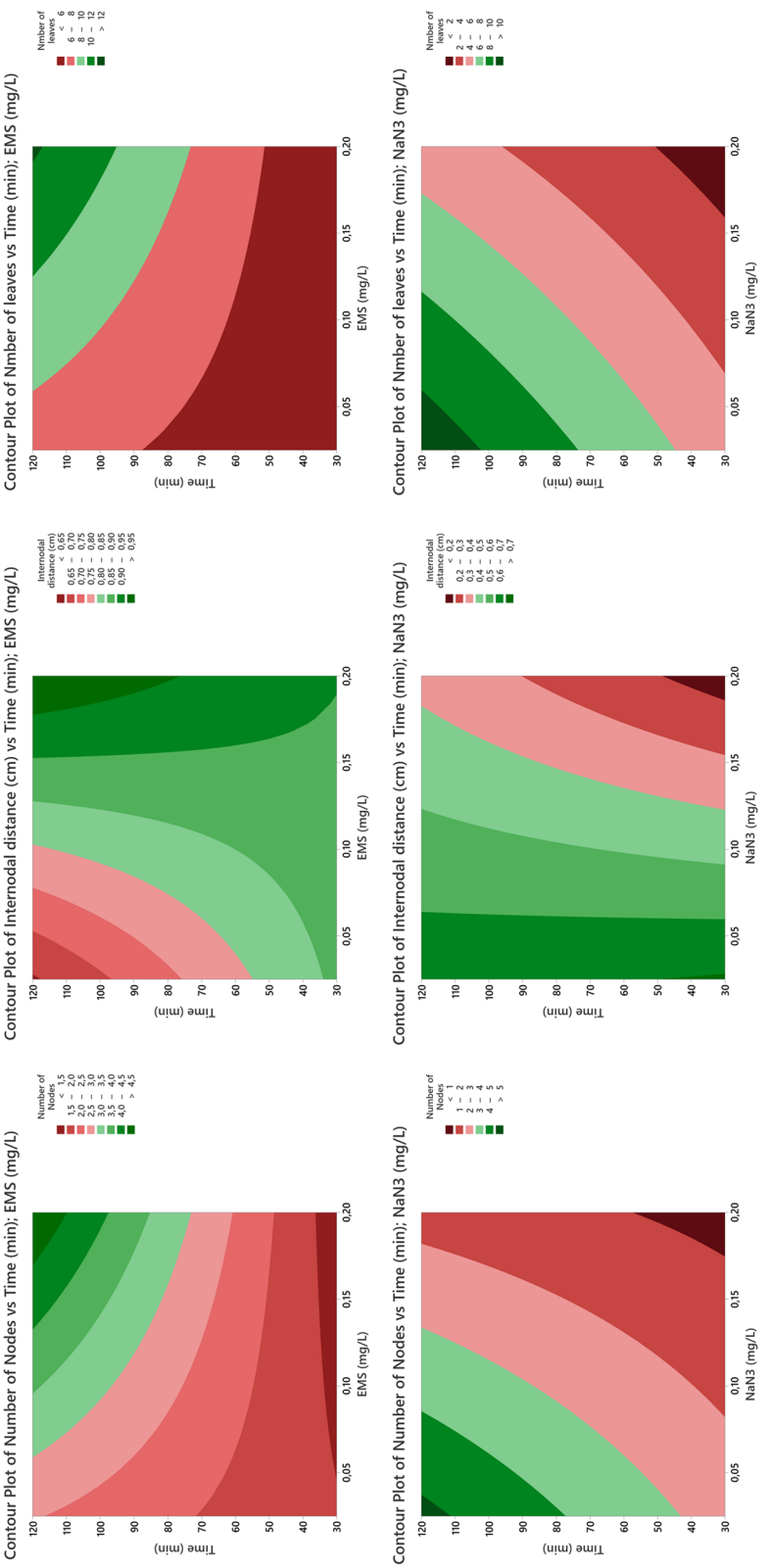

Fig. 5 Contour plots of morphogenic traits of in vitro induced shoots of *B. monnieri*

