A comparison of artificial intelligence techniques for predicting hyperforin content in *Hypericum perforatum* L. in different ecological habitats

Maryam Saffaria\(^1\) | Ali Jahani\(^2\) | Reza Jahani\(^3\)

\(^1\)College of Natural Resources, University of Tehran, Tehran, Iran
\(^2\)Assessment and Environment Risks Department, Research Center of Environment and Sustainable Development, Tehran, Iran
\(^3\)Department of Pharmacology and Toxicology, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran

**Abstract**

Hyperforin, a major bioactive constituent of *Hypericum* concentration, is impacted by various phenological phases and soil characteristics. We aimed to design a model predicting hyperforin content in *Hypericum perforatum* based on different ecological and phenological conditions. We employed artificial intelligence modeling techniques including multi-layer perceptron (MLP), radial basis function (RBF), and support vector machine (SVM) to examine the factors critical in predicting hyperforin content. We found that the MLP model \((R^2 = 0.9)\) is the most suitable and precise model compared with RBF \((R^2 = 0.81)\) and SVM \((R^2 = 0.74)\) in predicting hyperforin in *H. perforatum* based on ecological conditions, plant growth, and soil features. Moreover, phenological stages, organic carbon, altitude, and total N are detected in sensitivity analysis as the main factors that have a considerable impact on hyperforin content. We also report that the developed graphical user interface would be adaptable for key stakeholders including producers, manufacturers, analytical laboratory managers, and pharmacognosists.

**KEYWORDS**

artificial intelligence, ecological modeling, graphical user interface, hyperforin, *Hypericum perforatum*

---

**1 | INTRODUCTION**

*Hypericum perforatum* L., commonly known as St John’s wort, belongs to the family of Hypericaceae. This family has been studied widely as a source of bioactive compounds utilized as complementary therapeutics for weak to mild depression (Barnes et al., 2019). In addition, numerous preclinical and clinical studies have demonstrated anxiolytic, sedative, nootropic, antischizophrenic, anticonvulsant, anti-inflammatory, antibacterial, and antiviral activities (Galeotti, 2017). *H. perforatum* is one of the most important medicinal plant species of the family Hypericaceae and is listed as a top-selling herbal preparation due to presence of hyperforin, hypericin, and pseudohypericin along with other biologically active metabolites (Christenhusz & Byng, 2016). Hyperforin is a prenylated phloroglucinol derivative that contains a phloroglucinol skeleton with lipophilic isoprene chains (Ramalhete et al., 2016). For many years, it was believed that hypericin is responsible for the antidepressant properties of *H. perforatum*. However, recent clinical trials have demonstrated that hyperforin also contributes to alleviate depression symptoms (Szewczyk et al., 2019). Recently, hyperforin is studied to exhibit various other pharmacological benefits such as anticancer (Zhao et al., 2015), neuroprotective (Gaid et al., 2018), anti-inflammatory (Khan et al., 2019), antiangiogenic (Nabavi et al., 2018), and antibacterial properties. Therefore, accurate identification and
subsequent quantification of hyperforin is imperative for testing these activities and eventually developing a commercial preparation (Barnes et al., 2019). High-performance liquid chromatography (HPLC) with different detectors such as diode array, fluorescence detector, and mass spectrometry has been employed traditionally for routine detection and quantifications of hyperforin (Gitea et al., 2018). These instruments are accurate for identification of secondary metabolites (Saffariha et al., 2020), but they also present some challenges. For instance, complex data analysis procedures require time and trained human resource and consumption of high volume of reagents and solvents (Sahu et al., 2018). Thus, there is a dearth of relatively inexpensive and robust method and/or tools for accurate prediction of hyperforin content in *H. perforatum* applicable to various ecological habitats.

*H. perforatum* can adapt to a range of environmental conditions by changing its metabolic profile (Murch & Saxena, 2006). Various factors including soil composition, genetic, environmental conditions, and phenological stages are reported to impact the content of bioactive compounds in medicinal plants (Cirak & Radusiene, 2019; Radušienė et al., 2012). Similarly, content of hypericin, pseudohypericin, and hyperforin in *Hypericum perforatum* L., *Hypericum montbretii* Spach, and *Hypericum orignifolium* Willd is studied to be impacted significantly by different ecological conditions (Cirak et al., 2007; Cirak et al., 2008; Cirak & Radusiene, 2007; Maggi & Cecchini, 2010). We reported that the highest essential oil yield of *Salvia limbata* L. at flowering stage at the highest altitude (Saffariha et al., 2019). Radušienė et al. (2012) recorded the highest levels of hyperforin at flowering stage using liquid chromatography-mass spectrometry (LC-MS). Alternatively, artificial intelligence techniques are promising in accurately predicting content of secondary metabolites in medicinal plants (Saffariha et al., 2021).

Artificial neural network (ANN) modeling techniques, also known as artificial intelligence techniques, have been designed based on human brain functions using various mathematical algorithms to obtain maximum accuracy in outputs. Support vector machine (SVM), multilayer perceptron (MLP) neural network, and radial basis function (RBF) are the three most frequently used techniques of artificial intelligence in the field of chemical ecology (Kalantary et al., 2020). Recently, these nonlinear techniques have been compared to achieve the most accurate model in the prediction of the ecological process (Aghajani et al., 2014; Jahani & Saffariha, 2020b). For instance, the seed germination percentage of *S. limbata* was predicted under different ecological stresses (Saffariha et al., 2020) by comparing ANN models. ANN-based models can also potentially assist researchers to predict hyperforin content in *H. perforatum* with relatively less resources. These models also need to be validated across various ecological conditions to enhance their applicability. Saraiwa et al. (2018) determined the effects of CO₂ enrichment on the growth and biometal/nutrient content and accumulation in *Senna reticulate*. An ANN accurately predicted results suggesting that Mg, Na, and Fe contents display the most different behavior when comparing plants germinated at atmospheric and elevated CO₂ conditions. Also, Saffariha et al. (2021) measured hypericin content in *H. perforatum* and tested the potential of artificial intelligence techniques to correlate ecological factors with hypericin content. They found the MLP model ($R^2 = .87$) as the most accurate ANN technique in hypericin content prediction, but they believe that application of MLP technique for valid prediction of plants biochemical contents needs to be more explored in other biochemical contents to ensure the results. Rajkovic et al. (2013) used ANN techniques to optimize models for prediction of the sunflower oil transesterification. Authors compared the performances of the models as a decision support system tool during the investigated methanalysis process. The fatty acid methyl ester yield was predicted by ANN model much better than the predictions (±24.2%) obtained by the second-order polynomial equation.

