Heliyon 8 (2022) e11969

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Prediction of *In vitro* organogenesis of *Bacopa monnieri* using artificial neural networks and regression models

Pavitra Viswanathan [a,c], Jaabili S. Gosukonda [a,d], Samantha H. Sherman [a], Nirmal Joshee [a], Ramana M. Gosukonda [b,*]

[a] *Agricultural Research Station, Fort Valley State University, Fort Valley, GA 31030, USA*
[b] *Department of Agricultural Sciences, College of Agriculture, Family Sciences and Technology, Fort Valley State University, Fort Valley, GA 31030, USA*
[c] *Department of Bioinformatics, Boston University, Boston, MA 02115, USA*
[d] *Houston County High School, Warner Robins, GA 31088, USA*

ARTICLE INFO

Keywords:
Artificial neural networks
Modeling
Bacopa monnieri
Plant growth regulators
In vitro organogenesis

ABSTRACT

This study was conducted to determine if artificial neural networks (ANN) can be used to accurately predict *in vitro* organogenesis of *Bacopa monnieri* compared with statistical regression. Prediction models were developed for shoot and root organogenesis (outputs) on two culture media (Murashige and Skoog and Gamborg B5) affected by two explant types (leaf and node) and two cytokinins (6-Benzylaminopurine and Thidiazuron at 1.0, 5.0, and 10.0 μM levels) with and without the addition of auxin (1-Naphthaleneacetic acid 0.1 μM) (inputs). Categorical data were encoded in numeric form using one-hot encoding technique. Backpropagation (BP) and Kalman filter (KF) learning algorithms were used to develop nonparametric models between inputs and outputs. Correlations between predicted and observed outputs (validation dataset) were similar in both ANN-BP (R values = 0.77, 0.71, 0.68, and 0.48), and ANN-KF (R values = 0.79, 0.68, 0.75, and 0.49), and were higher than regression (R values = 0.13, 0.48, 0.39, and 0.37) models for shoots and roots from leaf and node explants, respectively. Because ANN models have the ability to interpolate from unseen data, they could be used as an effective tool in accurately predicting the *in vitro* growth kinetics of Bacopa cultures.

1. Introduction

Bacopa monnieri (L.), commonly known as Brahmi, is a medicinal Ayurvedic herb that has been used for improving memory and thinking skills, insomnia, mental health, seizures, and anxiety [1]. These therapeutic benefits of *B. monnieri* underscore the necessity for its use in industrial formulations. Tissue culture propagation supplies clean, identical, and compatible Brahmi plant material in a large-scale for downstream processing [2, 3]. Lately, the use of machine learning (ML) approaches such as artificial neural networks (ANN) to accurately predict *in vitro* growth kinetics have gained importance in validating plant tissue culture processes and helping the industry to scale up productions by fine-tuning the automation process [4, 5].

In tissue culture, plant organogenesis is primarily controlled by the presence of growth regulators, both *in vitro* and *in vivo* [6]. It is also affected by the type of explant, the constitution of culture media, the type of phytohormones such as auxins and cytokinins, and their interactions [7]. Often, the combination of these factors causes the process of plant *in vitro* organogenesis to be nonlinear and nondeterministic growth progressions [7, 8]. Consequently, any optimization procedure in plant tissue culture calls for numerous experiments which are often laborious and time-consuming. Therefore, an initial study followed by modeling based on the experimental data not only offer an easy and less expensive alternative but also expedite any such optimization procedures [9, 10]. However, use of statistical models for optimization of every stage in *in vitro* culture is challenging. Statistical models (e.g., non-linear multivariate modeling, polynomial regression method etc.) have limitations in discerning nonlinearity that exists in tissue culture processes [11]. For instance, these approaches have limited complexity and are not as appropriate for multifactorial and large datasets with complex relationships as they often produce poor curve fit during model development [12].

On the other hand, nonlinear, nonparametric machine learning (ML) techniques are more efficient in handling large amounts of complex and nondeterministic datasets with multiple independent versus multiple dependent variables which are often observed in plant tissue culture

* Corresponding author.
E-mail address: gosukonr@fvsu.edu (R.M. Gosukonda).

https://doi.org/10.1016/j.heliyon.2022.e11969
Received 19 July 2022; Received in revised form 28 September 2022; Accepted 22 November 2022

studies [13, 14]. Furthermore, nonparametric methods bring distinct advantages over statistical models, as they do not require a priori knowledge of the research problem and distribution or relationship between the input and output variables and thus are effective in modeling nonlinear plant tissue culture data [15, 16]. Modern ML methods come with a powerful set of algorithms that can self-learn, analyze complexities in datasets and predict, classify, estimate, simulate, the underlined trends and behaviors. Therefore, these methods provide better understanding of tissue culture processes and help us make correct decisions for optimization at every step [17, 18, 19].

Importantly, machine learning models have the ability to adapt robustly to new, previously unseen data, drawn from the same distribution as the one used to create the model. It means, the predictive capability of ML models continued to be useful in making educated guesses with data of same type from future plant tissue culture investigations [20]. Since machine learning models have the inherent ability to handle high biological variability and uncertainty associated with various biological processes, they could be of great use in the process analysis, optimization, product development, and quality assurances in tissue culture systems [9, 13, 21]. In addition, these models are non-invasive, stress-free to use, time efficient and lead to similar results obtained by means of other more invasive methods, thus offering a good alternative to use in such studies [5, 22].

Artificial neural networks (ANN), a popular machine learning approach, has led the modeling field in plant tissue culture for the past several years [23]. Artificial neural networks are reliable fast-computing techniques which yields accurate predictions even with incomplete and noisy data regularly observed in nonlinear problems [24, 25]. Furthermore, ANN models can be established rapidly with limited experimental data, and they could approximate any complex nonlinear systems [26]. However, the performance of an ANN model depends on various factors including the type of network architecture, the transfer function, and the learning algorithms used. Artificial neural networks trained with Kalman filter (KF) and backpropagation (BP) algorithms reported to have strong generalization capability compared to statistical models [27].