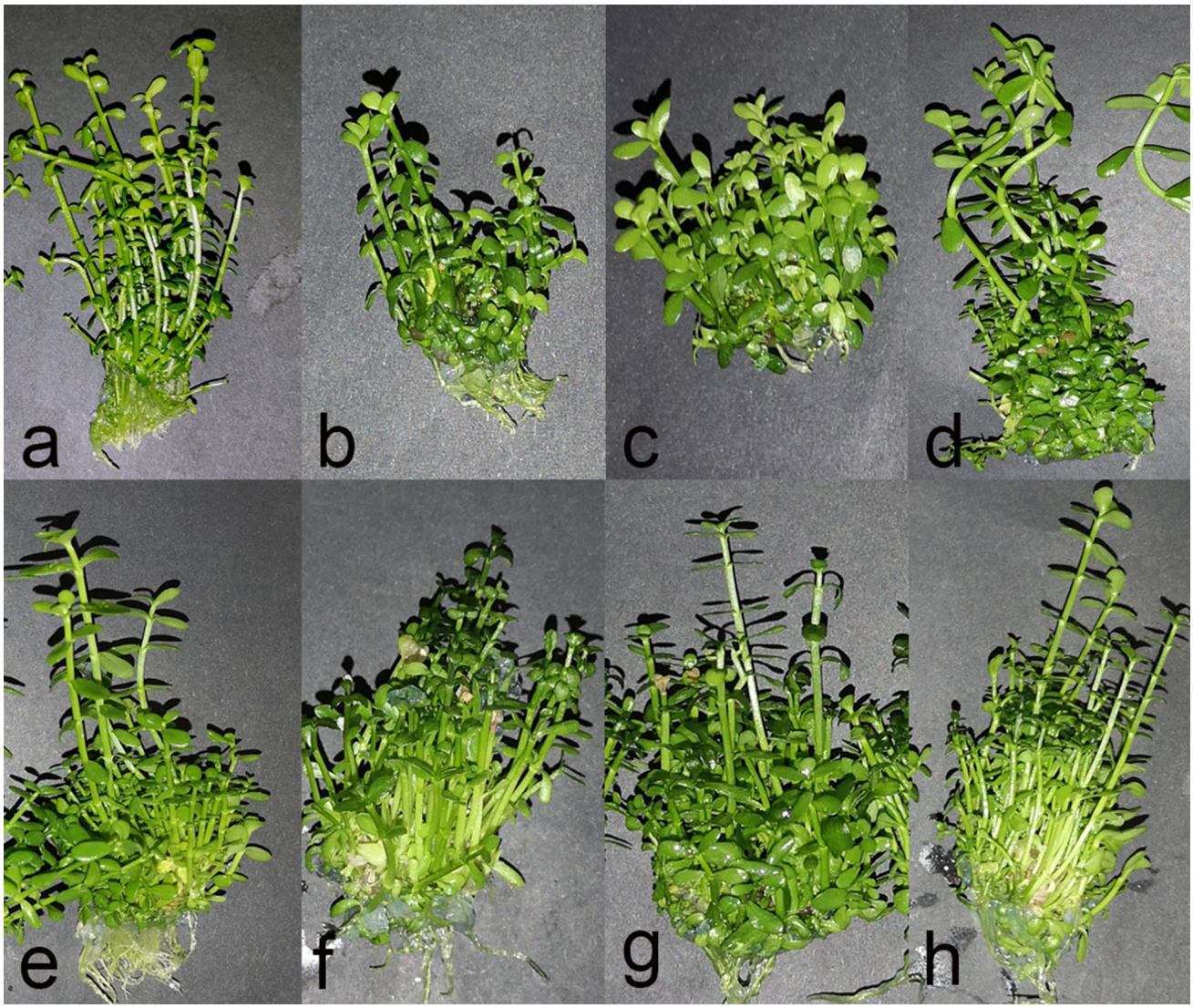

Fig. 6 In vitro multiple shoot induction from leaf explants exposed to 120 min at different mutagens and concentration **a** 0.025 mg l^{-1} EMS **b** 0.05 mg l^{-1} EMS, **c** 0.10 mg l^{-1} EMS, **d** 0.20 mg l^{-1} EMS **e** 0.025 mg l^{-1} EMS **f** 0.05 mg l^{-1} EMS, **g** 0.10 mg l^{-1} EMS, **h** 0.20 mg l^{-1} EMS

mean leaf numbers (Table 4) were recorded in the order of RF (1) > MLP (0.84) > GP (0.80) > SVC (0.77) > XGBoost (0.73).

