

## A SOFT COMPUTING QSAR ADAPTED MODEL FOR IMPROVEMENT OF GOLDEN ROOT *IN VITRO* CULTURE GROWTH

Valeriya Simeonova<sup>1</sup>, Krasimira Tasheva<sup>2</sup>, Georgina Kosturkova<sup>2</sup>, Dimitar Vasilev<sup>3</sup>

<sup>1</sup>Sofia University "St. Kliment Ohridski", Faculty of Mathematics and Informatics, Sofia, Bulgaria

<sup>2</sup>Bulgarian Academy of Sciences, Institute of Plant Physiology and Genetics, Sofia, Bulgaria

<sup>3</sup>AgroBioInstitute, Bioinformatics Group, Sofia, Bulgaria

Correspondence to: Valeriya Simeonova

E-mail: v.n.simeonova@me.com

### ABSTRACT

*Golden Root is a rare medicine plant, difficult for cultivation, with a wide potential in treating cardiovascular and other diseases because of its biologically active compounds. Therefore there are researches about in vitro cultivation of Rhodiola rosea. It is well known that biotechnological experiments provide slowly and in a non-efficient way the needed protocols about multiplication, growing and rooting in in vitro nutrient media. The previously collected data of such experiments were analyzed and computationally trained in order to identify the nutrient media that give the best results for growing and rooting, taking into account the limitations such as insufficiency of the plant material and in vitro nutrient media cost.*

*The proposed analysis contributes to the optimization of biotechnology experiments giving new directions for the theoretically fitting quantity of nutrient medium ingredients necessary for in vitro growth and rooting of Golden root, taking into account criteria such as biology and cost effectiveness concerning the experiment and results.*

*"Similar biological activity is provided by common structural properties" – this is the main concept of modeling by Quantitative structure–activity relationship model (QSAR), which is a common method used for property prediction of biochemistry molecules and drugs. Based on this we adapted the concept, making the following assumptions: "biological activity" can be every result from an in vitro experiment such as the type of developed tissue and percentage of necrotic tissues; "structural properties" can be the properties of the nutrient medium, i.e. the concentrations of chemical compounds in the medium. QSAR could be defined as a method based either on regression models, or on artificial intelligence models. Here Artificial Neural Networks is proposed as a soft computing method for creating a model of dependences between nutrient medium, type of explant, type of nutrient medium, cultivation days, medium price and the initial response during the in vitro cultivation. Forecast results were subjected to applied basic statistical analysis, clustering and graphical interpretations.*

Biotechnol. & Biotechnol. Eq. 2013, **27**(3), 3877-3884

**Keywords:** QSAR, AINN, multiplication coefficient, *Rhodiola rosea*, *in vitro* development, soft computing

**Abbreviations:** ABA: adapted biological activities; AC ratio: cytokinin /auxin ratio; AI: artificial intelligence; ANN: artificial neural network; AQSAR: adaptation of QSAR; ASP: adapted structural properties; BAP: N6-benzylaminopurine; CM: culture medium/a; 2,4-D: 2,4-dichlorophenoxyacetic acid; GA<sub>3</sub>: gibberellic acid; IAA: indolyl-3-acetic acid; 2-iP: 6-(γ,γ-dimethylallyl amino) purine; IBA: indole 3-butyric acid; IPM: initial plant medium/a; MC: media combination; NAA: α-naphthyl acetic acid; TDZ: thidiazuron; QSAR: Quantitative structure–activity relationship

### Introduction

Golden root is an endangered plant which is protected by law in Bulgaria and some other countries (UK, Finland, Russia, Mongolia). The extracts from the roots and rhizomes are used in prevention and treatment of socially important diseases of the cardiovascular and central nervous system, cancer and others. Difficult seed propagation (6) as well as intensive and unregulated harvesting/collection of the plants from their

natural habitat in many countries leads to disappearance of the species from its natural areas of existence (1).