Therefore, the aim of this research was to compare MLP, RBF, and SVM to predict the amount of hyperforin in *H. perforatum* at different growth stages and ecological conditions. The best model among the proposed models determines the most important ecological and phenological factors in the amount of hyperforin in *H. perforatum*. Also using the graphical user interface (GUI) tool will be able researchers to define the amount of hyperforin in *H. perforatum*. Moreover, our findings promoted a commercial consumption of active ingredients in *H. perforatum*.

### 2 | MATERIALS AND METHODS

#### 2.1 | Study area and sampling

The study area is located in the south of Alborz Mountain in Alborz Province in the north of Iran (35° 44'N to 36° 35'N and 51° 00'E to 51° 36'E). This area is considered as most suitable habitats of *H. perforatum* in the range of 1000–4000 m altitude. To include considerable variability in ecological conditions, 100 aerial tissue samples were collected along 15 transects of 1000 m at different altitudes at three phenological stages (vegetative growth, flowering, and seeding) over the period of 6 months. Some soil properties such as electrical conductivity (EC), absorbable phosphor (ppm), acidity (pH), sand (%), silt (%), clay (%), absorbable potassium (ppm), organic carbon (%), and total nitrogen (%) were recorded at each sample location. Other features such as altitude (m), slope (%), and hill aspect (four aspects including 1. North, 2. East, 3. South, and 4. West) for each sample were recorded. Finally, samples were sent to the Herbarium of Medicinal Plants and Drugs Research Institute, Shahid Beheshti University, Tehran, Iran, for further analysis.

#### 2.2 | Chemical and reagents

Hyperforin (Figure 1, 99% purity) was purchased from Alexis Corp. (Lausen, Switzerland). (2-Hydroxypropyl)-β-cyclodextrin and orthophosphoric acid (H₃PO₄) were obtained from Sigma (St. Louis, MO, USA). HPLC grade methanol and acetonitrile were purchased from Merck (Darmstadt, Germany), and pure water was obtained from a Milli-Q water purification system (Millipore, Bedford, MA, USA).
2.3 | Extraction of H. perforatum

Hyperforin extraction was performed using the method described by Soelberg et al. (2007). Briefly, fresh plant leaves were frozen with liquid nitrogen and ground into a very fine powder. Then, 250 mg of the prepared leaf powder was added to 2.5 ml of 80% aqueous methanol containing 0.073 M (2-hydroxypropyl)β-cyclodextrin in a centrifuge tube, and the pH was adjusted to 2.5 with H3PO4. The centrifuge tubes were sonicated for 10 min in an ultrasonic bath (Elma S30H, Germany) before 10-min centrifugation at 5000 rpm. The supernatant was decanted into a 10-ml volumetric flask. This procedure was repeated two times with the pellet, and the obtained supernatants were mixed. Finally, the volume of the collected supernatant was adjusted to the constant volume of 10 ml with methanol, and the sample was filtered through a 0.45-μm PFTE filter (Gelman Sciences, Ann Arbor, MI, USA).

2.4 | HPLC analysis of hyperforin

An Agilent Series 1200 HPLC system (Agilent Corporation, Palo Alto, CA, USA) that was equipped with a G1312A bin pump, a G1379B online degasser, a Mightysil RP-18 GP column (5 μm, 250 × 4.6 mm, Kanto Chemical, Tokyo, Japan) and a G1314B ultraviolet detector (Agilent Corporation, Palo Alto, CA, USA) was used for the analysis of hyperforin in the collected H. perforatum samples. For this purpose, a mixture containing acetonitrile and 0.3% phosphoric acid in water at a ratio of 90:10 was used as the mobile phase. Injection volume, flow rate, column oven temperature, and detector wavelength were set at 20 μl, 1.5 ml/min, 30 C, and 273 nm, respectively (Kuo et al., 2020). The hyperforin content was calculated using the plotted standard calibration curve (concentration vs. peak area) and expressed as mg/g dry mass following triplicate measurement.

2.5 | Modeling process

The current study aimed to study variations in the amount of hyperforin arising due to different landscapes, soils, and phenological stages in Alborz protected area, and current methods are incapable of providing high-precision data. According to studies, ANN models provide more accurate predictions in ecological phenomena studies (Jahani & Rayegani, 2020). Thus, the ANN function in MATLAB 2018 was used to design the structure of three models (MLP, RBF, and SVM).