In general, ANN applications in plant tissue culture are diverse and reported to have performed better than statistical regression models [21, 23]. For example, ANNs have been used in plant tissue culture to optimize the culture conditions of melon [25], measure and predict physical properties of embryogenic callus and number of somatic embryos in Ajowan [28], predict the hyoscyamine content in Datura [29], optimize *in vitro* shoot proliferation of wallflower [30], and prunus rootstocks [31]. Despite significant advantages, the use of ANN in Bacopa's plant tissue culture has not been explored.

Therefore, in this study ANN models were developed to predict *in vitro* shoot and root organogenesis of *B. monnieri* from leaf and node explants as a function of culture media and plant growth regulators, and to compare the prediction performances of ANN models with those of regression models. To our knowledge this is the first study to use artificial neural networks to predict *in vitro* organogenesis of *B. monnieri.*

2. Materials and methods

2.1. Media preparation, culture establishment, and organogenesis of *B. monnieri*

Based on earlier research at the Biotechnology Laboratory at Fort Valley State University, experiments were conducted to count the *in vitro* shoot and root production from leaf and node explants of *B. monnieri* on Murashige and Skoog (MS) [32] and Gamborg (B5) [33] culture media affected by two cytokinins (6-Benzylaminopurine (BAP) and Thidiazuron (TDZ) at 1.0, 5.0, and 10.0 μM levels) with and without the addition of auxin (1-Naphthaleneacetic acid (NAA) at 0.1 μM). Culture media preparation, growth regulator supplementation, explant preparation, and inoculations were carried out as outlined in the publications [34, 35]. Explants were cultured on the shoot induction media for three weeks and followed by subculture on MS basal elongation medium for four weeks. At the completion of elongation period, cultures were removed from vessels and destructive counting was performed to count adventitious shoots and roots in each explant. In this experiment, all cultures were grown under uniform environmental conditions using growth chambers. Therefore, factors such as temperature, photoperiod, humidity etc., were not included in the model development.

2.2. Collection of data for model development

Data from experiments discussed in section 2.1 were used in the development of models. Culture media and growth regulators data were used as inputs while *in vitro* shoot, or root organogenesis data, were used as outputs for developing prediction models. In total, there were four datasets, one for each output. i.e., shoot output and root output from leaf (leaf–shoots and leaf–roots) and node (node–shoots and node-roots) explants, respectively.

Because ANNs learning algorithms require data to be in numerical form, categorical data (such as culture media, which can be MS or B5 in this study) were converted into integer data using one-hot encoding technique (where each input label was represented by a binary vector (i.e., 0 or 1). Usually, this technique is suitable for categorical data in which the categories are not related to each other [36]. On the other hand, inputs with numeric data were presented as is to models.

Kurtosis and skewness were performed to understand where the most information was lying (distribution) and evaluated the outliers in the experimental data. While skewness is a degree of asymmetry observed in a probability distribution that deviates from the symmetrical normal distribution in data, Kurtosis refers to the degree of presence of outliers in the data distribution [37]. In this study, all datasets except node-shoot, had symmetrical skewness and flatter kurtosis (pytokurtic). However, node-shoot dataset presented with a positively high skewness and high peak kurtosis (leptokurtic) suggesting a higher probability of outlier values in this dataset as compared to others.

2.3. Artificial neural networks software

Artificial neural networks models were constructed using NeuralWorks Predict® (version 3.2; NeuralWare, Carnegie, PA, USA) software. Because the selection of a network configuration and its parameters is usually accomplished empirically [38], use of commercially available software such as Predict® is convenient for obtaining the best models with ease.

2.4. Multilayer perceptron network models

Prediction models were built using a multilayer perceptron (MLP) neural network with feedforward architecture (Figure 1), as it discerns nonlinear data [39]. Furthermore, it has universal approximation and the compact representation [20]. In a typical MLP, the input layer receives the signal to be processed from inputs. The output layer makes a prediction about the input, and the in between hidden layer, the true computational engine, provides a possible solution using a transfer function, in a feed forward approach [20, 39]. Usually, the software builds neural networks incrementally using cascade learning [40]. Cascade learning, also known as cascade–correlation, is a supervised learning algorithm, which begins with a nominal network and then trains and adds hidden units one at a time, and always connecting all the previous units to the current unit [38]. The "cascade" part refers to the stepwise mode of construction of the structure, and the "correlation" part refers to the way in which the hidden units are trained by maximizing the correlation between output of hidden units and the desired output of the network across the training data [41].

In this study, ANN models were constructed using two learning algorithms — a Kalman learning algorithm (ANN-KF) and an adaptive gradient learning algorithm (Backpropagation; ANN-BP). The objective

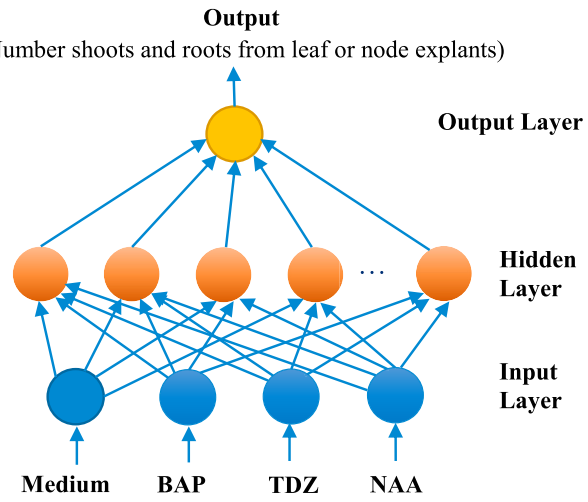

Figure 1. Schematic representation of multilayer perceptron neural network model. Inputs were culture media (MS and B5), BAP, TDZ and NAA and the outputs were number of shoot and roots produced from leaf and node explants of *B. monnieri*.

of these learning algorithms, during training phase, was to minimize the error between predicted and measured outputs for producing an optimal configuration [42].

