

Novel application of machine learning models to predict secondary metabolites in *Dracocephalum moldavica* L

sharareh najafian

sh.najafian@saadi.shirazu.ac.ir

Department of Natural Resources and Environment Engineering, School of Agriculture, Shiraz University, Shiraz, IR Iran.

Shahnaz Fathi

Department of Medicinal and aromatic Plants, Shahid Bakeri High Education Center of Miandoab, Urmia University, Urmia, Iran.

Sajad Heidari

Department of Biosystems Engineering, Shiraz University, Shiraz, Iran.

Roya Movlodzadeh

Department of Medicinal and aromatic Plants, Shahid Bakeri High Education Center of Miandoab, Urmia University, Urmia, Iran.

Research Article

Keywords: Biochemical, Cleaner production, Ecological sustainability, Secondary metabolites, *Dracocephalum moldavica*, Machine learning, Proline accumulation

Posted Date: December 9th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-8198840/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Novel application of machine learning models to predict secondary metabolites in *Dracocephalum moldavica* L

Sharareh Najafian^{1*}, Shahnaz Fathi², Sajad Heidari^{3*} and Roya Movlodzadeh²

¹ Department of Natural Resources and Environment Engineering, School of Agriculture, Shiraz University, Shiraz, IR Iran.

² Department of Medicinal and aromatic Plants, Shahid Bakeri High Education Center of Miandoab, Urmia University, Urmia, Iran.

³ Department of Biosystems Engineering, Shiraz University, Shiraz, Iran.

*Corresponding Authors: E-mail address: sh.najafian@saadi.shirazu.ac.ir; sajad_heidari@yahoo.com

Abstract

This study aimed to predict the levels of key secondary metabolites-carotenoids, the amino acid proline, and soluble sugars-in *Dracocephalum moldavica* under salinity stress using artificial neural networks (ANNs). Two ANN models, namely the Radial Basis Function (RBF) and Multilayer Perceptron (MLP), were employed.

The models used 13 morphological and physiological traits, including leaf length, leaf width, number of leaves, number of internodes, crown diameter, stem length, internode length, number of lateral stems, lateral stem length, root length, shoot fresh weight, shoot dry weight, and relative water content, as inputs. Model performance was evaluated using a 5-fold cross-validation approach to ensure robust and reliable predictions.

The RBF model consistently outperformed the MLP model, demonstrating a superior ability to capture nonlinear relationships between morphological traits and biochemical responses. While the MLP model provided reasonable estimates, it was less accurate for all target metabolites.

Within the scope of the current dataset and greenhouse conditions, the RBF model proved to be a reliable tool for predicting secondary metabolites. This approach offers a rapid and cost-effective method to estimate key biochemical compounds, although predictions should be interpreted with caution. Future studies using independent external datasets are encouraged to further assess the model's generalizability.

Keywords: Biochemical, Cleaner production, Ecological sustainability, Secondary metabolites, *Dracocephalum moldavica*, Machine learning, Proline accumulation.

33 Introduction

34 For centuries, aromatic and medicinal plants have been valued for their rich content of bioactive
35 compounds, which contribute to pharmaceutical, food, and cosmetic applications. *Dracocephalum*
36 *moldavica* L., commonly known as Moldavian balm, is an aromatic annual herb rich in essential
37 oils such as geranial, neral, and geranyl acetate, which impart its characteristic citrus aroma and
38 confer antioxidant and antimicrobial properties [1]. Environmental stresses, including salinity,
39 drought, and temperature fluctuations, can significantly affect plant metabolism, often reducing
40 both yield and quality. Plants respond to such stresses by accumulating secondary metabolites,
41 such as proline, carotenoids, and soluble sugars, which enhance stress tolerance through osmotic
42 balance, photoprotection, and antioxidant defense [2]. Traditional biochemical assays for
43 measuring these metabolites are often labor-intensive, time-consuming, and involve hazardous
44 chemicals. Recently, machine learning (ML) has emerged as a powerful, non-destructive tool in
45 plant science, capable of revealing hidden patterns in complex datasets and providing accurate
46 predictions of biochemical responses under varying environmental conditions [3-7].

47 While ML approaches such as Random Forest, Support Vector Regression (SVR), and XGBoost
48 have been applied in some plant studies, artificial neural networks (ANNs) are particularly suited
49 for capturing complex nonlinear interactions between multiple input traits and biochemical
50 outputs, which is essential for modeling stress-induced metabolite variations. Several studies have
51 successfully applied ML to predict plant secondary metabolites. For example, Tan et al. [3]
52 modeled and optimized secondary metabolite production in *Spirodela polyrhiza* using ANNs,
53 while Khanna et al. [4] predicted metabolite profiles in medicinal plants based on macronutrient
54 inputs. ML-integrated omics platforms have also enabled the identification of genes involved in
55 specialized metabolite biosynthesis [5], and deep learning tools such as Spec2Class have classified
56 plant metabolites into functional chemical super classes [7].

57 Despite these advances, the application of ML, particularly ANNs, to predict stress-related
58 secondary metabolites in *D. moldavica* remains limited. Most previous studies have focused on
59 microbial or fungal systems [8,9], leaving herbaceous medicinal plants underexplored.

To address this knowledge gap, the current study employs two ANN models-Multilayer Perceptron (MLP) and Radial Basis Function (RBF)-to predict major stress-related secondary metabolites in *D. moldavica*. Thirteen easily measurable morphological and physiological traits-including leaf length, leaf width, number of leaves, number of internodes, crown diameter, stem length, internode length, number of lateral stems, lateral stem length, root length, shoot fresh weight, shoot dry weight, and relative water content-were used as model inputs. This approach not only minimizes reliance on destructive chemical assays but also provides a framework for developing robust, data-driven tools to support sustainable cultivation and biochemical optimization of medicinal plants. The present study aims to: (1) implement RBF and MLP neural networks for predicting critical secondary metabolites in *D. moldavica* under salinity stress; (2) compare the predictive performance and reliability of these models; and (3) demonstrate the utility of easily measurable morphological traits for rapid, cost-effective, and non-destructive assessment of plant biochemical responses. By integrating AI-driven predictions into medicinal plant cultivation, this work contributes to sustainable agricultural practices while supporting efficient biochemical monitoring, stress evaluation, and optimization of medicinal plant products.

Materials and Methods

Experimental Conditions in the Greenhouse

The experiment was conducted at the research farm of Shahid Bakeri Miandoab Higher Education Center, Isfahan, Iran. Certified seeds of *Dracocephalum moldavica* L. were obtained from Pakan Bazr Company. Seeds were sterilized using distilled water followed by a 10% sodium hypochlorite solution before sowing. Seeds were sown in 4 kg plastic pots containing a homogeneous mixture of soil, sand, and well-rotted animal manure (1:1:1, v/v/v). Each pot received 20 seeds. Greenhouse conditions were maintained at 30°C during the day and 20°C at night, with relative humidity ranging from 60% to 80%.

Plant species identification was verified by Abolfazl Alirezalu (Urmia University), and a voucher specimen (Herbarium Code: 1536) was deposited at the Herbarium of Urmia University. Salinity stress was imposed using four levels of sodium chloride (0, 20, 40, and 60 mM). Four biostimulants-Kadostim, Fosnutren, Humiforte, and Aminolforte-e-were applied as foliar sprays at two growth stages: the 4-6 leaf stage and the onset of flowering. Bio stimulant solutions were prepared by dissolving the commercial formulations in distilled water to achieve a concentration

of approximately 3.75 g/L of total free amino acids, as indicated in Table 1. The spray volume was 10 mL per plant and was applied uniformly until runoff using a handheld sprayer. The chemical composition of the bio stimulants, including free amino acids, organic matter, and nutrient contents (N, P₂O₅, K₂O), is presented in Table 1.