Protocols for *in vitro* plant organogenesis, regeneration and propagation using plants growing in nature were established previously (8). *In vitro* cultures with different characteristics were obtained using a large number of nutrient media and explants, allowing flexibility in research. These schemes varied in length, time execution, laboratory chemicals and consumables as well as in their cost. The suitable concentration and combination of the plant growth regulators depends on the genotype, the ecotype, the explant type and its stage of development. The obtained results did not need to be further improved experimentally in order to establish efficient schemes for mass propagation and callusogenesis from the wild growing Golden root plants.

By definition the Quantity Structure–Activity Relationship (QSAR) analysis is a set of soft computing and statistical tools and approaches, applied to data from pharmacology, molecular biology, organic and quantum chemistry (2). QSAR is based on the hypothesis that similar biological activity is provided by common structural properties. It aims quantitative characterization of the relationship between the chemical

TABLE 1

Architecture and parameters of ANN for Stage 1 &amp; 2

	Characteristics of ANN	Stage 1	Stage 2
<b>Base statistics for ANN</b>	Learning cycles	6351	887
	Training error	0.00002	0.000065
	Input columns	15	36
	Output columns	6	10
	Excluded columns	4	7
	Training example rows	23	7
	Querying example rows	2415	2431
	Duplicated example rows	0	0
<b>Architecture of the ANN</b>	Input nodes connected	15	36
	Hidden layer 1 nodes	16	20
	Hidden layer 2 nodes	10	8
	Output nodes	6	10
<b>Parameters of the ANN</b>	Learning rate	0.6	0.6
	Momentum	0.8	0.8
	Target error	0.0001	0.0001

TABLE 2

Structure of the training tables

Used at Stage No.	Column name	Signal type	Data type	Filtered columns by Easy NN Plus, because of less data	Field type
1 & 2	IPM Type, IPM.AC ratio, IPM Value	Input	I,R,R	Included	Calculated field
1 & 2	Sucrose, Agar, Thidiazuron, GA3, IBA	Input	I,I,R,R,R	Excluded	Empirical data
1 & 2	Glutamine, Casein, Zeatin, BAP, Kinetin, 2-iP, IAA, NAA, 2,4-D, Explant	Input	I,I,R,R,R,R,R,R,Text	Included	Empirical data
1 & 2	Callus, Compact, Green, Leaf rosette, Plants, Cultivation days, Percentage	Output/ Input	B,B,B,B,I,R	Included	Empirical data
2	CM Type, CM.AC ratio, CM Value, MC Value	Input	I,R,R,R	Included	Calculated field
2	Casein.2	Input	I	Excluded	Empirical data
2	Kinetin.2	Input	R	Excluded	Empirical data
2	Glutamine.2, Sucrose.2Agar.2, Zeatin.2, BAP.2,2-iP.2Thidiazuron.2, GA3.2, IAA.2NAA.2, 2,4 D.2, IBA.2	Input	I,I,I,R,R,R,R,R,R,R	Included	Empirical data
2	Calus.2, Soft, Pale, Compact.2, Liquid, Green.2, Red points, Loose, Roots, Necrotic Tissue	Output	B	Included	Empirical data

structure of the compounds and the manifestations of their effects, expressed by 2D, 3D or multidimensional models spread in 2D or 3D. The 3D QSAR methods are used to find a quantitative relation between the spatial structure of a chemical compound and its activity (5).

The proposed soft computing method is an original adaptation of the QSAR (AQSAR) concept with applied

clusterizations over predicted results and base statistical analysis.

The semantic substitutions we made are that:

- “biological activity” can be every result from an *in vitro* experiment such as the type of developed tissue and percentage of necrotic tissues, hereunder called Adapted Biological Activities (ABA);

MC type

Nutrient Media Type	IPM.type = 0	IPM.type = 1	IPM.type = 2
CM.type = 0	0 – both media invoke shoots	10 – (roots, shoots)	20 – (all, shoots)
CM.type = 1	1 – (shoots, roots)	11 – both media invoke roots	21 – (all, roots)
CM.type = 2	2 – (shoots, all)	12 – (roots, all)	22 – both media invoke shoots, roots, or calli (all, all)

- “structural properties” can be the properties of the nutrient medium, i.e. the concentrations of the chemical constituents, hereunder called Adapted Structural Properties (ASP).