2.5. Train, test, and validation datasets for ANN models

Out of 878 available records for each output (i.e., leaf–shoots, leaf–-roots, node–shoots and node-roots), 618 were used to build prediction models, and the rest of 260 records were used to validate the built models. During the model building phase, the software split 618 records further into train (432 records) and test (186 records) datasets so that the two were statistically close to each other [38]. The objective of training models was to produce a formula to capture key relationships between inputs and outputs. Following, trained models were assessed for their interpolation performances with test datasets.

Although, models were trained using train dataset, heuristics related to their performance with test dataset were also used to guide choices for building an optimal model [38]. Because of this reason, a good ANN model mostly has similar performances with train and test datasets, but to have it perform well with unseen dataset is important. Since validation datasets were independent of the train and test datasets and had the distribution like that of the train datasets, they were used to further confirm the performances of ANN models. The number of records used in training, testing, and validation of models is shown in Table 1.

2.6. Selection of an optimal ANN configuration

The network configurations for models were approached empirically, and the model that performed well with validation dataset was selected. To evaluate the performance of ANN models, several statistical indices such as Pearson correlation (R), mean relative percentage residual (MRPR, %), bias factor, mean absolute relative residual (MARR, %), accuracy factor, and standard deviation were calculated [41]. Since R values measure the strength and direction of a linear relationship between predicted and observed outputs, they were used to select the best models (highest R values were considered as best) [26].

2.7. Statistical regression models

Similar to the datasets used in developing ANN models, regression models were also built (using MS Excel) with 618 records and validated with 260 records. The objective of regression models was to assess the quantitative impact of factors such as culture media, phytohormones and their combinations on *in vitro* organogenesis of *B. monnieri*. Artificial neural networks models are generalization of the standard linear and logistic regression methods [43], and a comparison between these approaches for prediction and data fit could lead to identification of better models.

The final architectures of ANN models and the equations of regression are shown in Table 2. Generally, the number of inputs in the architecture of ANN models does not tally with the number of inputs of the datasets due to data analysis and transformation, variable selection, and algorithm etc., used by Predict® during model development. Additionally, box and whisker plots were made to compare the prediction performances of ANN-BF, ANN-KF, and regression models and to visually display the data distribution through their quartiles.

3. Results and discussion

3.1. Performance of models with leaf-shoots organogenesis

Prediction performances of models for *in vitro* shoot organogenesis from leaf explants are shown in Table 3. Correlations between predicted

Table 2. Artificial neural networks (ANN) models' architecture and regression equations used to predict *in vitro* organogenesis responses as a function of culture media (MS and BS) and growth regulators (BAP, TDZ, and NAA) from leaf and node explants of *B. monnieri*.

Output (per explant)	Architecture (inputs-hidden nodes-output)		Regression Models
	ANN-BP	ANN-KF	
Leaf–Shoots	7–9–1	7–3–1	$Y = 8.24 + 0X_1 - 0.58X_2 + 0.058X_3 + 0.05X_4 + 4.54X_5$
Leaf–Roots	7–9–-1	7–7–1	$Y = 3.37 + 0X_1 - 0.75X_2 - 0.24X_3 + 0.04X_4 + 12.09X_5$
Node–Shoots	4–8–1	4–14–1	$Y = 1.70 + 0X_1 - 0.72X_2 + 0.14X_3 + 0.18X_4 + 5.28X_5$
Node–Roots	9–13–1	9–9–1	$Y = 4.67 + 0X_1 - 0.28X_2 - 0.23X_3 - 0.13X_4 + 7.40X_5$

Note: ANN-BP = Backpropagation algorithm; ANN-KF = Kalman filter algorithm; Y = shoot or root organogenesis; X_1 = B5 medium; X_2 = MS medium; X_3 = BAP; X_4 = TDZ; X_5 = NAA.

Table 1. Distribution of *in vitro* organogenesis data records of *B. monnieri* used in the development of the artificial neural networks and regression models.

Output (per explant)	Regression*	ANN-BP			ANN-KF		
		Training Set	Testing Set	Validation Set	Training Set	Testing Set	Validation Set
Leaf–Shoots	618/260	432	186	260	432	186	260
Leaf–Roots	618/260	432	186	260	432	186	260
Node–Shoots	618/260	432	186	260	432	186	260
Node–Roots	618/260	432	186	260	432	186	260

ANN-BP = Backpropagation algorithm; ANN-KF = Kalman filter algorithm.
* Model was developed using 618 records and validated with 260 records.

Table 3. Pearson correlations between measured and predicted outputs from regression and artificial neural networks models (ANN-BP and ANN-KF).

Output (per explant)	Regression		ANN-BP				ANN-KF			
	Validation		Training		Validation		Training		Validation	
	R	SD	R	SD	R	SD	R	SD	R	SD
Leaf–Shoots	0.13	3.22	0.75	4.42	0.77	4.09	0.76	4.41	0.79	4.10
Leaf–Roots	0.48	1.82	0.72	1.89	0.68	1.92	0.72	1.90	0.68	1.93
Node–Shoots	0.39	1.59	0.69	1.56	0.71	1.66	0.72	1.00	0.75	1.69
Node-Roots	0.37	1.65	0.54	1.87	0.48	1.76	0.54	1.21	0.49	1.77

ANN-BP = Backpropagation algorithm; ANN-KF = Kalman filter algorithm.
R = Pearson Correlation; SD = Standard deviation.

leaf-shoot and observed leaf-shoot outputs for regression model was 0.13 and for ANN-BP and ANN-KF models were 0.77and 0.79, respectively. Also, better performances of ANN models with validation datasets than with train datasets, signifying that these models have strong ability to generalize and interpolate unseen patterns within the domain of training.