Table 1. The compounds of amino acid, organic fertilizers

Bio stimulants	Free amino acids (mg/L)	Organic matters (%)	Nitrogen (%)	P ₂ O ₅ (%)	K ₂ O (%)
Kadostim	3750	2	-	3	-
Fosnutren	3750	2	3.8	6	-
Humiforte	3750	2	6	5	3
Aminolforte	3750	2	1.1	-	-

Amount and type of amino acids= Glysin 11.2 %, Valine 5.1%, Proline 8.3 %, Alanin 13.2 %, Aspartic acid 4.4 %, Arginine 8.3 %, Glutamic acid 0.9 %, Lysine 5.1 %, Lucine 16.4 %, Isolucine 4.4 %, Phenylalanin 5.1 %, Methionine 4.2 %, Serin 3.9 %, Treonine 0.3 %, Histidine 0.3 %, Tyrosine 1.5 %, Glutamine 0.9 %, Systeine 0.3 %, Asparagine 0.4 %, and Tryptophan 0.4 %.

Dataset Size, Experimental Design, and Statistical Analysis of Treatments

The experiment was arranged in a factorial design with four salinity levels (0, 20, 40, and 60 mM NaCl) and four amino acid-based bio stimulants (Kadostim, Fosnutren, Humiforte, and Aminolforte-e). Each treatment combination included three replicates, with seven plants per replicate, resulting in a total of 336 plants grown under greenhouse conditions. For each plant, 13 morphological and physiological traits-leaf length, leaf width, number of leaves, number of internodes, crown diameter, stem length, internode length, number of lateral stems, lateral stem length, root length, shoot fresh weight, shoot dry weight, and relative water content-were recorded, along with three biochemical traits: leaf proline, leaf carotenoids, and shoot soluble sugars. This resulted in a final dataset of 336 observations, each containing 13 input traits and 3 target variables for artificial neural network modeling, ensuring transparency and reproducibility.

To evaluate the effects of salinity and bio stimulant treatments, data were analyzed using two-way ANOVA to assess both main and interaction effects. Mean comparisons were performed using Duncan's multiple range test at a 5% significance level. Normality and homogeneity of variance were verified, and data were transformed when necessary. The results showed that higher salinity significantly reduced plant growth and relative water content, while increasing biochemical osmolytes (proline, carotenoids, and soluble sugars). Foliar application of amino acid-based bio

stimulants effectively alleviated the negative effects of salinity stress, with Kadostim showing the strongest positive impact, followed by Aminolforte. These classical statistical findings confirmed the biological significance of the treatments and complemented the ANN predictions, highlighting the potential of predicting key biochemical traits from easily measurable morphological and physiological characteristics.

Measurement of Growth Parameters

At the start of flowering, plant growth parameters were measured. Crown diameter (mm) was determined using a caliper, and stem length (cm) was measured with a ruler. Internode length (cm) was measured with a caliper, and the number of internodes was counted from the stem base to the tip. The number of lateral stems was recorded, and the length of each lateral stem was measured with a ruler. Leaf width (mm) was measured using a caliper, and leaf length (mm) was measured with a ruler. The total number of leaves per plant was recorded.

Root length (cm) was measured from the root tip to the root collar after carefully removing the surrounding soil. Shoot fresh weight was measured using an analytical balance, and shoot dry weight was determined after drying samples for 48 hours at 70°C. Relative water content (RWC) was calculated according to the method described by Moloudzadeh et al. [10].

Quantification of Soluble Sugars

Soluble sugars were determined using a modified method of Dubois et al. [11]. For each sample, 0.1 g of fresh plant material was weighed using a digital scale and homogenized in 10 mL of distilled water. The homogenate was filtered twice to obtain a clear extract.

One milliliter of the filtered extract was transferred to a sterile test tube and mixed with 1 mL of 5% phenol solution. Five milliliters of concentrated sulfuric acid (H_2SO_4) were then added slowly to the mixture. The reaction was allowed to proceed at room temperature for 10–15 minutes to ensure complete hydrolysis of sugars and the formation of a yellowish-brown complex.

Absorbance of the resulting solution was measured at 485 nm using a spectrophotometer. A standard glucose solution was prepared, and absorbance values of known concentrations were used to construct a calibration curve. Soluble sugar concentrations in the plant samples were calculated by comparison with the calibration curve and expressed as milligrams of glucose equivalents per gram of fresh weight. Careful preparation of the standard curve, accurate reagent measurement, proper incubation, and clean spectrophotometric tubes are critical to ensuring precise quantification of soluble sugars.

148

149 **Measurement of carotenoids in plant leaves**

150 Carotenoid content in plant leaves was determined using a modified method of Lichtenthaler and
151 Wellburn (1983) [12]. Fresh leaves (0.1 g) were homogenized in 10 mL of 80% acetone until a
152 uniform extract was obtained. The extract was transferred to centrifuge tubes and stored in darkness
153 at 4°C for 24 hours. After storage, samples were centrifuged at 4000 rpm for 10 minutes, and the
154 supernatant was collected. Absorbance was measured at 663, 645, and 470 nm to determine
155 chlorophyll a, chlorophyll b, and total carotenoids, respectively. Concentrations were calculated
156 using the equations provided in the original method [10].

157

158 **Measurement of Proline in plant leaves**

159 Proline content in plant leaves was determined using the method of Bates et al. [13]. Fresh leaf
160 tissue (0.2 g) was homogenized in 3 mL of 3% sulfosalicylic acid using a porcelain mortar. The
161 homogenate was centrifuged at 18,000 rpm for 15 minutes.

162 Two milliliters of glacial acetic acid and two milliliters of ninhydrin reagent were added to the
163 supernatant. The mixture was heated at 100°C in a water bath for 1 hour. After cooling, 4 mL of
164 toluene was added, and the mixture was vortexed for 15 seconds. The red chromophore containing
165 proline dissolved in toluene was separated, and its absorbance was measured at 520 nm using a
166 spectrophotometer. Proline concentration was calculated and expressed as mg per g of fresh leaf
167 tissue.

168 **Artificial neural network (ANN) models**

169 Artificial neural network (ANN) models were employed to predict target variables based on
170 measured plant traits. ANNs consist of interconnected nodes (neurons) that process input data
171 through weighted connections. Input variables included [specify your measured traits], and output
172 variables were [specify the predicted traits]. The models were trained using [training method, e.g.,
173 backpropagation] and validated with [validation approach, e.g., cross-validation or holdout
174 dataset]. Model performance was evaluated based on [specify metrics, e.g., R^2 , RMSE, MAE] to
175 assess predictive accuracy.

176 **Multilayer Perceptron Neural Network (MLPNN)**

177 The multilayer perceptron neural network (MLPNN) was implemented as a feedforward ANN
178 consisting of an input layer, one or more hidden layers, and an output layer (Fig. 1). Each layer

contained fully connected neurons characterized by input weights, a bias term, and a non-linear activation function [15]. The mathematical operation of each neuron (k) is described by equations (1) and (2):

$$y_k = f(u_k + b_k) \quad (1)$$

$$u_k = \sum_{i=1}^N w_{ki} x_i \quad (2)$$

It is important to note that in these equations, x_i denotes the input signals, w_{ki} are the corresponding weights, b is the bias, and f is a differentiable non-linear activation function. In this study, the sigmoid function was used due to its smoothness and effectiveness in gradient-based optimization.

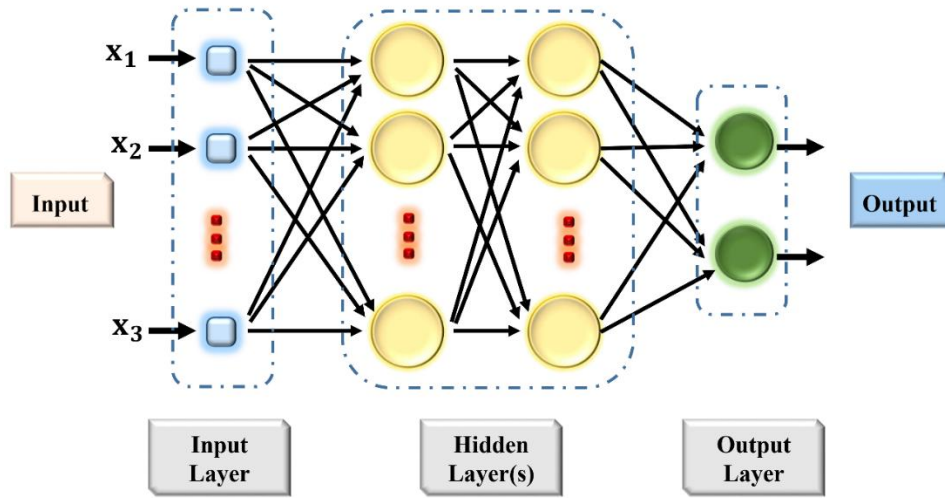


Fig.1. The structure of a Multilayer Perceptron Neural Network (MLPNN).