2.6 | Multilayer perceptron neural network

Neurons are the main elements used in the MLP model (Shams et al., 2020, 2021). To get an accurate model, hidden layers, transfer functions, and neurons must be carefully analyzed. In this paper, input variables include landform, phenological stages, and soil characteristics, and the output variable is the content of hyperforin. The MLP model by weighting variables and summarizing them produced the most accurate output in previous study by Shams et al. (2020, 2021) and Pourmohammad et al. (2020). At first, 60% of samples put in use in the training process. The remaining 40% of samples were divided equally in two data sets (20, 20) for validation and testing (Mosaffaei & Jahani, 2020). Most of studies apply 60–70% of samples for training, and the rest of them are tagged as validation and test data (Cline et al., 2018; Jahani & Rayegani, 2020; Kalantary et al., 2020; Saraiva et al., 2018). On the other hand, the ANN specialist and developer believe that the volume of validation data should be half of training samples (refer to Demuth & Beale, 2002). In MLP training, the weights (w) of the jth variable (x) in jth neuron are defined to calculate the output of jth neuron on the kth hidden layer (netj(k)) by Equation 1.

\[ net^j = \sum_{i=1}^{n} w_{ji} x_i \]  

We considered the output of Equation 1 as the input of a transfer function (f) in Equation 2. Many different functions are tested for the most accurate output (Demuth & Beale, 2002).

\[ Y_{net} = f(net^j) \]  

The accurate weighing of neurons and layers was carefully evaluated, and the most appropriate weights were selected. To accurately calculate hyperforin content in H. perforatum samples, we applied the back propagation method in Equation 3. In Equation 3, E presents the sum of squared errors, wji illustrates the weight of jth neuron in jth hidden layer, and α is the learning rate which is determined by a crisp value (refer to Demuth & Beale, 2002).

\[ w_{ji} = w_{ji}^{-1} + \left( -\alpha \frac{\partial E}{\partial W_{ji}} \right) \]  

2.7 | RBF neural network

Although a set of neurons was used in RBF, in general, RBF acts as a transfer function. In the modeling process, 80% of all samples were used to train the network, and the remaining 20% were allocated for
the accuracy assessment of the network. One of the most applicable RBF functions is Gaussian (Kalantary et al., 2020), and we used this function in our model. The center of circular classifiers, in multi-dimensional space, is measured by Equation 4.

\[ R_j(x) = \exp \left( \frac{-\|x - a_j\|^2}{2\sigma^2} \right) \]  

In Equation 4, \( R_j(x) \) is the RBF, \( \|x - a_j\| \) is the determined Euclidean distance between the total of \( a_j \) (RBF function center), \( x \) = (input vector or variables), and \( \sigma \) = a positive real number.

Finally, to predict hyperforin content, we employed Equation 5.

\[ y_k = \sum_{j=1}^{m} w_j R_j(x) + b_j \]  

In Equation 5, \( w_j \) = the weights of neurons, \( j \) = the number of each node in the hidden layer, \( m \) = the number of neurons, and \( b_j \) = bias (Kalantary et al., 2020).

### 2.8 Support vector machine

SVM is known as a classifier with a set of margins for determining the boundaries of classification which cover the uncertainties in the outputs of the model. In the modeling process, our ultimate goal was to optimize classification boundaries. As we explained in the RBF model, 80% of all samples were allocated to train the network, and the remaining samples were used for the accuracy assessment of the network. The SVM model (Equation 6) uses input variables in the structure of a kernel function (Equation 7).

\[ y(x) = \sum_{i=1}^{n} a_i K(x_i, x_j) + b \]  

There is a kernel function that is explained in Equation 7. The parameters of Equation 7 are \( x_i \) and \( x_j \) = samples and \( \gamma \) = kernel parameter.

\[ K(x_i, x_j) = \exp \left( -\gamma \|x_i - x_j\|^2 \right) \]  

The kernel function parameters are \( x_i \) and \( x_j \) = samples and \( \gamma \) = kernel parameter. By minimizing SVM network errors, we balanced the weight of the network to predict outputs (Equation 8). In Equation 8, the parameters were \( \Sigma \) = training errors, \( 1/2 ||w||^2 \) = the margin, and \( C \) = the tuning parameter.

\[ \frac{1}{2}||w||^2 + C \sum_{i=1}^{n} \xi_i \]  

### 2.9 Accuracy assessment of models

We employed the test data set to analyze the performance of the models. Based on recent research, statistical indicators have been used to assess the accuracy of the model (e.g., Jahani & Saffariha, 2020b). These indicators are MSE (Equation 9), RMSE (Equation 10), MAE (Equation 11), and \( R^2 \) (Equation 12). In these equations, \( y_i \) and \( \bar{y}_i \) = the targets and network outputs, respectively, \( \bar{y}_i \) = the mean of target values, and \( N \) = the number of samples.

\[ \text{MSE} = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n} \]  

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n}} \]  

\[ \text{MAE} = \frac{\sum_{i=1}^{n}|y_i - \bar{y}_i|}{n} \]  

\[ R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \]  

### 2.10 Sensitivity analysis

Different values of input variables in three developed models have a linear or logarithmic effect on the output. To prioritize the variables affecting the content of hyperforin in H. perforatum, we conducted a sensitivity analysis. In this method, each variable has its data set, and the range of standard deviation was variable. We changed each variable in the specified range, and other variables were fixed (equal to average). The sensitivity value for each variable was calculated. This value is equal to the standard deviation of outputs in response to variable changes.

### 2.11 GUI tool

A GUI as an EDSS tool was designed to determine the most precise model to predict the content of hyperforin in H. perforatum. In this tool, by applying ecological conditions of the site, landform, soil characteristics, and plant phenological stages, the content of hyperforin in H. perforatum was predicted.

### 3 RESULTS

#### 3.1 MLP performance

We used three datasets to obtain the best results from the MLP model. To predict the most precise MLP model, the number of neurons, hidden layers, training method, and activation function has been modified. Table 1 illustrates that the structure 13-8-1 has
been considered as the most authentic in the prediction of hyperforin content in *H. perforatum*. The factors used for the MLP model include 13 inputs, 8 neurons in the hidden layer, and 1 neuron as output.