Plots of predicted values for *in vitro* shoot organogenesis from leaf explants in Figure 2, indicate that ANN-BP and ANN-KF models very tightly followed the trends set by measured values at all treatment levels, while the regression model failed completely to follow the patterns of measured values. For instance, the average shoot production from node explants was highest (14.8) with B5 + BAP 5.0 μM treatment, followed by the second highest (13.3 shoots) with MS + BAP 1.0μM + NAA 0.1 μM treatment. Also, the lowest shoot organogenesis was observed in control treatments of original experiment. These observed trends were accurately predicted by ANN models with leaf-shoots dataset.

3.2. Performance of models with leaf-roots organogenesis

The results (Table 3) indicate that both ANN-KF and ANN-BP models have identical prediction performances (R = 0.68) for leaf-root organogenesis and are superior to the regression model (R = 0.48). In this case, ANN models had lower R values with validation dataset than with train dataset, suggesting that these models may have become overfitted and unable to generalize well to validation dataset. Overfitting occurs when the model memorizes the noise and fits too closely to the data points during its train phase. Also, it could occur if the training dataset has a lower error rate, and the validation dataset has a higher error rate [24].

Comparison of plots for leaf-root prediction in Figure 3 indicated that both ANN-BP and ANN-KF models closely followed the patterns set by measured values at all treatment levels, while the regression model failed to identify these patterns at certain treatments levels (e.g., B5 + TDZ 10 μM).

3.3. Performance of models with node-shoots organogenesis

Pearson correlations between predicted node-shoots and observed node-shoots values were better with ANN-BP (R = 0.71) and ANN-KF (R = 0.75) than with regression (R = 0.39) models (Table 3). Better correlations by ANN-KF over ANN-BP, suggesting that KF algorithm may have efficiently handled the inherent variability or noise in the node-shoots dataset.

Prediction plots in Figure 4 revealed that the patterns produced by ANN-BP and ANN-KF models were similar to those of measured values and are better than those of the regression model at majority of treatments. However, all models failed to detect the trends of certain treatments that exhibited large number of shoots (e.g., MS + TDZ 5.0μM + NAA 0.1 μM and B5 + BAP 1.0 μM). It is not unusual to ANN models to display better generalization in predicting certain outputs and not others [44].

3.4. Performance of models with node-roots organogenesis

Prediction performances for node-root organogenesis were better with ANN-BP (R = 0.48) and ANN-KF (R = 0.49) than with regression (R = 0.37) model (Table 3). However, both ANN models revealed stronger

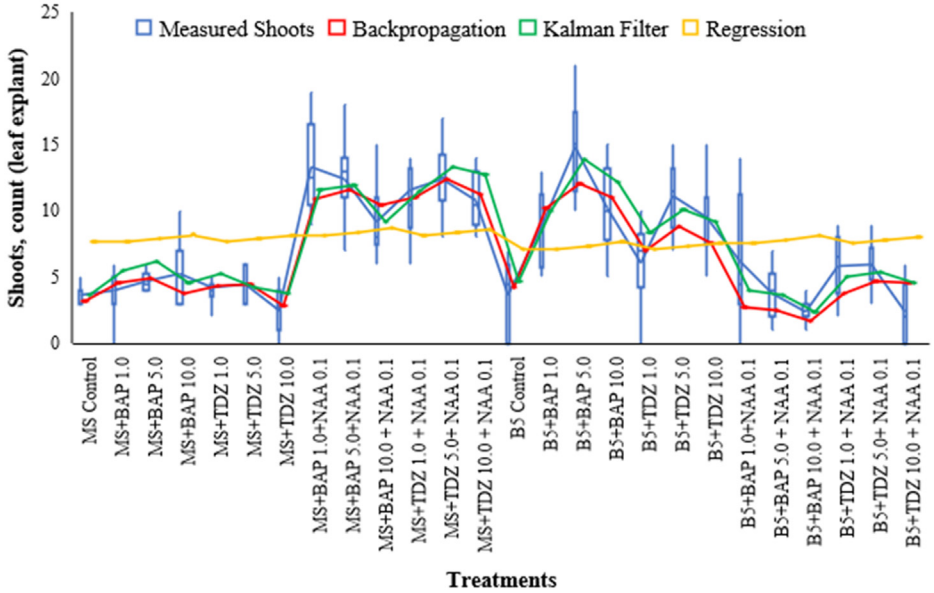

Figure 2. Trends in *in vitro* shoot production from leaf explants, in response to culture media (MS and B5) and combination of growth regulators (BAP, TDZ, and NAA) as predicted by regression, artificial neural networks with backpropagation algorithm (ANN-BP), and Kalman filter algorithm (ANN-KF) models on validation dataset.