MLPNNs were trained using the Levenberg–Marquardt optimization algorithm, a second-order method suitable for small to medium-sized datasets [16]. The algorithm adjusted network weights and biases iteratively to minimize prediction error. A validation-based stopping criterion was applied to prevent overfitting, terminating training when validation error increased over consecutive iterations. Optimal hyperparameters were selected based on model performance metrics, as reported in the Results section.

Radial Basis Function Neural Networks (RBFNN)

In this study, a Radial Basis Function Neural Network (RBFNN) was utilized as one of the two modeling strategies. RBFNNs are a class of feedforward artificial neural networks widely used in engineering and pattern recognition tasks due to their simple structure, fast learning capability, and robustness to noisy inputs. As shown in Fig 2, a typical RBFNN consists of three layers: an input

layer, a hidden layer that applies a non-linear transformation using radial basis functions (RBFs), and a linear output layer [17].

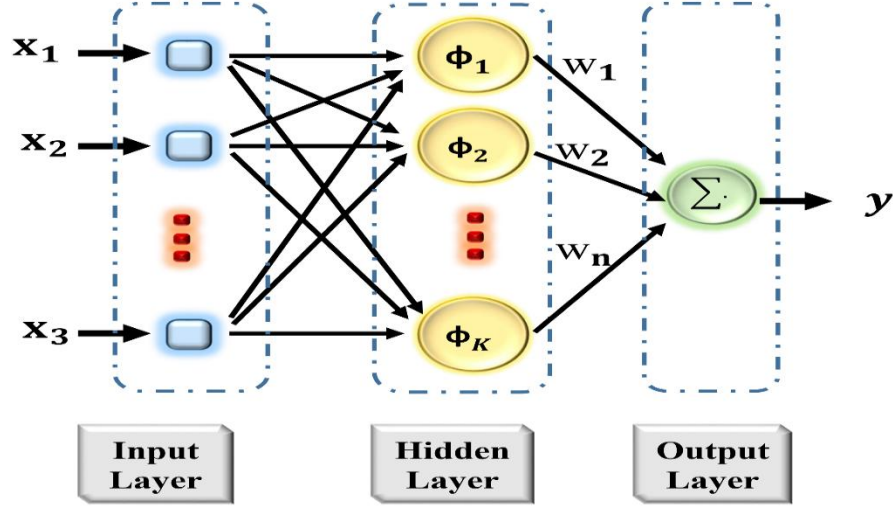


Fig. 2. The structure of a radial basis function neural networks.

The hidden layer is the core of the RBFNN. Each hidden neuron computes a radial basis function centered at a specific point in the input space. The output of the network is calculated as a weighted sum of these non-linear activations. The output of an RBFNN is mathematically described by Equation 3

$$y_i(x) = \sum_{j=1}^k w_{ij} \phi(\|x - c_j\|) \quad (3)$$

In the formula you see, the symbol x represents the input vector, the symbol y_i represents the i -th output of the network, the symbol k represents the number of neurons in the hidden layer, the symbol c_j is the center associated with the j -th hidden neuron, and finally w_{ij} corresponds to the weight that connects the j -th hidden neuron to the i -th output neuron, and $\| \cdot \|$ represents the Euclidean distance. The function ϕ serves as the radial basis function used by the hidden neurons, which quantifies the distance between the input vector and a designated center vector. Among the various types of radial basis functions commonly investigated, the Gaussian function is the most widely used. It is mathematically defined by Equation (4):

$$\phi(\|x - c_j\|) = e^{\left(-\frac{\|x - c_j\|^2}{2\sigma_j^2}\right)} \quad (4)$$

In this context, σ_j represents the width parameter of the j -th hidden neuron. Designing an effective RBF neural network involves selecting appropriate values for the centers, widths, and output

weights. In most cases, these parameters are estimated based on the characteristics of the input data. Additionally, the number of hidden neurons is not predetermined and is usually chosen through a trial-and-error process to achieve the best prediction performance [16, 18]. The selection of the most appropriate training function and optimal spread parameter was based on the comparative evaluation of model accuracy, as detailed in the Results section.

Data partitioning and K-fold cross-validation

It should be noted that some of the input variables (13 physiological and morphological traits) and the target outputs (proline, carotenoids, and soluble sugar concentrations) had different numerical scales; therefore, all variables were normalized before being used in two types of neural networks, MLP and RBF. This preprocessing step involved standardizing the data to a mean of zero and a standard deviation of one. Such adjustment improved training efficiency and contributed to better overall model performance.

To prevent overfitting, improve the generalizability of the developed models, and ensure more reliable usage of the available dataset, a k-fold cross-validation approach (with $k = 5$) was employed in this study. As illustrated in Figure 3, the entire dataset was randomly partitioned into five folds. At each iteration, one-fold was considered as the test set, while the remaining four folds were used as training data. After repeating the process k times, the overall performance of the models was evaluated by averaging the error metrics across all folds. By using K-fold cross-validation, all data samples will eventually go through both training and testing, which is an advantage. For computational analysis, we used MATLAB software (R2017b) in this study to develop RBF and MLP neural network models [19, 20].



Fig. 3. Illustration of the 5-fold cross-validation process used for training and evaluating neural network models

Model Performance Evaluation Metrics

metabolites in the medicinal plant *D. moldavica*, including leaf amino acids (proline), leaf carotenoids, and shoot soluble sugars. Three essential statistical criteria were employed to evaluate model performance: (1) Root Mean Square Error (RMSE), (2) Mean Absolute Percentage Error (MAPE), and (3) Coefficient of Determination (R^2). These criteria are mathematically defined in Equations (5), (6), and (7) and provide a comprehensive framework for assessing model accuracy.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{(Y_{est})_i - (Y_{meas})_i}{(Y_{meas})_i} \right| \times 100 \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n [(Y_{est})_i - (Y_{meas})_i]^2}{\sum_{i=1}^n (Y_{est} - \bar{Y})_i^2} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [(Y_{est})_i - (Y_{meas})_i]^2} \quad (7)$$

Here, n represents the total number of data points, Y_{est} and Y_{meas} represent the predicted and observed concentrations, and $\bar{Y}$ is the mean of observed values. Optimal model performance corresponds to minimized RMSE and MAPE and maximized R^2 , indicating the best agreement between predicted and observed values [21, 22].

It should be noted that some of the input variables (13 physiological and morphological traits) and the target outputs (proline, carotenoids, and soluble sugar concentrations) had different numerical scales; therefore, all variables were standardized (mean = 0, standard deviation = 1) before being used in the MLP and RBF neural network models. This approach was preferred over min–max scaling because it provided better numerical stability for variables with heterogeneous ranges and reduced the models' sensitivity to outliers. Such standardization improved training efficiency and contributed to better overall model performance. In contrast, min–max scaling can be more sensitive to outliers, which may exist in physiological data. Therefore, we standardized the data to improve model stability and generalization performance.

Feature Importance Analysis

To assess the relative contribution of each morphological and physiological trait to metabolite prediction, a sensitivity analysis was carried out using Olden's algorithm [23] implemented in MATLAB. This algorithm quantifies the importance of input variables based on the connection weights between the input–hidden and hidden–output layers of artificial neural networks.

In contrast to Garson's algorithm [24], which relies on the absolute magnitude of weights, Olden's method retains the sign of the connection weights, thus providing both the magnitude and direction of each predictor's influence.

Previous studies have demonstrated that Olden's approach produces reliable and interpretable estimates of variable importance compared with other connection-weight–based techniques [25].

In the present study, this analysis was performed separately for each target metabolite, including carotenoids, proline, and soluble sugars, and the resulting scores indicate how each input trait positively or negatively contributes to the predicted metabolite levels.

Results and Discussion

Predictive Modeling of Key Biochemical Traits

This study focused on modeling the concentrations of three key biochemical compounds-amino acid proline, carotenoids, and soluble sugars-in the stems of *Dracocephalum moldavica* using two artificial neural network (ANN) models: multilayer perceptron (MLP) and radial basis function (RBF). The experimental data were obtained from a salinity stress study approved by, and involving, the original researcher.