We applied the scatter plot to create a correlation between the target and MLP output (Jahani & Saffariha, 2020b). In Figure 2, the scatter plot of MLP outputs versus targets values of the hyperforin content for training, validation, test, and total data is shown. According to coefficient ($R^2$), there is a considerable correlation between MLP outputs and target values.

There is a comparison between the real (target) and simulated (output) values of MLP in the datasets shown in Figure 3.

### 3.2 RBF performance

In the process of training data, there are factors such as spread and neurons that needed to be optimized. In the training phase, we seek to reduce network error through spread values. Therefore, the number of neurons was 13, and the spread of RBF was 6.4 to attain the best performance. In Table 2, the results of RBF in training and test datasets are displayed. The best structure for the value of $R^2$ in training and test datasets is listed in Table 2. The structure of optimized RBF is 13-13-1 with 13 variables as inputs, 13 neurons in the hidden layer with Gaussian transfer function, and 1 neuron of output.

---

**TABLE 1** The results of parameters tuning in multilayer perceptron (MLP) structure

| Activation function | Training function | Structure | Test set | Training data |
|---------------------|-------------------|-----------|----------|---------------|
|                     |                   |           | $R^2$ | MSE | RMSE | MAE | $R^2$ | MSE | RMSE | MAE |
| Logsig-Purelin      | LM                | 13-8-1    | .90    | .14  | .37  | .32  | .91    | .10  | .01  | .26  |

**FIGURE 2** Scatter plots of targets values of hyperforin content generated through multilayer perceptron (MLP) outputs
A scatter plot of RBF outputs versus targets values of hyperforin content for training, test, and total data is shown in Figure 4.

Comparison of the real (target) and simulated (output) values of RBF in the datasets is shown in Figure 5.

3.3 | SVM performance

The kernel function, defined in SVM structure, classifies the data into a matrix of multidimensional space. The \( C \) parameter, epsilon (\( \varepsilon \)), and parameter gamma (\( \gamma \)) are key factors that influence SVM regression performance. We employed the value of the parameter \( \varepsilon \) to specify the number of support vectors as described by Laref et al. (2018). SVM regression with the Gaussian function includes \( \gamma \) parameter to define the width of bell-shaped curves. In this study, the linear function resulted in a more accurate prediction so the Gaussian function and \( \gamma \) parameter

| TABLE 2 | The RBF structure results (spread and neurons) in training and test data sets |
|----------------|------------------|------------------|------------------|------------------|------------------|
| Model          | Spread | Neurons | Test data | Training data |
| RBF            | 6.4    | 13      | \( R^2 \) | \( R^2 \) |
|                |        |         | MSE      | MSE        |
|                |        |         | RMSE     | RMSE       |
|                |        |         | MAE      | MAE        |
|                |        |         | .81      | .88        |
|                |        |         | 0.15     | 0.12       |
|                |        |         | 0.39     | 0.35       |
|                |        |         | 0.34     | 0.29       |

Abbreviation: RBF, radial basis function.
were later eliminated. In the SVM model, we looked for simple curves where the value of the parameter $C$ helps to achieve these curves. To acquire hyperforin content, $C$ and $\varepsilon$ factors in SVM regression are defined, and the most important SVM factors are described in Table 3.

Based on the values of $R^2$ in training and test datasets, the best $\varepsilon$ value was 0.004, and $C$ value was 922.6. The other models having various $\varepsilon$ and $C$ displayed over-fitting and under-fitting in models. The scatter plot of SVM outputs via target values of the hyperforin for training, test, and total data are shown in Figure 6. The values of coefficient ($R^2$) suggest the correlation between the SVM outputs and target values.

### Table 3

| $\varepsilon$ | $C$ | Test set | Training data |
|---------------|-----|----------|---------------|
|               |     | $R^2$    | MSE | RMSE | MAE | $R^2$ | MSE | RMSE | MAE |
| 0.004         | 922.6 | .74     | 4.46 | 2.11 | 1.71 | .76   | 3.43 | 1.85 | 1.29 |

Abbreviation: SVM, support vector machine.

Comparison between the simulated (output) values and the real target of SVM in datasets is shown in Figure 7.

### 3.4 Model selection

As we compared the output of the MLP, RBF, and SVM models in Figure 8, it is clearly demonstrated that the MLP model is the most suitable model measured by the highest value of $R^2$ in training, test, and all datasets.
The MLP model sensitivities for input variables are displayed in Figure 9. The standard deviations of MLP outputs for the content of hyperforin are shown in Figure 9. The most critical inputs that have an impact on MLP output are phenological stages, organic carbon, altitude, and total nitrogen, respectively.

Figure 10 displays that the amount of hyperforin in *H. perforatum* increases by increasing phenological stages, organic carbon, altitude, and total nitrogen, but the opposite trend was observed for the slope variable.

To determine the content of hyperforin in *H. perforatum* in other areas, a GUI was designed to assist robust prediction of the hyperforin content in varied habitat conditions. The EDSS tool will be operated by pressing the “hyperforin content prediction” function in the GUI tool.

### 4 | DISCUSSION

Several factors drive the spatiotemporal distribution of plant’s chemical footprints, and it is almost impossible to single out a particular factor solely responsible of observed changes in composition of metabolites (Radušienė et al., 2012). The current study evaluated the effect of phenological stages, landform characteristics, and soil features on hyperforin content in *H. perforatum*. Our findings suggest that hyperforin content was significantly dependent on phenological stages, organic carbon, altitude, and total N. Out results are in agreement to another study by Ruzalepp et al., 2017 comparing chemical composition of secondary metabolites was impacted in *H. perforatum* and *Hypericum maculatum* by various phenological stages, organic carbon levels, and altitude. Results of our study are important to understand biosynthesis of secondary metabolites with higher medicinal value including hyperforin under a range of ecological habitats.