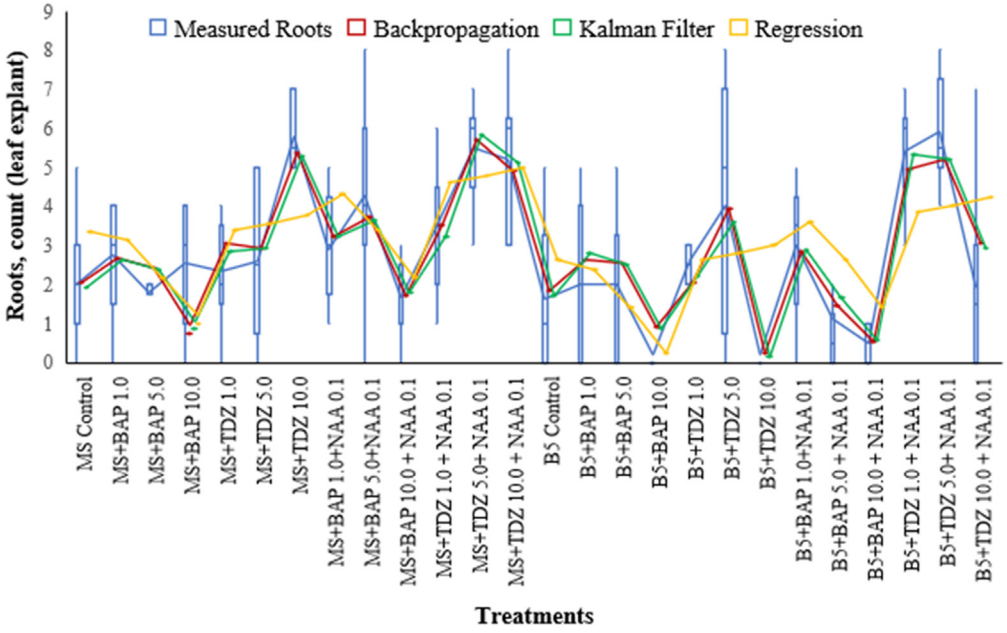

Figure 3. Trends in *in vitro* root production from leaf explants in response to culture media (MS and B5) and combination of growth regulators (BAP, TDZ, and NAA) as predicted by regression, artificial neural networks with backpropagation algorithm (ANN-BP), and Kalman filter algorithm (ANN-KF) models on validation dataset.

performances (R = 0.54) with the train dataset than with the validation dataset, suggesting an overfit of the data points by these models.

Prediction performances of ANN-BP, ANN-KF, and regression models were plotted against the measured values of node-root organogenesis (Figure 5). Compared to the patterns produced by regression models, the patterns produced by ANN models were better aligned with the measured values of original experiment. For instance, the average root count was high with treatment MS + BAP 1.0μM + NAA 0.1 μM and low with treatments MS + BAP 5.0 μM and B5 + TDZ 10 μM. These observed trends were correctly reproduced by the ANN models with node-roots dataset.

Overall, both ANN-BP and ANN-KF models showed higher correlation values for shoot organogenesis than for root organogenesis across the datasets. Also, prediction performances of ANN models were better for leaf-based organogenesis compared to node-based organogenesis, which suggests the possibility of higher levels of inherent variability in node explants. It agrees with the trends observed in original experiment in which the leaf-based organogenesis was stronger than the node-based organogenesis. The reason could be due to differences in inherent variability between node and leaf explants as they differ in their structural complexity and levels of endogenous phytohormones. Usually, endogenous phytohormones regulate tissue differentiation *in vitro* and thus varying levels would carry varying degrees of competence in explants for organogenesis [45].

Despite high biological variabilities associated with plant material and tissue culture process, ANN models were able to generalize well with leaf-shoots and node-shoots validation datasets. Normally, data transformation is needed for count data (such as number of roots or shoots observed in this study) to improve normality for regression methods

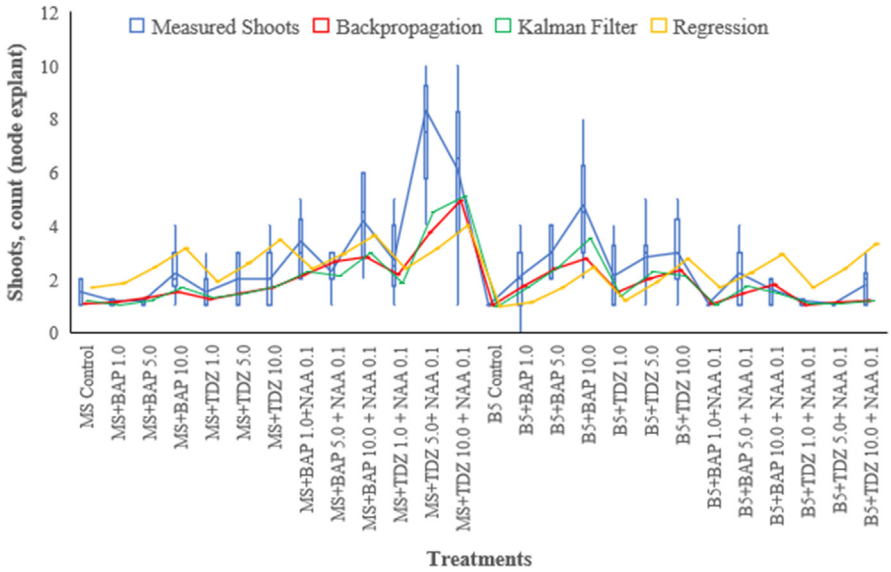

Figure 4. Trends in *in vitro* shoot production from node explants in response to culture media (MS and B5) and combination of growth regulators (BAP, TDZ, and NAA) as predicted by regression, artificial neural networks with backpropagation algorithm (ANN-BP), and Kalman filter algorithm (ANN-KF) models on validation dataset.

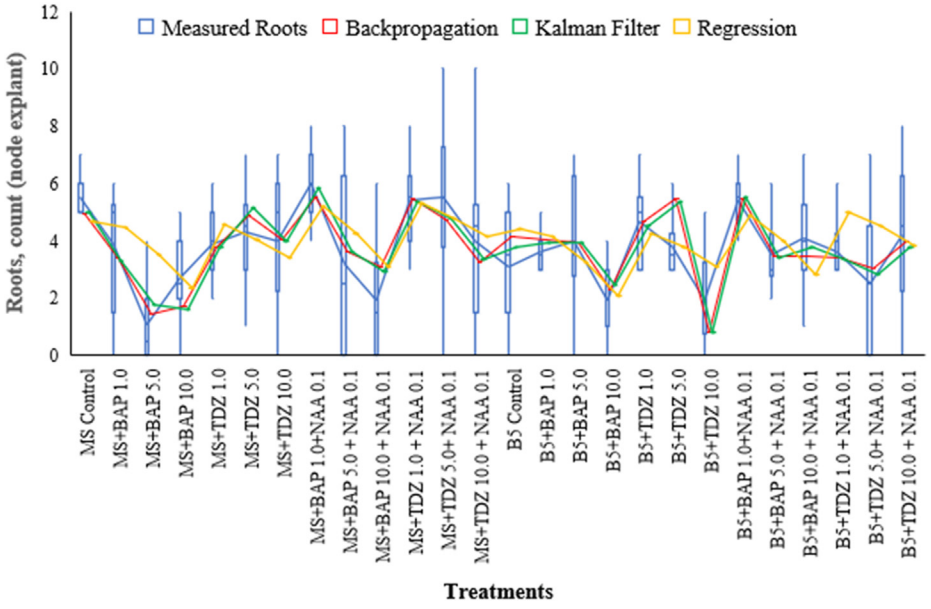

Figure 5. Trends in *in vitro* root production from node explants in response to culture media (MS and B5) and combination of growth regulators (BAP, TDZ, and NAA) as predicted by regression, artificial neural networks with backpropagation algorithm (ANN-BP), and Kalman filter algorithm (ANN-KF) models on validation dataset.