Thirteen physiological and morphological traits were selected as input variables based on their practical measurability under field or greenhouse conditions. These included stem length (cm), crown diameter (mm), internode length (cm), number of internodes, number of lateral stems, lateral stem length (cm), leaf width (mm), leaf length (mm), number of leaves, stem relative weight, root weight, dry weight of branches (g), and dry stem length (cm). These features were chosen due to their potential associations with biochemical responses under abiotic stress.

The output variables-proline, carotenoids, and soluble sugars-are widely recognized as reliable biochemical markers of plant responses to salinity stress. These compounds are typically quantified using time-consuming, costly, and chemically intensive laboratory methods. Therefore, this study aimed to develop a predictive framework that enables indirect, yet accurate, estimation

of these metabolites using readily measurable morphological inputs. Such an approach may reduce the dependency on destructive biochemical assays and enhance the efficiency of plant stress response evaluations.

Performance Evaluation of the RBF Model

Prior to training the RBF neural network, as detailed in the *Materials and Methods* section, all input variables (13 physiological and morphological traits) and target outputs (proline, carotenoids, and soluble sugars) were normalized to a mean of zero and a standard deviation of one. In addition, a five-fold cross-validation strategy was employed to ensure reliable model evaluation and minimize overfitting. Achieving optimal performance in the RBF model requires careful and systematic tuning of its architectural and functional parameters, particularly the choice of the training algorithm and the spread value. To identify the most suitable RBF configuration, several training algorithms available in MATLAB were evaluated, including the radial basis network (newrb), exact radial basis network (newrbe), generalized regression neural network (newgrnn), and probabilistic neural network (newpnn). Among these configurations, the RBF model implemented using the newgrnn function demonstrated superior predictive accuracy, yielding the lowest mean squared error (MSE) and the highest coefficient of determination (R^2) across all tested configurations. The spread parameter, which determines the width of the Gaussian basis functions in the RBF network and plays a crucial role in balancing model smoothness and generalization, was optimized through a systematic evaluation. Spread values ranging from 0.10 to 1.00 (with a step size of 0.01) were examined, and for each value, the model was trained and validated under the 5-fold cross-validation framework. The RMSE exhibited a convex trend across the tested values, reaching its minimum at a spread of 0.529 with an RMSE of 0.484, which was consequently selected as the optimal spread for the final GRNN configuration, as illustrated in Figure 4.

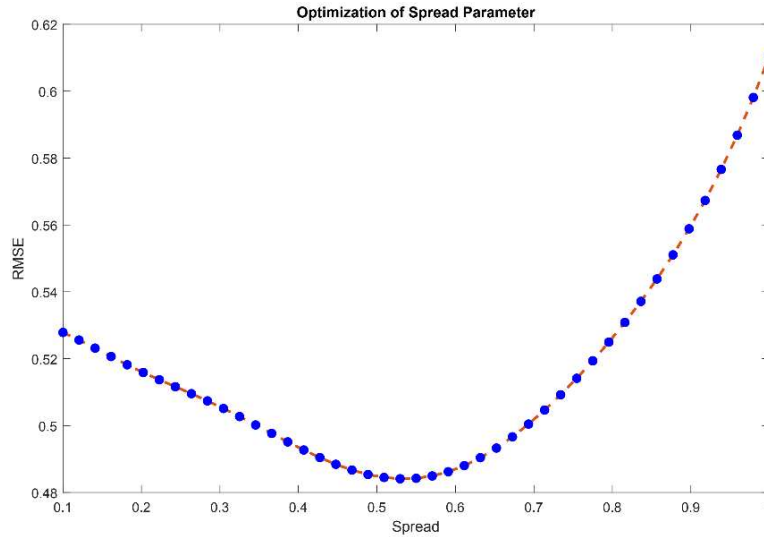


Fig. 4. Optimization of the spread constant in the RBF model using RMSE as a performance criterion.

The optimal configuration of the RBF network was achieved through an iterative evaluation process using MATLAB's `newgrnn` function. The model was trained and validated under a five-fold cross-validation framework, and its predictive performance was systematically assessed based on the error metrics obtained from the test sets. The finalized GRNN-based RBF model achieved a consistent balance between accuracy and generalization, confirming its reliability for modeling complex nonlinear relationships in the dataset.

Performance Evaluation of the MLP Model

The MLP model was trained and evaluated using the same normalization procedure and five-fold cross-validation strategy as those used for the RBF model. The optimal architecture of the MLPNN model was identified through a grid search across various network configurations by tuning key hyperparameters, including the number of hidden layers, neurons per layer, training algorithms, and activation functions. The most effective network structure was determined by evaluating each configuration based on R^2 , RMSE, and MAPE metrics. The optimal number of neurons in this layer was determined by testing configurations ranging from 5 to 25, with the best overall performance achieved with 13 neurons. Several training algorithms were evaluated, including Levenberg–Marquardt, Bayesian regularization, scaled conjugate gradient, steepest descent, elastic backpropagation, and Newton's method. Among these, the Levenberg–Marquardt algorithm yielded the lowest validation error and was thus selected as the most effective optimizer, consistent with the findings of Nait Amar et al. (2022). Furthermore, among the tested activation

functions (log-sigmoid, tanh, and ReLU), the log-sigmoid function resulted in the highest predictive accuracy and was consequently adopted for the hidden layer [26]. In addition, a validation monitoring strategy was incorporated into the training process to prevent overfitting by halting the training if the validation error increased over several consecutive predefined iterations. The key structural and training parameters of both the RBF and MLP neural network models are summarized in Table 2 for clarity.

Table 2. Summary of the key training parameters of the developed RBFNN and MLPNN models.

Hyperparameters Models	Hidden Layers	Neurons per Hidden Layer	Activation Function	Training Algorithm	Spread
RBF	1	Equal to number of training samples	Radial basis (Gaussian)	Generalized Regression Neural Network (newgrnn)	0.529
MLP	1	13	Log- sigmoid	Levenberg– Marquardt (LM)	-

Evaluation of Neural Network Models for Predicting Key Biochemical Compounds

Table 3 presents the performance metrics of the MLP and RBF models in predicting carotenoid, proline, and shoot soluble sugar concentrations, based on a 5-fold cross-validation strategy. Model performance was evaluated using three standard indicators: coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE). These metrics were calculated separately for the training set, test set, and the combined dataset.

The use of cross-validation allowed for a more robust assessment of each model’s ability to generalize across different data partitions. This evaluation framework helps to reduce the risk of overfitting and provides a clearer understanding of how well each network architecture performs under varied data conditions.

Table 3. Performance comparison of RBFNN and MLPNN models for predicting carotenoid, proline, and shoot soluble sugar.

	Model types	Train			Test			Total		
		R ²	RMSE	MAPE	R ²	RMSE	MAP	R ²	RMSE	MAPE
Carotenoid	MLPNN	0.948	0.257	2.010	0.858	0.386	2.612	0.935	0.299	2.291
	RBFNN	0.996	0.066	0.412	0.902	0.295	1.971	0.983	0.148	0.783
Proline	MLPNN	0.963	0.170	1.097	0.903	0.289	1.869	0.957	0.199	1.371
	RBFNN	0.999	0.052	0.556	0.975	0.235	1.124	0.996	0.115	0.831
Shoot soluble sugar	MLPNN	0.944	1.166	1.338	0.876	1.619	1.965	0.937	1.276	1.463
	RBFNN	0.999	0.178	0.206	0.941	0.948	1.229	0.992	0.455	0.411

Carotenoid Prediction

The prediction results for carotenoid content in *D. moldavica* indicated that the RBFNN model achieved better predictive accuracy than the MLPNN model. Specifically, the RBFNN model produced an R² of 0.983, RMSE of 0.148, and MAPE of 0.783, while the MLPNN model resulted in an R² of 0.935, RMSE of 0.299, and MAPE of 2.291. These findings suggest that RBFNN had lower prediction error and stronger goodness-of-fit, particularly in modeling carotenoid levels under the given experimental conditions.