To the best of our knowledge, this study is the first report to predict hyperforin content in *H. perforatum* using artificial intelligence techniques. Recently, the use of the artificial intelligence techniques remained a subject of active research (Cline et al., 2018; Jahani et al., 2011; Rawlins et al., 2012). This technique not only is applicable in predicting active compounds in *H. perforatum* but also could be useful for other plants with various environmental conditions. We compared three models (MLP, SVM, and RBF) and revealed that the MLP as an EDSS tool outperformed in predicting hyperforin content more accurately ($R^2 = .9$) compared with RBF ($R^2 = .81$) and SVM ($R^2 = .74$) under plant growth, landform conditions, and soil features considered in this study. Because identification and quantification of hyperforin are expensive and require tandem mass spectrometry and skillsets, the ANN modeling could potentially serve as an alternative approach (Eftekhari et al., 2018). ANN is a mathematical approach to obtain the closest result to the expected value (Suárez et al., 2015). These models have been recently compared in another study looking at aesthetic quality prediction and vegetation density (Jahani & Saffariha, 2020a). In addition, Savić et al. (2013) reported that MLP, along with central composite design (CCD), was the most suitable model in predicting the total flavonoid extraction from green tea. Moreover, the ANN modeling approach has several advantages over...
regression modeling as discussed in detail previously (Jamshidi et al., 2016).

We demonstrated that the most influential factors impacting hyperforin content were phenological stages, organic carbon, altitude, and total N as measured by sensitivity analysis. This observation is in agreement with several other reports studying influence of altitude, growth stages, and genotype on hyperforin content (Büter et al., 2008; Filippini et al., 2010; Xenophontos et al., 2008; Zobayed et al., 2006). Similarly, other factors including nitrogen availability, phenologic stage, drought stress, altitude, and soil feature are also reported to impact hyperforin content (Cirak & Radusiene, 2019; Murch et al., 2003). We observed a positive correlation between organic carbon, altitude, total N, and the hyperforin content. Bozin et al. (2013) and Gioti et al. (2009) reported the highest concentration...
of hyperforin at floral budding stage, which was later confirmed by Kladar et al. (2017) reporting the highest level of active ingredients in \textit{H. perforatum} between floral budding and flowering stage. However, other reports provided evidence of highest concentration at ripening stage (Cirak et al., 2008) and fruit development (Filippini et al., 2010). This discrimination in the literature could be explained by instability and sensitivity of hyperforin to light (Zidorn, 2010). Yesaghi (2006) cal can potentially benefit for these models.

The recent emphasis on plant secondary metabolite and their use in pharmaceutical, medicinal and food industry require study of ecological factors for maximum yield of the active ingredients. It also requires complementary approaches for accurate prediction of bioactive constituents. Our study compared prediction of hyperforin content with the help of three models and profoundly suggests that the MLP was the most accurate model in predicting hyperforin content defined by MATLAB 2018 software. Furthermore, we observed a positive correlation between phenological stages, organic carbon, altitude, and total N with hyperforin content. Developing such approaches can greatly reduce the cost and resources required for traditional analytical platforms, and various industries such as pharmaceutical and agrochemical can potentially benefit for these models.

ACKNOWLEDGMENTS
Not applicable.

CONFLICT OF INTEREST
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.

AUTHOR CONTRIBUTIONS
Data gathering, Data analysis, Article writing: M. S.; Data analyzing, Method curation, Article editing: A. J.; Lab analysis, Visualization (if applicable): R. J.

DATA AVAILABILITY STATEMENT
Data will be available by authors in reasonable request.

ORCID
Maryam Saffariha https://orcid.org/0000-0003-1981-3389
Ali Jahani https://orcid.org/0000-0003-4965-3291
Reza Jahani https://orcid.org/0000-0002-5058-1283

REFERENCES
Aghajani, H., Marvie Mohadjer, M. R., Jahani, A., Asef, M. R., Shirvany, A., & Azaryan, M. (2014). Investigation of affective habitat factors affecting on abundance of wood macrofungi and sensitivity analysis using the artificial neural network. \textit{Iranian Journal of Forest and Poplar Research}, 21(4), 9–19.

Barnes, J., Arnason, J. T., & Roufogalis, B. D. (2019). \textit{St John’s wort} (\textit{Hypericum perforatum} L.): Botanical, chemical, pharmacological and clinical advances. \textit{The Journal of Pharmacy and Pharmacology}, 71, 1–3. https://doi.org/10.1111/jphp.13053

Bozin, B., Kladar, N., & Grujić, N. (2013). Impact of origin and biological source on chemical composition, anticholinesterase and antioxidant properties of some St. John’s wort specie (\textit{Hypericum} spp., \textit{Hypericaceae}) from the central Balkans. \textit{Molecules}, 18(10), 11733–11750. https://doi.org/10.3390/molecules181011733

Büter, B., & Büter, B. (2008). Ontogenetic variation regarding hypericin and hyperforin levels in four accessions of \textit{Hypericum perforatum} L. \textit{Journal of Herbs, Spices & Medicinal Plants}, 9(2–3), 95–100.