This semantic substitution allows to create an Artificial Neural Network (ANN) with input signals ASP and output signals ABA, which describes the relationships between nutrient media (as quantity of compounds and characteristics) and the developed *in vitro* tissue. The general idea of the study was to train the provided experimental data and use it to make “artificial experiments” with other nutrient media (different from those in the original biotechnological experiments) in order to collect predicted results to be analyzed and give theoretical directions for real *in vitro* experiments with optimized parameters of the biotechnological scheme.

Biotechnological experiments are time-consuming and limited by the empirical method of selecting the conditions for effective callusogenesis, organogenesis, regeneration, and morphogenesis. Methods such as AQSAR could be used as models producing quick, better and cheap results, with economy of materials and resources.

The objective of the study was to develop a soft computing model based on adapted QSAR methodology in order to model the biological experiment data for improvement of the established *in vitro* system for *Rhodiola rosea* cultivation as a rare medicine plant with proved treatment potential.

## Materials and Methods

### Experimental background

A series of experiments were performed with the aim to observe different *in vitro* responses and plant development, using 21 combinations of explants and culture media. The initial plant material was collected from wild growing plants in the Rilla National Park and the area Beglika, Rhodope Mountains. The explants were: apical buds (10 mm to 12 mm), adventitious shoots (10 mm to 15 mm), stem segments (8 mm to 10 mm), rhizome buds (10 mm to 12 mm) and rhizome segments without buds (10 mm to 12 mm) isolated from wild growing plants. Nutrient media variants of the basic Murashige and Skoog medium (4) were made, containing different concentrations and combinations of phytohormones – zeatin, BAP, kinetin, IAA, NAA, IBA, 2-iP, 2,4-D; and supplements – casein hydrolysate and glutamine. The concentration of sucrose was 30 g/L and of agar-agar 7 g/L in the medium for organogenesis,

and 10 g/L to 20 g/L and 5 g/L to 6 g/L, for callusogenesis and root induction, respectively.

The data was used to create two data sets (depending on the biological process – growing/rooting). These data sets were set up in order to provide ANN training and forecasting, as the validation of the training process is assured by the desired maximum error level which was set to be 0.0001. This means that the predicted results could produce a bigger error if the trained data is insufficient or incomplete. One of the advantages of using artificial neural networks is that the network could provide training until reaching criteria of maximum or average error. This is the method’s proof for correctness, because these error rates represent the error between predicted and original data. The validation tool is another option which can guarantee optimal results, but in this case it was not used because of the problem with overtraining the data. The validation tool is good to be used only if we have large data sets and if there is a need of a closed network with strong upper and down limits concerning the predictors.

### Soft computing method AQSAR – modeling and analyses

AQSAR is a soft computing method (3) that involves ANN, clusterization techniques and graphical interpretation of the data. It facilitates the extraction of hidden information about the relationship between the quantity of different phytohormones, explant type, nutrient medium type and/or media combination type as well as *in vitro* response (development of explant as Boolean outputs for each type of response). Creating the ANN is the core of the method used in both biotechnology related stages. The analysis of the forecasted data is done in two groups of interest: 1) ANN analysis – includes Analysis of Importance & Sensitivity, and analysis of error rates; 2) multidimensional clusterizations and graphical representations over forecasted data.

The first group of interest is essential because the analysis or the error rate proves how reliable the training is, as the Importance and Sensitivity give information about which the most important predictors in the ANN are, and how the predicted data will be affected if there is a change in one of the predictors.