The improved performance of RBFNN may be attributed to its structure, which facilitates the modeling of non-linear relationships between easily measurable morphological traits and carotenoid content. Unlike the MLPNN, which uses multiple hidden layers, RBFNN relies on localized radial basis functions to approximate input-output mappings more efficiently. This structural difference likely contributed to the higher predictive accuracy observed in this study.

The superior performance of RBFNN in this context is consistent with findings from previous studies. For instance, Li et al. [27] developed an RBF-based model for shelf-life prediction in grapes, demonstrating its high predictive reliability. Similarly, González-Camacho et al. [28] highlighted the effectiveness of RBFNN in estimating plant metabolite levels, particularly when dealing with complex, nonlinear traits. While other modeling approaches, such as random forests or traditional regression, have been explored in different contexts [27], the current study did not evaluate them directly. Thus, further comparative studies may be warranted to validate these findings across different plant systems and modeling techniques.

Overall, the results support the use of RBFNN as a robust tool for predicting carotenoid content in medicinal plants using easily measurable morphological inputs, though future research should investigate broader model comparisons and validate the findings under diverse environmental conditions.

Proline Prediction

The RBFNN model outperformed the MLPNN model in predicting proline concentrations. Based on the complete dataset, RBFNN achieved a higher R^2 (0.996), and lower RMSE (0.115) and MAPE (0.831) compared to MLPNN ($R^2 = 0.957$, RMSE = 0.199, MAPE = 1.371). On the experimental dataset, RBFNN also showed better predictive ability ($R^2 = 0.975$, RMSE = 0.235, MAPE = 1.124) than MLPNN ($R^2 = 0.903$, RMSE = 0.289, MAPE = 1.869).

These findings align with previous research by Zhang and Wang [29], who demonstrated the robustness of RBFNN in predicting secondary metabolites in medicinal plants. Proline, as a key Osmo protectant under stress conditions, has been the subject of accurate modeling using machine learning. Our results reinforce the potential of RBFNN as a reliable tool for such predictions. Prior studies have also highlighted the role of machine learning in phytochemical modeling supporting our approach.

Shoot Soluble Sugar Prediction

The RBFNN model demonstrated a higher predictive accuracy for shoot soluble sugar levels, achieving an overall R^2 of 0.992, RMSE of 0.455, and MAPE of 0.411 across the complete dataset. In contrast, the MLPNN model showed lower performance with an overall R^2 of 0.937, RMSE of 1.276, and MAPE of 1.463. The superiority of the RBFNN model was consistent across other data subsets, further validating its effectiveness.

The regression plots in Figures 5 and 6 visually confirm this trend, with the RBFNN model yielding R^2 values of 0.983, 0.996, and 0.992 for carotenoid, proline, and shoot soluble sugar, respectively, compared to the MLPNN.

This observation is consistent with previous studies, such as Tan et al. [30], who employed RBFNN in modeling metabolite production in *Spirodela polyrhiza*. Other studies [29, 31] have similarly highlighted the capability of RBFNN in predicting secondary metabolites in medicinal plants. While these studies differ in species and target metabolites, they collectively support the general utility of RBFNN for modeling complex biochemical traits in plants. Our results align with

this trend and further demonstrate the effectiveness of RBFNN under the tested conditions, particularly for soluble sugar prediction.

In contrast, the MLP model (Fig. 6) yielded R^2 values of 0.935, 0.957, and 0.937 for carotenoid, proline, and shoot soluble sugar, respectively. Although both models demonstrated strong predictive capabilities, the RBF model showed marginally higher R^2 values for carotenoid and shoot soluble sugar. These results suggest that RBF may provide slightly better fit for this dataset, but both models are suitable for predicting these biochemical traits.

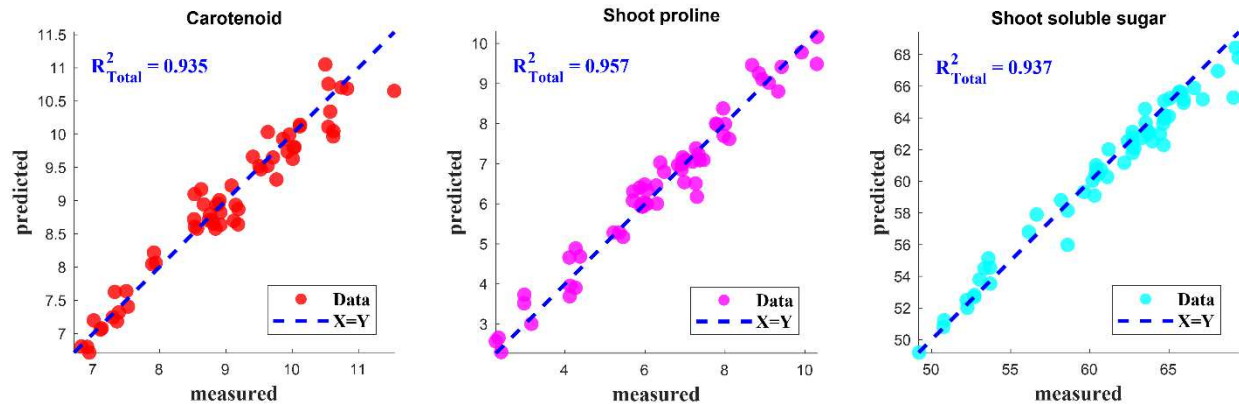


Fig.5. Regression of measured vs. predicted values for carotenoid, shoot proline, and shoot soluble sugar using the RBF neural network.

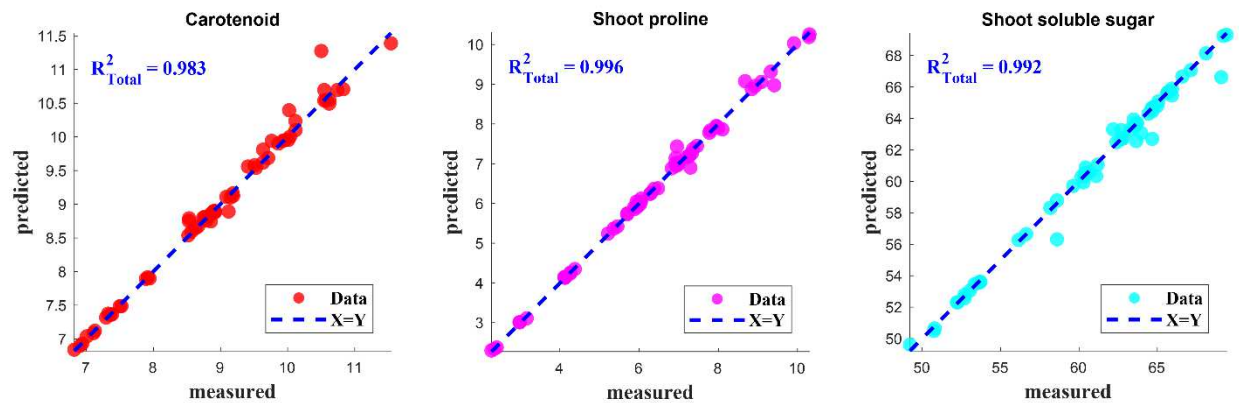


Fig. 6. Regression of measured vs. predicted values for carotenoid, shoot proline, and shoot soluble sugar using the MLP neural network.

To reduce the potential risk of overfitting, we implemented a 5-fold cross-validation procedure and normalized all input variables. Hyperparameters were optimized based on validation set

performance. Although an independent external dataset was not available, the consistent predictions across folds, along with low RMSE and MAPE values, suggest that the models are reliable within the scope of this study. Nevertheless, external validation using independent datasets is necessary in future work to fully confirm model generalizability and real-world applicability.

Comparative Evaluation of Neural Network Models for Predicting Carotenoid, Proline, and Shoot Soluble Sugar

This study compared the predictive performance of two neural network models, multilayer perceptron (MLP) and radial basis function (RBF), for estimating carotenoid, shoot proline, and shoot soluble sugar concentrations. The models were evaluated using RMSE, MAPE, and R^2 metrics based on the test dataset, as illustrated in Figures 7, 8, and 9.