Camas, N., Radusiene, J., Ivanauskas, L., Jakstas, V., & Cirak, C. (2014). Altitudinal changes in the content of bioactive substances in \textit{Hypericum orientale} and \textit{Hypericum pallens}. \textit{Acta Physiologiae Plantarum}, 36, 675–686. https://doi.org/10.1007/s11738-013-1446-z

Christenhusz, M. J. M., & Byng, J. W. (2016). The number of known plants species in the world and its annual increase. \textit{Phytotaxa}, 261(3), 201–217. https://doi.org/10.11646/phytotaxa.261.3.1

Cirak, C., & Radusiene, J. (2007). Variation of hyperforin in \textit{Hypericum montbretii} during its phenological cycle. \textit{Natural Product Research}, 21, 1151–1156. https://doi.org/10.1080/14787447.2015.1058758

Cirak, C., & Radusiene, J. (2019). Factors affecting the variation of bioactive compounds in \textit{Hypericum} species. \textit{Biologia Futura}, 70, 1–15. https://doi.org/10.1556/019.70.2019.25

Cirak, C., Radusiene, J., Janulis, V., & Ivanauskas, L. (2007). Secondary metabolites in \textit{Hypericum perforatum}: Variation among plant parts and phenological stages. \textit{Botanica Helvetica}, 117, 29–36. https://doi.org/10.10.30035-007-0777-z

Cirak, C., Radusiene, J., Janulis, V., & Ivanauskas, L. (2008). Pseudohypericin and hyperforin in \textit{Hypericum perforatum} from Northern Turkey: Variation among populations, plant parts and

5 | CONCLUSION
The recent emphasis on plant secondary metabolite and their use in pharmaceutical, medicinal and food industry require study of ecological factors for maximum yield of the active ingredients. It also requires complementary approaches for accurate prediction of bioactive constituents. Our study compared prediction of hyperforin content with the help of three models and profoundly suggests that the MLP was the most accurate model in predicting hyperforin content defined by MATLAB 2018 software. Furthermore, we observed a positive correlation between phenological stages, organic carbon, altitude, and total N with hyperforin content. Developing such approaches can greatly reduce the cost and resources required for traditional analytical platforms, and various industries such as pharmaceutical and agrochemical can potentially benefit for these models.
phenomenological stages. *Journal of Integrative Plant Biology*, 50(5), 575–580. https://doi.org/10.1111/j.1744-7909.2008.00657.x

Cline, L., Roundy, B., & Christensen, F. (2018). Using germination prediction to inform seeding potential: I. Temperature range validation of germination prediction models for the Great Basin, USA. *Journal of Arid Environments*, 150, 71–81. https://doi.org/10.1016/j.jaridenv.2017.11.018

Demasi, S., Caser, M., Lonati, M., Cioni, P. L., Pistelli, L., Najar, B., & Scarici, V. (2018). Latitude and altitude influence secondary metabolism production in peripheral Alpine populations of the Mediterranean species *Lavandula angustifolia* Mill. *Frontiers in Plant Science*, 9, 983–993. https://doi.org/10.3389/fpls.2018.00983

Demuth, H., & Beale, M. (2002). *Neural network toolbox for use with MATLAB®* (4th ed., pp. 1–840). MathWorks, Inc.

Eftekhar, M., Yadollahi, A., & Ahmad, H. (2018). Development of an artificial neural network as a tool for predicting the targeted phenolic profile of grapevine (*Vitis vinifera*) foliar wastes. *Frontiers in Plant Science*, 9, 837. https://doi.org/10.3389/fpls.2018.00837

Filippini, R., Plovan, A., Borsari, A., & Caniato, R. (2010). Study of dynamic accumulation of secondary metabolites in three subspecies of *Hypericum perforatum*. *Fitoterapia*, 81, 115–119. https://doi.org/10.1016/j.fitote.2009.08.002

Gaid, M., Biedermann, E., Fuller, J., & Haas, P. (2018). Biotechnological production of hyperforin for pharmaceutical formulation. *European Journal of Pharmaceutics and Biopharmaceutics*, 126, 10–26. https://doi.org/10.1016/j.ejpb.2017.03.024

Galeotti, N. (2017). *Hypericum perforatum* (St John’s wort) beyond depression: A therapeutic perspective for pain conditions. *Journal of Ethnopharmacology*, 200, 136–146. https://doi.org/10.1016/j.jep.2017.02.016

Giotti, E. M., Fiamegos, Y. C., Skalkos, D. C., & Stalikas, C. D. (2009). Antioxidant activity and bioactive components of the aerial parts of *Hypericum perforatum* L. from Epirus, Greece. *Food Chemistry*, 117(3), 398–404.

Gitea, D., Vicas, S., Gitea, M. A., Nemeth, S., Tit, D. M., & Pasca, B. (2018). HPLC screening of bioactive compounds and antioxidant capacity of different hypericum species. *Revista de Chimie*, 69(2), 305–309. https://doi.org/10.37358/RC.18.2.6095

Jahani, A., Etemad, V., Doolati, M., & Avani, N. (2011). Effects of planting depth and planting density on height growth, basal diameter and via-tality of *Haloxylon persicum* L. from Epirus, Greece. *Revista de Chimie*, 69(2), 305–309. https://doi.org/10.37358/RC.18.2.6095

Jahani, A., Rayegani, B. (2020). Forest landscape visual quality evaluation with artificial intelligence techniques as a decision support system. *Stochastic Environmental Research and Risk Assessment*, 34(10), 1473–1486. https://doi.org/10.1007/s00477-020-01832-x

Jahani, A., & Saffariha, M. (2020a). Aesthetic preference and mental restoration prediction in urban parks: An application of environmental modeling approach. *Urban Forestry & Urban Greening*, 54, 126775. https://doi.org/10.1016/j.ufug.2020.126775

Jahani, A., & Saffariha, M. (2020b). Human activities impact prediction in vegetation diversity of Lar National Park in Iran using artificial neural network model. *Journal of Integrated Environmental Assessment and Management*, 17(1), 42–52. https://doi.org/10.1002/ieam.4349

Jamshidi, S., Yadollahi, A., Ahmadi, H., Arab, M. M., & Eftekhar, M. (2016). Predicting in vitro culture medium macro-nutrients composition for pear rootstocks using regression analysis and neural network models. *Frontiers in Plant Science*, 7, 274–285.