[11]. However, in the present study, better prediction performances were observed in ANN models (without data transformation) than in regression models with shoot and root count data.

Comparison of statistical indices to evaluate prediction performances of models (Table 4) showed that both ANN-BP and ANN-KF performed better than regression models (Table 4). The mean relative percentage residual (MRPR, %) values indicated an overprediction of the outputs by all models with validation datasets. Regression models displayed greater amounts of overprediction as compared to ANN models. For instance, regression models overpredicted the leaf–shoots outputs (validation dataset), on an average 80.51% (MRPR = −80.51), whereas ANN-BP and ANN-KF overpredicted the same, on average 15.20% (MRPR = -15.20) and 32.28% (MRPR = -32.28), respectively. However, with node-shoots dataset, the overprediction rates of models were markedly lower as compared to leaf–shoots outputs.

In general, overprediction by ANN models was found to be greater with validation datasets than with train datasets. The bias factor calculations for ANN models were in contrast with MRPR (%) trends, with most values being less than or closer to 1 (exception; leaf–shoot by ANN-KF) indicate an underestimation of outputs by these models [46]. A bias factor of 1 indicates that model has no bias, while bias factors of greater than 1 and less than 1 indicates an overestimation and underestimation of outputs by models, respectively [47].

Table 4. Comparison of statistical indices between artificial neural networks and regression models for each output.

Statistic	Model	Dataset	Output (by explant type)			
			Leaf–Shoots	Leaf–Roots	Node–Shoots	Node–Roots
MRPR (%)	Regression	Validation	-80.51	-18.83	-14.73	-16.74
	ANN-BP	Training	-6.08	-6.76	0.46	-20.84
		Validation	-15.20	-19.53	-5.17	-27.36
	ANN-KF	Training	-22.76	-5.27	-0.14	-4.99
		Validation	-32.28	-0.35	-3.00	-9.48
Bias factor	Regression	Validation	1.33	1.02	1.10	1.06
	ANN-BP	Training	0.92	0.97	1.00	1.01
		Validation	0.99	0.94	0.93	1.02
	ANN-KF	Training	1.07	0.98	1.00	1.00
		Validation	1.08	0.90	0.86	0.93
MARR (%)	Regression	Validation	104.89	36.71	28.09	26.82
	ANN-BP	Training	41.56	7.62	3.88	10.75
		Validation	44.12	25.07	13.01	21.84
	ANN-KF	Training	46.41	7.57	3.95	10.41
		Validation	49.94	25.38	12.58	21.61
Accuracy factor	Regression	Validation	1.82	1.35	1.26	1.21
	ANN-BP	Training	1. 46	1.08	1.04	1.08
		Validation	1.44	1.28	1.16	1.20
	ANN-KF	Training	1.43	1.08	1.04	1.08
		Validation	1.42	1.30	1.15	1.20

ANN-BP = Backpropagation algorithm; ANN-KF = Kalman filter algorithm.
MRPR = Mean relative percentage residual; MARR = Mean absolute relative residual.

The computed values of mean absolute relative residual (MARR, %) and accuracy factor also showed that the prediction performances of ANN models were better than regression. Both MARR (%) and accuracy factor measures the average deviations of predicted outputs from observed outputs. In case of leaf–shoots with validation dataset, the MARR value of 49.94 for ANN-KF model indicate that the predicted outputs of this model were 49.94% above or below the observed outputs. Likewise, an accuracy factor of 1.42 for this model and dataset suggests that the predicted outputs deviated on average 42% from observed outputs.

The discrepancy among different statistical indices for effectively identifying prediction bias could be due to a difference in the normalization of parameters or the methods of mathematical calculation (e.g., logarithm and exponential) [46]. Normalization of residuals is involved in the computations of MRPR and MARR, while normalization of predicted values is included in the computation of bias and accuracy factor. Because the normalized residuals have homogeneity of the variance, MRPR and MARR are more reliable statistics than bias and accuracy factor for assessing prediction performances of models [41].

Normally, the use of multiple statistical indices is recommended to assess precision, bias, and accuracy of prediction performances of models. In particular, these assessments are necessary for models dealing with plant *in vitro* organogenesis because it is highly difficult to predict due to inherent variability among explants combined with nonlinear tissue culture processes [17].

Furthermore, plant tissue culture is very intricate, and finding the factors such as culture media, type of explants, and plant growth regulators that influence the processes of organogenesis cannot always be well comprehended.

4. Conclusions

It is evident from this study that ANN models can predict *in vitro* shoot and root organogenesis of *B. monnieri* better than regression models, despite high variability associated with plant tissues and their complex interaction with culture media and growth regulators. For instance, Pearson correlations between predicted and observed outputs (on validation dataset) were similar in both ANN-BP (R values = 0.77, 0.71, 0.68, and 0.48), and ANN-KF (R values = 0.79, 0.68, 0.75, and 0.49), and were higher than regression (R values = 0.13, 0.48, 0.39, and 0.37) models for shoots and roots from leaf and node explants, respectively. Also, comparison of prediction plots indicated that both BP and KF learning algorithms had better precisions than regression models in all datasets. Since, ANN models have the ability to interpolate unseen patterns, they offer an effective tool for accurately predicting the *in vitro* growth kinetics of *B. monnieri*. Furthermore, ANN approaches have the potential to replace laborious and time-consuming experimental research.