Figure 7 shows that the RBF model consistently achieved lower RMSE values across all biochemical traits compared to the MLP model, indicating a reduction in prediction error. Similarly, Figure 8 illustrates that the RBF model produced lower MAPE values, with the most notable difference observed in shoot soluble sugar prediction. Figure 9 presents R^2 values, where the RBF model demonstrated higher explanatory power in capturing variance within the test data.

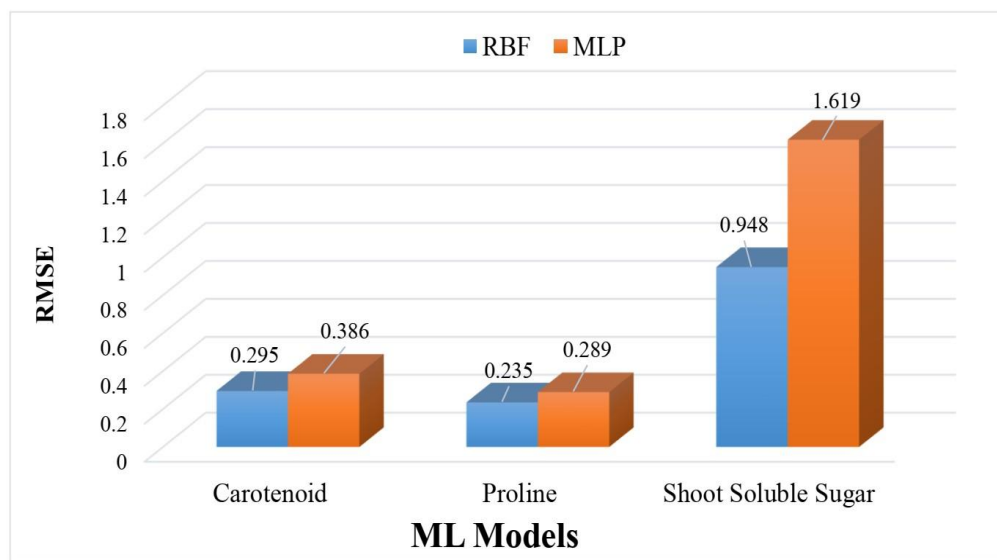


Fig. 7. Comparison of RMSE values for RBF and MLP models on test datasets for predicting carotenoid, shoot proline, and shoot soluble sugar levels.

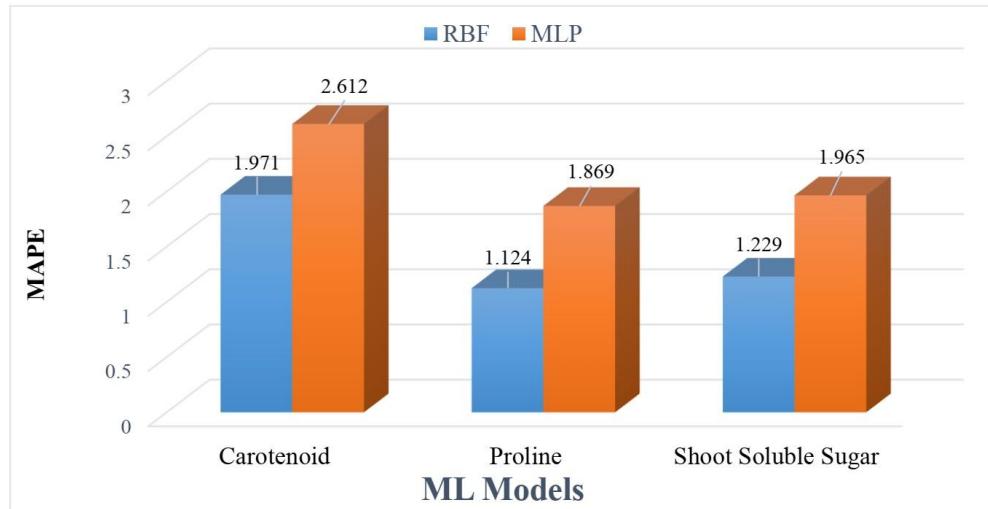


Fig. 8. Comparison of MAPE values for RBF and MLP models on test datasets for predicting carotenoid, shoot proline, and shoot soluble sugar levels.

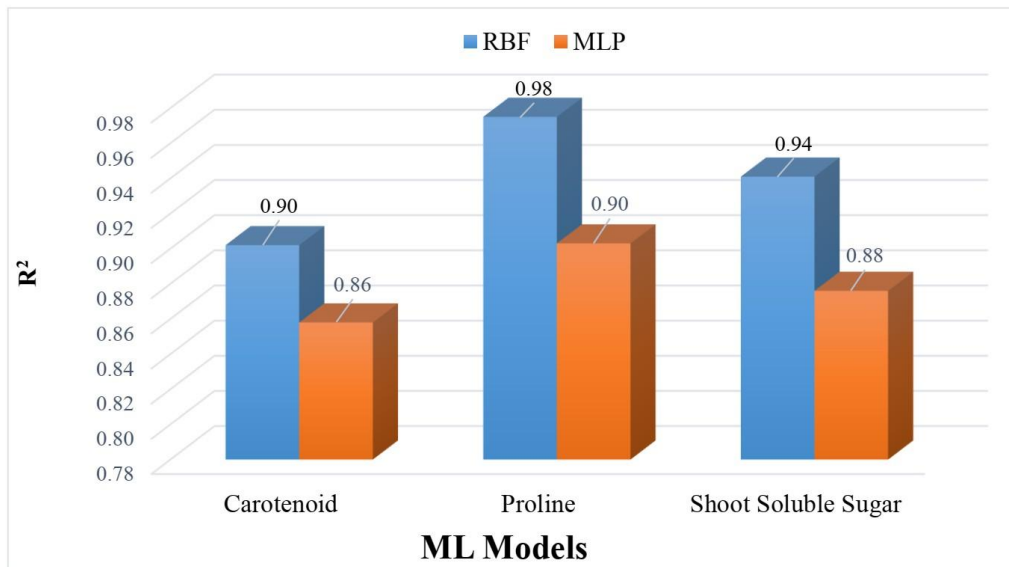


Fig. 9. Comparison of R^2 values for RBF and MLP models on test datasets for predicting carotenoid, shoot proline, and shoot soluble sugar levels.

These results suggest that, under the conditions of this study, the RBF model provided more accurate and reliable predictions than the MLP model for the biochemical traits considered. Previous research supports the applicability of RBF networks for modeling biochemical traits in

plants, largely due to their capability to model nonlinear relationships effectively [32]. Although MLP networks also accommodate nonlinearities, they may require larger datasets and more extensive tuning to reach comparable performance levels [33]. This is consistent with our findings, where the RBF model showed superior performance across all measured variables. Moreover, the flexibility of RBF networks in adapting to complex plant biochemical responses under stress conditions, such as salinity and drought, has been documented in various studies [34]. Such attributes highlight the potential of RBF models for applications in plant biotechnology and agronomy, particularly when precise prediction of biochemical parameters is critical.

Evaluation of Predicted and Actual Data Using the Optimized RBF Model

Table 4 summarizes key statistical parameters-including sum, mean, standard deviation, minimum, and maximum values-for both the actual measurements and the RBF model predictions across the three biochemical variables. The close agreement between predicted and observed values across these metrics indicates that the optimized RBF model reliably captured the variation in the dataset. Furthermore, the relatively small deviations observed in the statistical parameters support the model's accuracy and consistency in predicting carotenoid, proline, and soluble sugar concentrations.

Table 4. Comparison of observed and predicted values for carotenoid, proline, and shoot soluble sugar using the final RBF model.