Kalantary, S., Jahani, A., & Jahani, R. (2020). MLR and ANN approaches for prediction of synthetic/natural nanofibers diameter in the environmental and medical applications. *Journal of Scientific Reports*, 10, 8117. https://doi.org/10.1038/s41598-020-65121-x

Khan, H., Sureda, A., Belwal, T., Çetinkaya, S., Sünatar, I., Tejada, S., Devkota, H. P., Ullah, H., & Aschner, M. (2019). Polyphenols in the treatment of autoimmune diseases. *Autoimmunity Reviews*, 18, 647–665. https://doi.org/10.1016/j.autrev.2019.05.001

Kladar, N., Mršanović, J., Anačkova, G., Šolajić, S., Gavarić, N., Srđenović, B., & Božin, B. (2017). *Hypericum perforatum*: Synthesis of active principles during flowering and fruitation—Novel aspects of biological potential. *Evidence-based Complementary and Alternative Medicine*, 2017, 2865610. https://doi.org/10.1155/2017/2865610

Kleemann, B., Loos, B., Scriba, T., Lang, D., & David, S. (2014). St John’s wort (*Hypericum perforatum L*) photomedicine: Hypericin-photodynamic therapy induces metastatic melanoma cell death. *PLoS ONE*, 9(7), 103762.

Kuo, C., Chou, Y., Liao, K., Shieh, C., & Deng, T. (2020). Optimization of light intensity, temperature, and nutrients to enhance the bioactive content of hyperforin and rutin in St John’s wort. *Molecules*, 25(18), 4256.

Laref, R., Losson, E., Sava, A., & Siadat, M. (2018). On the optimization of the support vector machine regression hyperparameters setting for gas sensors array applications. *Chemometrics and Intelligent Laboratory Systems*, 184, 22–27.

Maggi, F., & Cecchini, C. (2010). Chemical composition and antimicrobial activity of the essential oils from several *Hypericum Taxa* growing in central Italy. *Chemistry & Biodiversity*, 7, 447–466.

Mir, M., Hamid, S., Kamili, A., & Hassan, O. (2019). Sneak peek of *Hypericum perforatum* L: Phytochemistry, phytochemical efficacy and bio-technological interventions. *Journal of Plant Biochemistry and Biotechnology*, 28, 357–373. https://doi.org/10.1007/s13562-019-00490-7

Mosaffeia, Z., & Jahani, A. (2020). Modeling of ash (Fraxinus excelsior) bark thickness in urban forests using artificial neural network (ANN) and regression models. *Modeling Earth Systems and Environment*, 7, 1443–1452. https://doi.org/10.1007/s40808-020-00869-9

Murch, S. J., Haq, K., Vasantha Rupasinghe, H. P., & Saxena, P. K. (2003). Nickel contamination affects growth and secondary metabolite composition of St John’s wort (*Hypericum perforatum L*). *Environmental and Experimental Botany*, 49, 251–257. https://doi.org/10.1016/S0098-8742(02)00090-4

Murch, S. J., & Saxena, P. K. (2006). A melatonin-rich germplasm line of St John’s wort (*Hypericum perforatum L*). *Journal of Pineal Research*, 41, 284–287.

Nabavi, S. M., Nabavi, S. F., & Sureda, A. (2018). The water extract of Tutsan (*Hypericum androsaemum*), red berries exerts antidepressive like effect and in vivo antioxidant activity in a mouse model of post-stroke depression. *Biomedicine & Pharmacotherapy*, 99, 290–298. https://doi.org/10.1016/j.biopha.2018.01.073

Odbas, M. S., Radusiene, J., Camas, N., Janulis, V., Ivanauskas, L., & Cirak, C. (2009). The quantitative effects of temperature and light intensity on hyperforin and hypericins accumulation in *Hypericum perforatum L*. *Journal of Medicinal Plant Research*, 3(7), 519–525.

Pourmohammad, P., Jahani, A., Zare Chahooki, M. A., & Goshhtasb Meigooni, H. (2020). Road impact assessment modeling on plant diversity in national parks using regression analysis in comparison with artificial intelligence. *Modeling Earth Systems and Environment*, 6(3), 1281–1292. https://doi.org/10.1007/s40808-020-00799-6

Radusiene, J., Karpaviciene, K., & Stanisius, Z. (2012). Effect of external and internal factors on secondary metabolites accumulation in St John’S worth. *Botanica Lithuanica*, 18(2), 101–108. https://doi.org/10.2478/v10279-012-0012-8

Rajkovic, K. M., Avramovic, J. M., Milic, P. S., Stamenkovic, O. S., & Veljkovic, V. B. (2013). Optimization of ultrasound-assisted base-catalyzed methanolation of sunflower oil using response surface and artificial neural network methodologies. *Chemical Engineering Journal*, 215–216, 82–89.