Results on precision also highlighted the supremacy of RF models, which were followed by MLP model. Among the tested models, XGBoost was least effective for all output variables. The precision values of in vitro regeneration traits variables (Table 3) were registered as shoot regeneration frequency (0.72–1), shoot counts (0.75–1), and shoot length (0.76–1). The precision scores of in induced shoots (Table 4) were registered as 0.69–1 for both internode length and number of nodes, and 0.71–1 for the leaf numbers. Recall values exhibited variable scores, with RF model ranked first for all output variables (Tables 3, 4). Whereas, the performance of remaining four models changed with respective to output variable. Recall scores for shoot regeneration frequency was scored in the order of RF (0.98) = SVC (0.98) > GP (0.87) = XGBoost (0.87) > MLP (0.85). Shoot counts scores were recorded as RF (1) > MLP (0.83) > GP (0.69) = XGBoost (0.69) > SVC (0.66). Whereas, recall scores of shoot length were in accordance of RF (1) > MLP (0.84) > GP (0.76) > SVC (0.74) > XGBoost (0.68). Outcomes on morphogenic traits were recorded as RF (1) = XGBoost (1.0) > GP (0.85) = MLP (0.85) > SVC (0.82) for number of nodes, RF (1) = XGBoost (1.0) > GP (0.94) > MLP (0.90) > SVC (0.86) for intermode length, and RF (1) > MLP (0.85) = GP (0.85) > SVC (0.80) > XGBoost (0.75).

Table 3 Confusion matrix and performance score of the ML algorithms models trained for in vitro shoot regeneration traits

Confusion Matrix		Y-predicted		Shoot regeneration frequency (%)			
Model	Y-true	0	1	Accuracy	F1 score	Precision	Recall
MLP	0	16	9	0.81	0.81	0.78	0.83
	1	7	40				
SVC	0	8	17	0.75	0.72	0.79	0.66
	1	1	46				
GP	0	9	16	0.74	0.72	0.75	0.69
	1	6	41				
XGBoost	0	10	15	0.69	0.69	0.69	0.69
	1	6	41				
RF	0	25	0	1	1	1	1
	1	1	46				
Confusion Matrix		Y-predicted		Number of shoots			
MLP	0	29	6	0.81	0.81	0.78	0.83
	1	8	29				
SVC	0	31	12	0.75	0.72	0.79	0.66
	1	6	23				
GP	0	29	8	0.74	0.72	0.75	0.69
	1	11	24				
XGBoost	0	26	11	0.69	0.69	0.69	0.69
	1	11	24				
RF	0	37	0	1	1	1	1
	1	0	35				
Confusion Matrix		Y-predicted		Shoot length (cm)			
MLP	0	25	9	0.79	0.81	0.78	0.84
	1	6	32				
SVC	0	25	9	0.74	0.75	0.76	0.74
	1	10	28				
GP	0	26	8	0.76	0.77	0.78	0.76
	1	9	29				
XGBoost	0	21	13	0.65	0.68	0.67	0.68
	1	12	26				
RF	0	34	0	1	1	1	1
	1	0	38				