Element	Phases	Data	Sum	Average	Standard deviation	Minimum	Maximum
Carotenoid	Total	Actual	539.407	8.990	1.178	6.822	11.540
		Predicted	540.836	9.014	1.197	6.841	11.390
	Train	Actual	428.075	8.918	1.247	6.822	11.540
		Predicted	428.066	8.918	1.238	6.841	11.390
	Test	Actual	111.332	9.278	0.786	7.943	10.622
		Predicted	112.770	9.397	0.925	7.899	11.275
Proline	Total	Actual	387.041	6.451	1.939	2.272	10.302
		Predicted	386.940	6.449	1.932	2.301	10.254
	Train	Actual	299.823	6.246	2.050	2.272	10.302
		Predicted	299.816	6.246	2.047	2.301	10.254
	Test	Actual	87.218	7.268	1.075	5.704	9.412
		Predicted	87.124	7.260	1.041	5.748	9.077

soluble sugar	Total	Actual	3660.866	61.014	5.175	49.214	69.420
		Predicted	3652.846	60.881	5.104	49.610	69.318
	Train	Actual	2893.552	60.282	5.327	49.214	69.420
		Predicted	2893.414	60.279	5.313	49.610	69.318
	Test	Actual	767.314	63.943	3.108	58.600	69.218
		Predicted	759.433	63.286	3.180	56.312	69.248

Evaluation of Predicted and Actual Data for Key Biochemical Traits

Carotenoids

For the carotenoid pigment dataset, the predicted mean concentration by the RBF model was 9.014, closely matching the observed mean of 8.990. The standard deviations were also similar (1.197 predicted vs. 1.178 observed), indicating that the model effectively replicated the distribution of carotenoid levels across the dataset. This consistency was evident in both training and testing subsets.

Proline

Similarly, the predicted mean proline concentration (6.449) closely approximated the actual mean (6.451). Standard deviations for predicted and observed values were nearly identical (1.932 vs. 1.939), demonstrating the RBF model’s reliable performance in estimating proline levels throughout the dataset.

Shoot Soluble Sugar

The soluble sugar concentration predictions also showed close agreement with measured values. The predicted mean (60.881) was very near the actual mean (61.014), with small differences in standard deviation (5.104 predicted vs. 5.175 observed). This pattern held true for both training and experimental subsets, supporting the robustness of the model.

Overall, the minimal differences between predicted and actual statistical parameters across all three biochemical traits suggest that the optimized RBF model provides accurate and consistent estimates, reinforcing its suitability for modeling key biochemical characteristics in *D. moldavica*.

Feature Importance and Sensitivity Analysis

The sensitivity analysis based on Olden’s algorithm revealed distinct patterns of trait influence for each predicted metabolite (Figure 10).

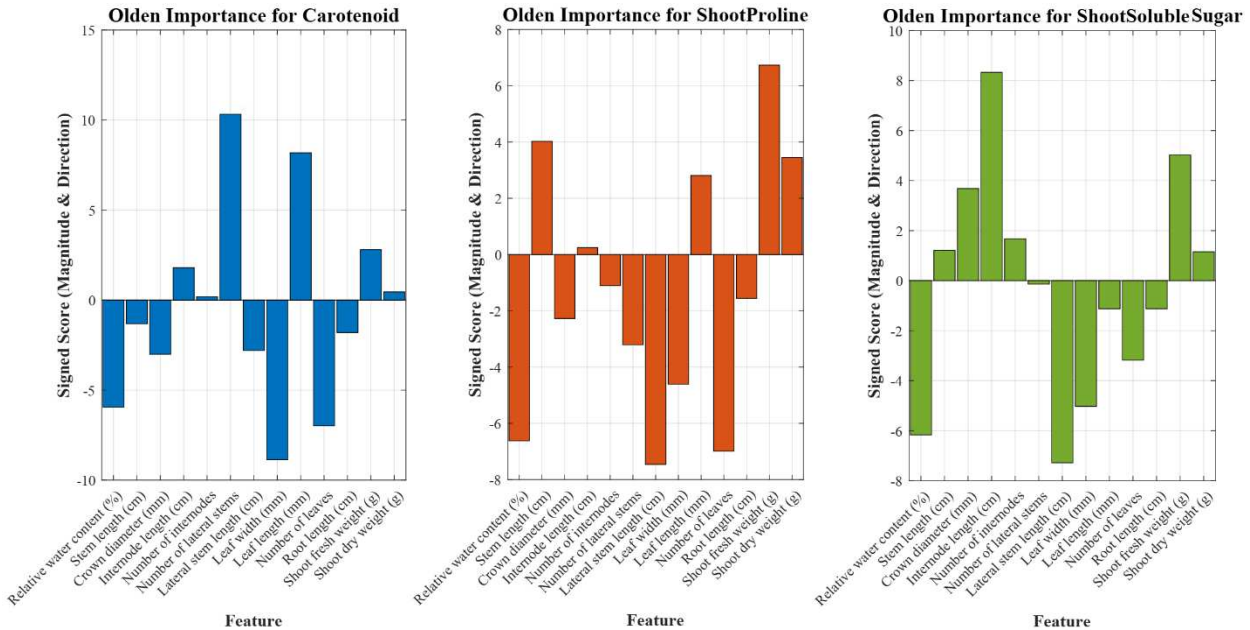


Fig. 10. Relative importance of morphological and physiological predictors for the artificial neural network (ANN) models based on Olden’s algorithm. Carotenoid, Proline, and Shoot soluble sugar.

For carotenoids, Olden’s algorithm assigned positive importance scores (in descending order) as follows: number of lateral stems, leaf length (mm), shoot fresh weight (g), internode length (cm), shoot dry weight (g), and number of internodes. These ordered positive contributions suggest that plants with greater branching, larger leaf area (as reflected by leaf length), and higher shoot biomass may have an enhanced capacity for carotenoid accumulation under salinity stress, possibly due to increased photosynthetic surface and greater demand for photoprotective pigments. Negative importance scores (in descending order of absolute influence) were observed for leaf width (mm), number of leaves, relative water content (%), crown diameter (mm), lateral stem length (cm), root length (cm), and stem length (cm), indicating that wider leaves, higher tissue water content, and elongation-related growth traits were inversely associated with carotenoid levels in stressed plants.

For proline, traits with the strongest positive influence were shoot fresh weight (g), stem length (cm), shoot dry weight (g), leaf length (mm), and internode length (cm). These results suggest that

greater shoot biomass and elongation growth are positively associated with proline accumulation, likely reflecting enhanced osmotic adjustment and structural adaptation under salinity stress. In contrast, lateral stem length (cm), number of leaves, relative water content (%), leaf width (mm), number of lateral stems, crown diameter (mm), root length (cm), and number of internodes showed negative importance values, indicating that canopy expansion and root elongation traits contributed less-or even inversely-to proline accumulation under stress conditions.

For soluble sugars, traits showing the strongest positive influence included internode length (cm), shoot fresh weight (g), crown diameter (mm), number of internodes, shoot dry weight (g), and stem length (cm). These associations suggest that plants with greater structural development and shoot biomass tend to accumulate more soluble sugars, likely due to enhance assimilate transport and storage capacity under salinity stress. Conversely, lateral stem length (cm), relative water content (%), leaf width (mm), number of leaves, leaf length (mm), root length (cm), and number of lateral stems exhibited negative contributions, indicating that excessive vegetative expansion or higher tissue hydration was inversely related to sugar accumulation in stressed plants.

Overall, the sensitivity analysis showed that biomass-related traits such as shoot fresh and dry weight, stem length, and internode length had the greatest positive influence on metabolite accumulation, indicating their key role in stress adaptation and resource allocation. In contrast, traits related to canopy expansion and tissue hydration, including root length, relative water content, and leaf width, generally exhibited negative effects. These findings suggest that compact, biomass-efficient plants tend to accumulate higher levels of secondary metabolites under salinity stress, highlighting the potential of simple morphological measurements for rapid and non-destructive metabolite prediction.

Conclusion

This study compared the performance of two neural network models, multilayer perceptron (MLP) and radial basis function (RBF), in predicting key biochemical metabolites, carotenoids, proline, and soluble sugars, in *Dracocephalum moldavica* under salinity stress conditions. The results showed that the RBF model exhibited superior predictive performance under the tested conditions, as reflected by lower RMSE and MAPE values and higher R^2 scores across all evaluated datasets. These outcomes suggest that the RBF model can effectively represent complex nonlinear relationships between morphological traits and biochemical responses in this species.

The findings also indicate that using easily measurable physiological parameters to predict important biochemical compounds may offer a rapid and cost effective alternative to conventional laboratory analyses. Such modeling approaches could support more efficient monitoring and management of medicinal plants, though their broader agronomic and sustainability implications require further validation.