Declarations

Author contribution statement

Pavitra Viswanathan: Performed the experiments, Contributed analysis tools or data.

Jaabili S. Gosukonda: Analyzed and interpreted the data; Contributed analysis tools or data; Wrote the paper.

Samantha H. Sherman: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Nirmal Joshee; Ramana M. Gosukonda: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

This work was supported by USDA National Institute of Food and Agriculture, Capacity Building Grant [grant no. 2021-38821-34512/project accession no. 1026022] and Evans Allen Grant (Accession no. 1018491).

Data availability statement

Data will be made available on request.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

[1] J.M. Bokelmann, 28 - Bacopa (*Bacopa monnieri*): leaf, in: J.M. Bokelmann (Ed.), Medicinal Herbs in Primary Care, Elsevier, USA, 2022, pp. 211–216.
[2] D. Aggarwal, S.K. Upadhyay, R. Singh, N. Sehrawat, M. Yadav, M. Singh, V. Kumar, Tissue culture propagation of a medicinal plant *Bacopa monnieri* (L.) Pennell, Adv. Bio. Res. 11 (2020) 97–103.
[3] M.P. Dharishini, K. Balasubramanian, A. Radha, *In vitro* micropropagation of *Bacopa monnieri* and detection of bacosides from secondary callus, J. Acad. Ind. Res. 3 (2014) 233–236.
[4] P. García-Pérez, E. Lozano-Milo, M. Landín, P.P. Gallego, Combining medicinal plant *in vitro* culture with machine learning technologies for maximizing the production of phenolic compounds, Antioxidants 9 (2020) 2–14.
[5] H. Singh, N. Bharadvaja, Treasuring the computational approach in medicinal plant research, Prog. Biophys. Mol. Biol. 164 (2021) 19–32.
[6] G.C. Phillips, M. Garda, Plant tissue culture media and practices: an overview, *in Vitro* Cell, Dev. Biol.-Plant 55 (2019) 242–257.
[7] Y. Long, Y. Yang, G. Pan, Y. Shen, New insights into tissue culture plant-regeneration mechanisms, Front. Plant Sci. 13 (2022), 926752.
[8] T.A. Arteta, R. Hameg, M. Landin, P.P. Gallego, M.E. Barreal, Artificial neural networks elucidated the essential role of mineral nutrients versus vitamins and plant growth regulators in achieving healthy micropropagated plants, Plants 11 (2022) 1–22.
[9] R.W. Maschke, K. Geipel, T. Bley, Modeling of plant *in vitro* cultures: overview and estimation of biotechnological processes, Biotechnol. Bioeng. 112 (2015) 1–12.
[10] J.C. Lorenzo, M. Garcia-Borroto, Use of regression analysis in plant cell, tissue, and organ culture experiments, *in Vitro* Cell, Dev. Biol. Plant 44 (2008) 229–232.
[11] J. Gago, L. Martınez-Nunez, M. Landın, P.P. Gallego, Artificial neural networks as an alternative to the traditional statistical methodology in plant research, J. Plant Physiol. 167 (2010) 23–27.
[12] P. Sharma, B.B. Sahoo, Precise prediction of performance and emission of a waste derived biogas–biodiesel powered dual–fuel engine using modern ensemble boosted regression tree: a critique to Artificial neural network, Fuel 321 (2022), 124131.
[13] S. Zielinska, E. Kępczyńska, Neural modeling of plant tissue cultures: a review, Biotechnologia 94 (2013) 253–264.
[14] B. Kia, A. Mendes, A. Parnami, R. George, K. Mobley, W.L. Ditto, Nonlinear dynamics based machine learning: utilizing dynamics-based flexibility of nonlinear circuits to implement different functions, PLoS One (2020), e0228534.
[15] S.J. Russell, P. Norvig, Learning from example, in: Artificial Intelligence: A Modern Approach, fourth ed., Pearson, USA, 2022, pp. 704–778.
[16] M. Hesami, A.M.P. Jones, Application of artificial intelligence models and optimization algorithms in plant cell and tissue culture, Appl. Microbiol. Biotechnol. 104 (2020) 9449–9485.
[17] M. Aasim, R. Katrici, O. Akgur, B. Yildirim, Z. Mustafa, M.A. Nadeem, F.S. Baloch, T. Karakoy, G. Yılmaz, Machine learning (ML) algorithms and artificial neural network for optimizing *in vitro* germination and growth indices of industrial hemp (*Cannabis sativa* L.), Ind. Crop. Prod. 181 (2022), 114801.
[18] P. Kaur, R.C. Gupta, A. Dey, T. Malik, D.K. Pandey, Optimization of salicylic acid and chitosan treatment for bitter secoiridoid and xanthone glycosides production in shoot cultures of *Swertia paniculata* using response surface methodology and artificial neural network, BMC Plant Biol. 20 (2020) 9–13.
[19] M. Pepe, M. Hesami, F. Small, A.M.P. Jones, Comparative analysis of machine learning and evolutionary optimization algorithms for precision micropropagation of *Cannabis sativa*: prediction and validation of *in vitro* shoot growth and development based on the optimization of light and carbohydrate sources, Front. Plant Sci. 12 (2021), 757869.
[20] O.I. Abiodun, A. Jantan, A.E. Omolara, K.V. Dada, N.A. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: a survey, Heliyon 4 (2018), e00938.
[21] M. Hesami, R. Naderi, M. Yoosefzadeh-Najafabadi, M. Rahmati, Data-driven modeling in plant tissue culture, J. Appl. Environ. Biol. Sci. 7 (2017) 37–44.
[22] R.D. Malangsa, E.A. Maravillas, Performance comparison of naïve bayes and K-NN algorithms on contamination grading for Abaca tissue culture (*in vitro*), Int. J. Comput. Sci. Inf. Technol. 5 (2017) 5–10.
[23] M. Niazian, G. Niedbała, Machine learning for plant breeding and biotechnology, Agriculture 10 (2020) 2–23.
[24] P. Sharma, Z. Said, A. Kumar, S. Nižetić, A. Pandey, A.T. Hoang, Z. Huang, A. Afzal, C. Li, A.T. Le, X.P. Nguyen, V.D. Tran, Recent advances in machine learning