Y-true 0 = TN; Y-true 1 = FN; Y-predicted 0 = TP; Y-predicted 1 = FP

Discussion

The application of mutagens under in vitro conditions is highly significant for inducing mutations. Type of mutagens, concentration, treatment time, time of application, plant type, and explants are some eminent parameters for inducing in vitro mutants (Spencer-Lopes et al. 2018; Aasim et al. 2019; GMO et al. 2021). Bacopa is an important medicinal plant due to containing Bacoside and limited studies revealed the use of chemical mutagens like EMS, MMS (Vajpayee et al. 2006), and colchicine (Kharde et al. 2017). These studies revealed the significant role of mutagens type and concentration on in vitro regeneration and organogenesis parameters. In this study, mutagens type and concentration on in vitro organogenesis and morphogenic traits of in vitro induced shoots is conducted in detail. Results illustrated the variable impact of both mutagens on in vitro regeneration frequency. The results are contrary to the findings on apple rootstock, where decreased regeneration frequency in response to mutagens was observed (Rayan et al. 2014), with possibility of different input treatments and genotypes. Whereas, enhanced shoot counts and shoot length were observed accordingly with an increase of concentration and exposure time. Results further exhibited that both in vitro

Table 4 Confusion matrix and performance score of the ML algorithms models trained for in vitro morphogenic traits of in vitro induced shoots

Confusion matrix		Y-predicted		Number of nodes			
Model	Y-true	0	1	Accuracy	F1 score	Precision	Recall
MLP	0	26	7	0.82	0.84	0.82	0.85
	1	6	33				
SVC	0	25	8	0.79	0.81	0.8	0.82
	1	7	32				
GP	0	25	8	0.81	0.83	0.8	0.85
	1	6	33				
XGBoost	0	0	22	0.69	0.82	0.69	1
	1	0	50				
RF	0	33	0	1	1	1	1
	1	0	39				
Confusion matrix		Y-predicted		Internodal distance (cm)			
MLP	0	11	11	0.78	0.85	0.80	0.90
	1	5	45				
SVC	0	12	10	0.76	0.83	0.81	0.86
	1	7	43				
GP	0	5	17	0.72	0.82	0.73	0.94
	1	3	47				
XGBoost	0	0	22	0.69	0.82	0.69	1
	1	0	50				
RF	0	22	0	1	1	1	1
	1	0	50				
Confusion matrix		Y-predicted		Number of leaves			
MLP	0	25	7	0.82	0.84	0.83	0.85
	1	6	34				
SVC	0	21	11	0.74	0.77	0.74	0.80
	1	8	32				
GP	0	21	11	0.76	0.80	0.76	0.85
	1	6	34				
XGBoost	0	20	12	0.69	0.73	0.71	0.75
	1	10	30				
RF	0	32	0	1	1	1	1
	1	0	40				

Y-true 0 = TN; Y-true 1 = FN; Y-predicted 0 = TP; Y-predicted 1 = FP

regeneration and morphological traits responded in variable ways to both mutagens and their concentration. The results on node numbers revealed the need for high NaN_3 and low EMS concentration along with more exposure time. However, internode length remained statistically insignificant to all mutagens and treatment time. On the contrary, variable response of both mutagens were observed on leaf numbers (Rayan et al. 2014). The results evidently exhibited that both mutagens, concentration and treatment time exerted variable impact on morphological traits of in vitro induced shoots of *B. monnieri*. Previous results on using individual mutagens (EMS, NaN_3) also revealed the variable impact on the morphological traits of other plants under in vitro conditions. Variable internode length of in vitro induced mutants (plantlets) of *Asteracantha longifolia* (L.), exposed to different concentrations and treatment time of EMS has also been registered (Behera et al. 2012).

The impact of mutagens type and concentration was also analyzed by 2D contour plots. Contour plots are powerful tools that classified the results into different subunits and can be used to optimize the desired traits. Results depicted the distribution of output variables into different sub-classes, and can be used to optimize the results or interactions between two input variables (Kelly et al. 2021). The results

confirmed that it is possible to obtain in vitro somaclones by exposing explants to both mutagens but need more work to check them on molecular or phytochemical basis. Most importantly, in this study we used machine learning algorithms to validate the results using five different ML or ANN models using a confusion matrix system.