It is important to note that these conclusions are specific to *D. moldavica* and the experimental conditions applied. Future studies should focus on validating model generalizability using independent external datasets and exploring applicability to other plant species and stress environments. Additionally, integrating hybrid or ensemble models, such as ANN combined with genetic algorithms (ANN-GA) or particle swarm optimization (ANN-PSO), may further enhance predictive accuracy and robustness for practical implementation in precision agriculture and phytochemical research.

Future Research Directions

Given the promising performance of the RBF model in predicting key secondary metabolites in *Dracocephalum moldavica*, future studies should explore its applicability across a broader range of medicinal plants under various environmental stresses. Additionally, comparative assessments involving advanced or hybrid machine learning algorithms such as ensemble methods (e.g., XGBoost, LightGBM) or deep learning approaches could offer further insights into optimizing predictive accuracy. Longitudinal studies incorporating seasonal or developmental variation may also enhance the generalizability and robustness of these models in practical agricultural or pharmaceutical settings.

Declarations

Clinical trial number: not applicable.

Authors' Contributions

Sh Najafian: Writing – original draft, Visualization, Investigation, Data curation. **Shahnaz Fathi and Roya Movlodzadeh:** Resources, Methodology. **Sajad Heidari:** Conceptualization, Formal analysis, Methodology, Software, original draft, Editing. **Shahnaz Fathi** review & editing, Project administration, funding acquisition.

Ethics declarations

Ethical approval and consent to participate.

Not applicable.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Compliance with ethical standards

The presented manuscript represents the honest research work of the authors.

Funding

This research received no external funding.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

References

1. Kavitha K, Rajkumar S, Vijay N, Usha SGA, et al. Role of deep learning and machine learning algorithms in identifying and classifying secondary metabolites in Ashwagandha plant. In: Secondary Metabolites in Medicinal Plants. IGI Global. 2025.
2. Chagas-Paula DA, Oliveira TB, Zhang T, et al. Prediction of anti-inflammatory plants and discovery of their biomarkers by machine learning algorithms and metabolomic studies. *Planta Med.* 2015; 81(16):1451–1460.
3. Tan WH, Tong CY, Chua MX, Derek CJC. Modelling and optimizing secondary metabolites production in *Spirodela polyrhiza* using machine learning. *J Clean Prod.* 2024; [Ahead of print].
4. Khanna A, Jain S, Choudhury T, et al. Predicting secondary metabolites from diverse macronutrients using machine learning and AI. *OPJU Int Conf.* 2024; [Conference proceedings].
5. Bai W, Li C, Li W, et al. Machine learning assists prediction of genes responsible for plant specialized metabolite biosynthesis by integrating multi-omics data. *BMC Genomics.* 2024; 25:123.
6. Mukherjee D, Mukherjee A. Understanding plant secondary metabolism using bioinformatics tools: recent advances and prospects. *Wiley Interdiscip Rev Syst Biol Med.* 2024; [Ahead of print].
7. Poltorak V, Shachaf N, Aharoni A, Zeevi D. Spec2Class: Accurate prediction of plant secondary metabolite class using deep learning. *bioRxiv.* 2024; Preprint.
8. Riedling O, Walker AS, Rokas A. Predicting fungal secondary metabolite activity from biosynthetic gene cluster data using machine learning. *Microbiol Spectr.* 2024; 12(3):e123456.
9. Aghdam SA, Brown AMV. Deep learning approaches for natural product discovery from plant endophytic microbiomes. *Environ Microbiome.* 2021; 16(1):15.
10. Moloudzadeh R, Fathi S, Yari F, Najafian S, Seyedi A. The potential of coated iron nanoparticles for modulating negative effects of salinity stress in Ajowan. *Hortic Environ Biotechnol.* 2024; 65:747-760.
11. Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F. Colorimetric method for the determination of sugars and related substances. *Anal Chem.* 1956; 28(3):350–356.
12. Lichtenthaler HK, Wellburn AR. Determinations of total carotenoids and chlorophylls a and b of leaf extracts in different solvents. *Biochem Soc Trans.* 1983; 11:591-592.

13. Gholamzadeh E, Ghaemi A, Shokri A, Heydari B. Investigation of boiler energy consumption in the gas refinery units using RSM ANN and Aspen HYSYS. *Heliyon*. 2024; e41450.
14. Ghritlehre HK, Prasad RK. Exergetic performance prediction of solar air heater using MLP, GRNN and RBF models of artificial neural network technique. *J Environ Manag*. 2018; 223:566–575.
15. Kruse R, Mostaghim S, Borgelt C, Braune C, Steinbrecher M. Multi-layer perceptrons. In: *Computational Intelligence: A Methodological Introduction*. Cham: Springer International Publishing; 2022: 53-124.
16. Hashemi Fath A, Madanifar F, Abbasi M. Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems. *Petroleum*. 2020; 6:80–91.
17. Fernández-Navarro F, Hervás-Martínez C, Gutiérrez PA, Peña-Barragán JM, López-Granados F. Parameter estimation of q-Gaussian radial basis functions neural networks with a hybrid algorithm for binary classification. *Neurocomputing*. 2012;75(1):123-134.
18. Mehrabi S, Maghsoudloo M, Arabalibeik H, Noormand R, Nozari Y. Application of multilayer perceptron and radial basis function neural networks in differentiating between chronic obstructive pulmonary and congestive heart failure diseases. *Expert Syst Appl*. 2009;36(3):6956-6959.
19. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*. 2017; 40:913–929.
20. White J, Power SD. K-Fold cross-validation can significantly overestimate true classification accuracy in common EEG-based passive BCI experimental designs: An empirical investigation. *Sensors*. 2023; 23(13):6077.
21. Heidari S, Mirzaee-Ghaleh E, Rabbani H, Vesali F. Development of an Android app for estimating the water quality parameters in fish pond. *Environ Sci Pollut Res*. 2021; 28:34501-34510.
22. Yan Z, Zhu X, Wang X, Ye Z, Guo F, Xie L, Zhang G. A multi-energy load prediction of a building using the multi-layer perceptron neural network method with different optimization algorithms. *Energy Explor Exploit*. 2023; 41:273–305.
23. Olden JD, Joy MK, Death RG. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*. 2004; 178:389-397.
24. Garson DG. Interpreting neural network connection weights. *AI Expert*. 1991; 6:47-51.
25. Pizarroso J, Portela J, Muñoz A. NeuralSens: sensitivity analysis of neural networks. *J Stat Softw*. 2022; 102:1-36.
26. Nait Amar M, Jahanbani Ghahfarokhi A, Shang Wui Ng C. Predicting wax deposition using robust machine learning techniques. *Petroleum*. 2022; 8:167-173.
27. Li Y, Chu X, Fu Z, Feng J, Mu W. Shelf-life prediction model of postharvest table grape using optimized radial basis function (RBF) neural network. *Br Food J*. 2019.
28. González-Camacho JM, de Los Campos G, Pérez P, Gianola D, Cairns JE, Mahuku G, Babu R, Crossa J. Genome-enabled prediction of genetic values using radial basis function neural networks. *Theor Appl Genet*. 2012; 125:759-771.
29. Zhang Y, Wang Y. Recent trends of machine learning applied to multi-source data of medicinal plants. *J Pharm Anal*. 2023. 13(5):675-689.
30. Tan WH, Tong CY, Chua MX, Derek CJC. 2024. Modelling and optimizing secondary metabolites production in *Spirodela polyrhiza* using machine learning. *J Clean Prod*. 2024; 478:143986.
31. Jamir L, HP P. 2024. Employing Machine Learning Models to Predict Potential α -Glucosidase Inhibitory Plant Secondary Metabolites Targeting Type-2 Diabetes and Their In Vitro Evaluation. *J Chem Inf Model*. 2024; 64(2):221-233.
32. Zhao Y, Zheng S, Pei J, Yang X. Multiple discriminants preserving support subspace RBFNNs with graph similarity learning. *Inf Sci*. 2023; 619:421-438.
33. Lee WJ, Lee EH. Runoff prediction based on the discharge of pump stations in an urban stream using a modified multi-layer perceptron combined with meta-heuristic optimization. *Water*. 2022; 14(1):99.
34. Zarbakhsh S, Shahsavari AR. Artificial neural network-based model to predict the effect of γ -aminobutyric acid on salinity and drought-responsive morphological traits in pomegranate. *Sci Rep*. 2022