research for nanofluid-based heat transfer in renewable energy system, Energy Fuels 36 (2022) 6626–6658.
[25] Q. Zhang, D. Deng, W. Dai, J. Li, X. Jin, Optimization of culture conditions for differentiation of melon based on artificial neural network and genetic algorithm, Sci. Rep. 10 (2020) 3524.
[26] R. Gosukonda, A.K. Mahapatra, X. Liu, G. Kannan, Application of artificial neural network to predict *Escherichia coli* O157:H7 inactivation on beef surfaces, Food Control 47 (2015) 606–614.
[27] R. Gosukonda, A.K. Mahapatra, D. Ekefre, M. Latimore Jr., Prediction of thermal properties of sweet sorghum bagasse as a function of moisture content using artificial neural networks and regression models, Acta Technol. Agric. 2 (2017) 29–35.
[28] M. Niazian, S.A. Sadat-Noori, M. Abdipour, M. Tohidfar, S.M.M. Mortazavian, Image processing and artificial neural network-based models to measure and predict physical properties of embryogenic callus sand number of somatic Embryos in Ajowan (*Trachyspermum ammi* (L.) Sprague). *In Vitro* Cell, Dev. Biol. Plant 54 (2018) 54–68.
[29] R. Amdoun, E. Benyoussef, A. Benamghar, L. Khelifi, Prediction of hyoscyamine content in *Datura stramonium* L. hairy roots using different modeling approaches: response Surface Methodology (RSM), Artificial Neural Network (ANN) and Kriging, Biochem. Eng. J. 144 (2019) 8–17.
[30] F. Fakhrzad, A. Jowkar, J. Hosseinzadeh, Mathematical modeling and optimizing the *in vitro* shoot proliferation of wallflower using multilayer perceptron non-dominated sorting genetic algorithm-II (MLP-NSGAII), PLoS One 17 (2022), e0273009.
[31] M.M. Arab, A. Yadollahi, M. Eftekhari, H. Ahmadi, M. Akbari, S.S. Khorami, Modeling and optimizing a new culture medium for *in vitro rooting* of G× N15 prunus rootstock using artificial neural network-genetic algorithm, Sci. Rep. 8 (2018) 1–18.
[32] T. Murashige, F. Skoog, A revised medium for rapid growth and bioassays with tobacco tissue cultures, Plant Physiol. 15 (1962) 473–497.
[33] O.L. Gamborg, R.A. Miller, K. Ojima, Nutrient requirements of suspension cultures of soybean root cells, Exp. Cell Res. 50 (1968) 151–158.
[34] N. Joshee, D. Harris, A. Yadav, A.K. Yadav, Influence of explant selection and culture conditions on organogenesis and germplasm conservation in *Bacopa monnieri* (L), Wettst, Acta Hortic. 756 (2007) 119–128.
[35] A.L. Croom, C. Jackson, B.N. Vaidya, P. Parajuli, N. Joshee, Thin cell layer (TCL) culture system for herbal biomass production and genetic transformation of *Bacopa monnieri* L, Wettst. Am. J. Plant Sci. 7 (2016) 1232–1245.
[36] A. Zheng, A. Casari, in: R. Roumeliotis, J. Bleiel (Eds.), Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists, O'Reilly Media, Sebastopol, CA, 2018, pp. 77–96.
[37] Z. Said, P. Sharma, R.M. Elavarasan, A.K. Tiwari, M.K. Rathod, Exploring the specific heat capacity of water-based hybrid nanofluids for solar energy applications: a comparative evaluation of modern ensemble machine learning techniques, J. Energy Storage 54 (2022), 105230.
[38] NeuralWorks Predict®, The Complete Solution for Neural Data Modeling User Guide, Neuralware, Carnegie, PA, 2018.
[39] J. Zou, Y. Han, S.S. So, Overview of artificial neural networks, Methods Mol. Biol. 458 (2008) 15–23.
[40] S.E. Fahlman, C. Lebiere, The Cascade-correlation learning architecture, in: D.S. Touretzky (Ed.), Advances in Neural Information Processing Systems, 2, Morgan-Kaufmann, Los Altos, CA, 1990.
[41] G. Kannan, R. Gosukonda, A.K. Mahapatra, Prediction of stress responses in goats: comparison of artificial neural network and multiple regression models, Can. J. Anim. Sci. 100 (2020) 102–110.
[42] L. Luttmann, P. Mercorelli, Comparison of backpropagation and Kalman filter-based training for neural networks, Preprints (2021), 2021040523.
[43] W.S. Sarle, Neural Networks and Statistical Models, Proceedings of the Nineteenth Annual SAS Users Group International Conference, SAS Institute, USA, 1994, pp. 1538–1550. Cary, NC.
[44] R. Dong, G. Zhao, The use of artificial neural network for modeling *in vitro* rumen methane production using the CNCPS carbohydrate fractions as dietary variables, Livest. Sci. 162 (2014) 159–167.
[45] M. Lotfi, M. Mars, S. Werbrouck, Optimizing pear micropropagation and rooting with light emitting diodes and trans-cinnamic acid, Plant Growth Regul. 88 (2019) 173–180.
[46] S. Jeyamkondan, D.S. Jayas, R.A. Holley, Microbial modeling with artificial neural networks, Int. J. Food Microbiol. 64 (2001) 343–354.
[47] T. Ross, Indices for performance evaluation of predictive models in food microbiology, J. Appl. Bacteriol. 81 (1996) 501–508.