Ramalheira, N., Machado, A., & Serrano, R. (2016). Comparative study on the in vivo antidepressant activities of the Portuguese *Hypericum foliosum*, *Hypericum androsaemum* and *Hypericum perforatum*.
medicinal plants. *Industrial Crops and Products*, 82, 29–36. https://doi.org/10.1016/j.indcrop.2015.12.014
Rawlins, K., Roundy, B., Davis, S., & Eggett, D. (2012). Predicting germination in semi-arid wildland seedbeds. I. Thermal germination models. *Environmental and Experimental Botany*, 76, 60–67. https://doi.org/10.1016/j.envexpbot.2011.10.003
Ruzaleza, L., Raal, A., & Pussa, T. (2017). Comparison of chemical composition of *Hypericum perforatum* and *H. maculatum* in Estonia. *Biochemical Systematics and Ecology*, 73, 41–46. https://doi.org/10.1016/j.bse.2017.06.004
Saffariha, M., Azarnivand, H., Zare, M., Tavili, A., Nejadebrahimi, S., & Potter, D. (2019). The effect of flowering on the quality and quantity of essential oils of *Salvia limbata* in Taleghan. *Journal of Rangeland and Watershed Management*, 72, 139–149.
Saffariha, M., Jahani, A., Jahani, R., & Latif, S. (2021). Prediction of hypericin content in *Hypericum perforatum* L. in different ecological habitat using artificial neural networks. *Journal of Plant Methods*, 17, 10. https://doi.org/10.1186/s13007-021-00710-z
Saffariha, M., Jahani, A., & Potter, D. (2020). Seed germination prediction of *Salvia limbata* under ecological stresses in protected areas: An artificial intelligence modeling approach. *BMC Ecology*, 20, 48. https://doi.org/10.1186/s12898-020-00316-4
Sahu, P. K., Ramisetty, N. R., Cecchi, T., Swain, S., Patro, C. S., & Panda, J. (2018). An overview of experimental designs in HPLC method development and validation. *Journal of Pharmaceutical and Biomedical Analysis*, 147, 590–611. https://doi.org/10.1016/j.jpba.2017.05.006
Saraiva, A. C. F., Mesquita, A., de Oliveira, T. F., & Hauser-Davis, R. A. (2018). High CO2 effects on growth and biometal contents in the pioneer species *Senna reticulata*: Climate change predictions. *Journal of Trace Elements in Medicine and Biology*, 50, 130–138. https://doi.org/10.1016/j.jtemb.2018.06.018
Savić, I., Nikolić, L., Stanković, M., & Moder, K. (2013). Optimization of total flavonoid compound extraction from *Camellia sinensis* using the artificial neural network and response surface methodology. *Chemical Industry*, 67, 249–259. https://doi.org/10.2298/HEMIND120313066S
Seyis, F., Yurteri, E., Ozcan, A., & Cirak, C. (2020). Alloidal impacts on chemical content and composition of *Hypericum perforatum*, a prominent medicinal herb. *South African Journal of Botany*, 135, 391–403.
Shams, S. R., Jahani, A., Kalantary, S., Moiaddini, M., & Khorasani, N. (2021). The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO2 concentration. *Journal of Urban Climate*, 37, 100837. https://doi.org/10.1016/j.jucim.2021.100837
Shams, S. R., Jahani, A., Moiaddini, M., Khorasani, N., & Kalantary, S. (2020). Prediction of air carbon monoxide in Tehran air using intelligence data approach. *Journal of Health and Safety at Work* *Health and Safety at Work*, 10(4), 385–396.
Soelberg, J., Jorgensen, L. B., & Jager, A. K. (2007). Hyperforin accumulates in the translucent glands of *Hypericum perforatum*. *Annals of Botany*, 99(6), 1097–1100. https://doi.org/10.1093/aob/mcm057
Suárez, M. H., Dopezo, G. A., López, D. L., & Espinosa, F. (2015). Identification of relevant phytochemical constituents for characterization and authentication of tomatoes by general linear model linked to automatic interaction detection (GLM-AID) and artificial neural network models (ANNs). *PLoS ONE*, 10, e0128566.
Szewczyk, B., Pochwat, B., Muszyńska, B., Opoka, W., Krakowska, A., Rafalo-Ulińska, A., Friedland, K., & Nowak, G. (2019). Antidepressant-like activity of hyperforin and changes in BDNF and zinc levels in mice exposed to chronic unpredictable mild stress. *Behavioural Brain Research*, 372, 112045. https://doi.org/10.1016/j.bbr.2019.112045
Walker, L., Sirvant, T., Gibson, D., & Vance, N. (2001). Regional differences in hypericin and pseudohypericin concentrations and five morphological traits among *Hypericum perforatum* plants in the northwestern United States. *Canadian Journal of Botany*, 79(10), 1248–1255.
Xenophonatos, M., Stavropoulou, S., Avramakis, E., Navakoudis, E., Dörnemann, D., & Kotzabasis, K. (2008). Influence of the habitat altitude on the (proto) hypericin and (proto) pseudohypericin levels of *Hypericum* plants from Crete. *Planta Medica*, 74(12), 1496–1503. https://doi.org/10.1055/s-2008-1081337
Yang, S., Li, J., Zhao, Y., Chen, B., & Fu, C. (2011). Harpagoside variation is positively correlated with temperature in *Scrophularia ningpoensis* Hems. *Journal of Agricultural and Food Chemistry*, 59, 1612–1621. https://doi.org/10.1021/jf1014702u
Yesaghi, S. (2006). An investigation of ecophysiological, phenological, ethnological, and antibacterial needs of two widely-grown species of the herbs, *Hypericum perforatum* in Golestan province. MS Thesis, Islamic Azad University.
Zhao, J., Liu, W., & Wang, J. (2015). Recent advances regarding constituents and bioactivities of plant from the genus *Hypericum*. *Chemistry & Biodiversity*, 12, 309–349. https://doi.org/10.1002/cbdv.201300304
Zidorn, C. (2010). Alloidal variation of secondary metabolites in flowering heads of the Asteraceae, trends and causes. *Phytochemistry Reviews*, 9, 197–203. https://doi.org/10.1007/s11101-009-9143-7
Zobayed, S. M. A., Afreen, F., Goto, E., & Kozi, T. (2006). Plant–environment interactions: Accumulation of hypericin in dark glands of *Hypericum perforatum*. *Annals of Botany*, 98, 793–804. https://doi.org/10.1093/aob/mcl169

How to cite this article: Saffariha, M., Jahani, A., & Jahani, R. (2021). A comparison of artificial intelligence techniques for predicting hyperforin content in *Hypericum perforatum* L. in different ecological habitats. *Plant Direct*, 5(11), e363. https://doi.org/10.1002/pld3.363