Application of ML models in the field of agriculture and plant sciences is prevailing for better prediction and optimization of the collected data. Some of the examples include predicting insects based on data collection on environmental factors like temperature and humidity (Sagar et al. 2017) or counting the number of insects trapped in a certain period (Skawsang et al. 2019). Similarly, edaphic, climatic, water, agronomic, and crop properties using ML algorithms can be employed for yield prediction of different crops (Elavarasan et al. 2018; Shahhosseini et al. 2019; Whitmire et al. 2021). These results suggest that ML algorithms can be employed successfully to analyze the data by binarizing the data into two sub-groups (Visa et al. 2011; Marković et al. 2021). In this way, ML models evaluate and predict the results using confusion matrix and cross-validation techniques. In the present study, the confusion matrix of different ML models was used for predicting the impact of different mutagens type, concentration, and their treatment time on in vitro regeneration and morphogenic traits of in vitro induced shoots. The confusion matrix is a powerful tool used for predicting the data and was recently successfully reported for predicting the insect appearance (Marković et al. 2021). Cross-validation is another technique in ML that is used to estimate the accuracy of the employed model for the unseen data, and allow to select the best one or two models (Suganya 2020). It also helps to get the error rate of the given dataset. The results achieved confirmed the significance of the cross validation technique and clearly illustrated the better performance of the RF model for all output variables followed by the MLP model over the remaining three other models (Aasim et al. 2022).

The use of ML and ANN models in the field of agriculture and plant biotechnology is limited (Silva et al. 2019), with some successful reports on predicting the results in plant tissue culture studies (Hesami and Jones 2021; Salehi et al. 2021). Application of different ML models presented the variable prediction based on input variables like genotypes, culture conditions, or variable outputs (TU et al. 2018; Hesami et al. 2020a; Jafari and Shahsavar 2020; Saffariha et al. 2020). Among the models utilized in this study, the RF algorithm was the best model to estimate the outputs of in vitro organogenesis and morphological traits of regenerated shoots. This model showed outstanding performance in all predictions with a performance score of 1.0 for 5 output variables and 0.99 for the sixth variable. Similarly, studies in plant tissue culture using ML models also revealed the better performance of the RF model compared to other models used for optimizing in vitro callus growth and development of hemp (Hesami and Jones 2021). The MLP ranked second with a better prediction of the five out of six output variables as compared to other models. The performance of the other models was relatively less than RF Model and varied according to the type of the outputs. The results confirmed that ML algorithms may predict in variable ways to the given dataset (Suganya 2020). These results further illustrated that the response of applied models is depending on the target and mode of the model applied in plant tissue culture (Hesami and Jones 2021; Salehi et al. 2021).

Conclusion

Application of ML and ANN models are powerful tools for predicting the attained results more precisely. Analysis of data followed by optimizing and predicting the results can be useful for complex biological processes like in vitro mutagenesis. In this study, the effect of sodium azide (NaN_3) and ethyl methanesulfonate (EMS) in different concentrations and treatment time on the outputs was investigated. Machine learning algorithms were utilized to predict the outputs. The RF model was the best model predicting the outputs correctly and displayed outstanding performance. The results attained from this study will alow to apply ML models for plant tissue culture and other areas of plant biotechnology.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00344-022-10808-w.

Acknowledgements The experiments reported in this paper were fully performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources).

Author Contributions KM: Conducted research and data tabulation, MA: conceived idea, Data analysis, Article Writing, RK: Machine Learning Analysis, Article writing, MK: Methodology, Supervision, SAA: Data Analysis, Article writing.

Declarations

Conflict of Interest The authors declare that there is not any conflict of interest regarding publication of this paper.

References

Aasim M, Sameeullah M, Karataş M et al (2019) An insight into biotechnological approaches used for the improvement of secondary metabolites from the medicinal aquatic plant, water hyssop (Bacopa monnieri L.). In: Akhtar MS, Swamy MK (eds) Natural Bio-active compounds. Springer, pp 123–152