The second group of interest represents the analysis over forecasted results, which show how the perfect *in vitro* biotechnological scheme should look in theory. In this line, several criteria were followed:

- to find the nutrient media which provide the development of explants to plants and leaf rosettes so they could be rooted during the next passage and define the appropriate explant types (concerning the first biotechnology stage);
- to find the media combinations which provide rooting of plants and leaf rosettes (concerning the second biotechnology stage);
- more MCs that guarantee rooting of plant and/or leaf rosettes at less costs.

The main advantages of the AQSAR are:

- strong relation to the biotechnology experiments and biotechnology scheme which have to be optimized;
- trusted and accurate method for predicting – ANN, which is the core of the method;
- calculated predictors which represent the weighted value of the nutrient medium;
- calculated predictors which classify every nutrient medium (Culture Medium – CM)
- calculated predictors which classify the media combination and support the relationship between both biotechnological stages;
- innovative scatter plot analysis of Importance & Sensitivity of the predictors of ANNs;
- easy-to-do clusters of data by explant type, *in vitro* response and price range of CM and/or MC with Pivot table;
- innovative application of wind rose diagrams which shows the relationships between explant type, MC type and MC price and gives the final conclusions;
- applicability to other *in vitro* experiments with the same or different plants;
- possibility to expand the ANNs with genotype and/or ecotype information and any other classifiers of the data or experiment.

## Modeling

**Architecture and parameters of the neural networks for both Stages 1 and 2.** EasyNN Plus (<http://www.easynn.com/>) was used for creation, training and querying the ANN as well as for obtaining parameters of the trained networks, and statistical and clusterization analysis, done with Microsoft Excel 2007 (<http://office.microsoft.com>). EasyNN Plus is based on a back-propagation algorithm. The verification mechanism of the results obtained by the training of the neural network is statistical – based on the required error levels set as input parameters. The maximum error level was set to be 0.0001 at each stage of predicting.

The initial information for the AQSAR analysis was taken from both biotechnological passages. The following signifiers were chosen to be included in construction of the ANNs: explant type; nutrient medium – the concentrations of different hormones and other nutrients, each of them presented

separately with its quantity. Auxin–Cytokinin (AC) ratio – the calculated coefficient was based on concentrations of medium nutrients; medium type based on the AC ratio; nutrient medium price – as signification for “excluded” hormones from the training process of neural networks, i.e. play a role of weighted value of the nutrient medium; obtained results from the biological experiments (like percentage of expression, development of buds, shoots, plants, calli, roots, necrotic tissues). For each biotechnological passage there is a stage of predicting, and the second stage uses also the input and output parameters from the first stage. This ensures that the connection between the initial data and their future development is not lost, if some of the phyto regulators are missing. They save the synchronization with the processes of control of shoot apical meristem and root apical meristem formation as well as their role in *in vitro* organ regeneration (7).

A total of 488 nutrient media were created and also 35 variations based on standard nutrient media were used. The artificial experimental query row for each type of explant (apical bud, rhizome buds, stem segment, rhizome segments, and leaf node) was constructed. Thus, more than 2 400 artificial experimental query rows were obtained. As there are too many possibilities to make different combinations of initial plant media and culture media, the decision made was not to additionally expand the range of query rows. Accordingly, only one variant was used for creating the media combinations, but was applied for each explant type, i.e. the combination (N235; N345) will occur exactly 5 times – one for each explant type (apical bud, leaf node, etc.). The term Initial Plant Medium/a (IPM) is used to define the first stage and Culture Medium/a (CM) is used to define the second stage. The combination between IPM and CM is called “Media Combination” (MC).