Aasim M, Katırcı R, Akgur O et al (2022) 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
Angel Deborah S, S Milton Rajendram, and T T Mirnalinee (2017) SSN_MLRG1 at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis Using Multiple Kernel Gaussian Process Regression Model. In:·Proceedings of the 11thInternational Workshop on Semantic Evaluation (SemEval-2017), Associationfor Computational Linguistics (ACL), Vancouver, Canada. p 823–826
Aguiar S, Borowski T (2013) Neuropharmacological review of the nootropic herb Bacopa monnieri. Rejuvenation Res 16:313–326
Bado S, Forster BP, Nielen S et al (2015) Plant mutation breeding: current progress and future assessment. Plant Breed Rev 39:23–88
Bairu MW, Aremu AO, Van Staden J (2011) Somaclonal variation in plants: causes and detection methods. Plant Growth Regul 63:147–173
Behera M, Panigrahi J, Mishra RR, Rath SP (2012) Analysis of EMS induced in vitro mutants of Asteracanthalongifolia (L.) Nees using RAPD markers. Indian J Biotech 11:37-39
Chaudhari KS, Tiwari NR, Tiwari RR, Sharma RS (2017) Neurocognitive effect of nootropic drug Brahmi (Bacopa monnieri) in Alzheimer's disease. Ann Neurosci 24:111–122
Chen T, Guestrin C (2016) XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, pp 785–794
Elavarasan D, Vincent DR, Sharma V et al (2018) Forecasting yield by integrating agrarian factors and machine learning models: a survey. Comput Electron Agric 155:257–282
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
Gago J, Landín M, Gallego PP (2010) Artificial neural networks modeling the in vitro rhizogenesis and acclimatization of Vitis vinifera L. J Plant Physiol 167:1226–1231
GMO Ep on GMO, Mullins E, Bresson J et al (2021) In vivo and in vitro random mutagenesis techniques in plants. EFSA J 19:e06611
Hesami M, Jones AMP (2020) Application of artificial intelligence models and optimization algorithms in plant cell and tissue culture. Appl Microbiol Biotechnol 104:1–37
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 105:1–12
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-NSGAII). Front Plant Sci 10:1–13. https://doi.org/10.3389/fpls.2019.00282
Hesami M, Condori-Apfata JA, Valencia MV, Moham M (2020a) Application of artificial neural network for modeling and studying in vitro genotype-independent shoot regeneration in wheat. Appl Sci 10:1–10. https://doi.org/10.3390/APP10155370
Hesami M, Pepe M, Alizadeh M et al (2020b) Recent advances in cannabis biotechnology. Ind Crops Prod 158:113026. https://doi.org/10.1016/j.indcrop.2020.113026
Hesami M, Yoosefzadeh Najafabadi M, Adamek K, Torkamaneh D, Jones AMP (2021). Synergizing offtargetpredictions for in silico insights of CENH3 knockout in cannabis through CRISPR/CAS. Molecules 26(7):2053
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, Shahsavar 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
Jankowicz-Cieslak J, Till BJ (2017) Chemical mutagenesis and chimera dissolution in vegetatively propagated banana. Biotechnologies for plant mutation breeding. Springer, Cham, pp 39–54
Karataş M, Aasim M (2014) Efficient in vitro regeneration of medicinal aquatic plant water hyssop (Bacopa monnieri L. PENNELL). Pakistan J Agric Sci 51
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
Katırcı R, Yılmaz EK, Kaynar O, Zontul M (2021) Automated evaluation of Cr-III coated parts using Mask RCNN and ML methods. Surf Coatings Technol 422:127571. https://doi.org/10.1016/j.surfcoat.2021.127571
Kean JD, Downey LA, Stough C (2017) Systematic overview of Bacopa monnieri (L.) Wettst. dominant poly-herbal formulas in children and adolescents. Medicines 4:86
Kelly R, Healy K, Anand M et al (2021) Climatic and evolutionary contexts are required to infer plant life history strategies from functional traits at a global scale. Ecol Lett 24:970–983
Kharde AV, Chavan NS, Chandre MA et al (2017) In vitro enhancement of bacoside in brahmi (Bacopa monnieri) using colchicine. J Plant Biochem Physiol 5:1–6
Kharkwal MC (2012) A brief history of plant mutagenesis. Plant Mutat Breed Biotechnol. https://doi.org/10.1079/9781780640853.0021
Kirtis A, Aasim M, Katırcı 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 21:1–12
Marković D, Vujičić D, Tanasković S et al (2021) Prediction of pest insect appearance using sensors and machine learning. Sensors 21:4846
Metlek S, Kayaalp K (2020) Detection of autistic spectrum disorder with machine learning algorithms. J Intell Syst Theory Appl 3:60–68. https://doi.org/10.38016/jista.755481
Qi C, Diao J, Qiu L (2019) On estimating model in feature selection with cross-validation. IEEE Access 7:33454–33463. https://doi.org/10.1109/ACCESS.2019.2892062
Rayan AO, Mohamed SY, El-Habashy S (2014) In vitro studies of different chemical mutagens on some apple rootstock (Malus domestica Borkh). J Agric Chem Biotechnol 5:305–318
Saffariha M, Jahani A, Potter D (2020) Seed germination prediction of Salvia limbata under ecological stresses in protected areas: an artificial intelligence modeling approach. BMC Ecol 20:1–14
Sagar D, Nebapure SM, Chander S (2017) Development and validation of weather based prediction model for Helicoverpa armigera in chickpea. J Agrometeorol 19:328–333
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
Shahhosseini M, Martinez-Feria RA, Hu G, Archontoulis SV (2019) Maize yield and nitrate loss prediction with machine learning algorithms. Environ Res Lett 14:124026
Sharma A, Jain A, Gupta P, Chowdary V (2020) Machine learning applications for precision agriculture: a comprehensive review. IEEE Access 9:4843–4873
Silva JCF, Teixeira RM, Silva FF et al (2019) Machine learning approaches and their current application in plant molecular

biology: a systematic review. Plant Sci 284:37–47. https://doi.org/10.1016/j.plantsci.2019.03.020
Sivaramakrishna C, Rao CV, Trimurtulu G et al (2005) Triterpenoid glycosides from Bacopa monnieri. Phytochemistry 66:2719–2728
Skawsang S, Nagai M, Tripathi NK, Soni P (2019) Predicting rice pest population occurrence with satellite-derived crop phenology, ground meteorological observation, and machine learning: a case study for the Central Plain of Thailand. Appl Sci 9:4846
Spencer-Lopes MM, Forster BP, Jankuloski L (2018) Manual on mutation breeding. Food and Agriculture Organization of the United Nations
Suganya M (2020) Crop yield prediction using supervised learning techniques. Int J Comput Eng Technol 11
Suprasanna P, Vitthal SB, Yadav PV (2012) In vitro mutagenesis and selection in plant tissue cultures and their prospects for crop improvement. Bioremediation, Biodivers Bioavailab 6:6–14
Suprasanna P, Mirajkar SJ, Patade VY, Jain SM (2014) Induced mutagenesis for improving plant abiotic stress tolerance. In Mutagenesis: exploring genetic diversity of crops. Wageningen Academic Publishers, p e30765
Tu K, Li L, Yang L et al (2018) Selection for high quality pepper seeds by machine vision and classifiers. J Integr Agric 17:1999–2006
Vajpayee P, Dhawan A, Shanker R (2006) Evaluation of the alkaline comet assay conducted with the wetlands plant Bacopa monnieri L. as a model for ecogenotoxicity assessment. Environ Mol Mutagen 47:483–489
Visa S, Ramsay B, Ralescu AL, Van Der Knaap E (2011) Confusion matrix-based feature selection. MAICS 710:120–127
Whitmire CD, Vance JM, Rasheed HK et al (2021) Using machine learning and feature selection for alfalfa yield prediction. AI 2:71–88
Yan S, Ye L, Han S et al (2020) Speech interactive emotion recognition system based on random forest. Int Wirel Commun Mob Comput IWCMC 2020:1458–1462. https://doi.org/10.1109/IWCMC48107.2020.9148117
Zhang Q, Deng D, Dai W et al (2020) Optimization of culture conditions for differentiation of melon based on artificial neural network and genetic algorithm. Sci Rep 10:1–8. https://doi.org/10.1038/s41598-020-60278-x