Standard parameters for creating the ANNs like Learning rate 0.6 and Momentum 0.8 were used. The best training time and error levels were obtained for architectures Input nodes – Hidden Layer 1 nodes – Hidden Layer 2 nodes – Output nodes: 15-16-10-6 and 36-20-8-10 in respect of predicting stages. The number of Learning cycles at Stage 1 is 6351, and at Stage 2, 887. The excluded columns are 4 at Stage 1 and 7 at Stage 2. These columns contain the same value. That is why they could not take part in the training process. The excluded columns at Stage 1 are GA<sub>3</sub>, IBA, Sucrose and Agar. In addition, at Stage 2 also the TDZ phyto regulator from IPM, as well as Kinetin and Casein – compounds of the CM, were excluded because of the less raw data. Because they are phyto regulators, it was decided to include other calculated columns, which represent the weighted value of the media such as Medium AC ratio and Medium price (Table 1).

**Input tables.** Table 2 shows the features of each column, used for creating the artificial neural network. The “Used at Stage No.” shows which column is used at Stage 1 and Stage 2 of the predicting process, and which only at the second one. The “Signal type” shows if the column is used as Input (Predictor) or as Output (Predicted). When it is labeled as “Output/Input”, it means that at the first stage it is predicted, and then at the



second stage the predicted value is used as predictor. The data type can be Integer(I), Real(R), Boolean(B), or Text.

The output columns are the initial response of the experiments using IPM. Most of them were defined as Boolean because of the experiment itself: there are cases that show that we could obtain more than one type of response, i.e. for example it is possible to find at the same time plants and leaf rosettes. Media classification is done by calculating the IPM. AC ratio. The high ratio of cytokinin to auxin favors shoot production, whereas a high auxin to cytokinin ratio favors root production (8). Therefore the established criteria were as follows:

Interval [0; 0.5] – root apical meristem formation medium, type 1;

Interval (0.5; 1) – shoot and root apical meristem formation and calli formation medium, type 2;

Interval [1; 24] – shoot apical meristem formation medium, type 0.

According to this, we created a calculated field, named MC type (Table 3). It shows whether the Medium Combination is clear type 0/11/22 or types 1, 2, 10, 12, 20, 21. The formula is:

$$\text{MC Type} = 10 * \text{IPM type} + \text{CM type}$$

After the training processes were done, a basic statistical and clustering analysis was completed with initial and predicted data. The analysis includes the following steps: 1) Error rates analysis; 2) Importance & Sensitivity analysis; 3) Analysis of the explant development in IPM and MC (IPM; CM) – filter media by biotechnological effectiveness; 4) Analysis of cost effectiveness over biotechnologically effective media; 5) Analysis of MC which meet the criteria of effectiveness.

## Results and Discussion

### Error rates analysis

The obtained maximum error rate of both stages about the learning process was under 0.00001 (Fig. 1, Fig. 2). We believe that such an error does not have the effect of overtraining because the validating mechanism is dismissed. After the analysis of the forecast results, a small part of them (generating an error of about 0.0028) was found, about 5 to 7, which could be defined as “calli” at Stage 1 but were not predicted as calli. Such an error is small enough and in terms of the ANN is acceptable to consider that there is no overtraining. The main property of ANNs to not be overtrained is that they can predict even values out of ranges. As we do not have a big data set, they could not be defined as comprehensive.

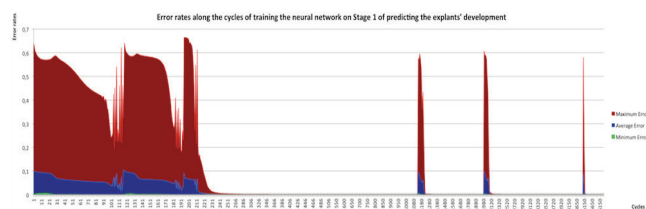


Fig. 1. Error rates at Stage 1.

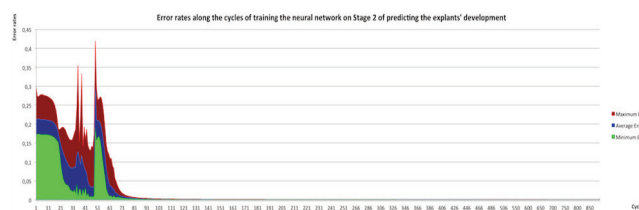


Fig. 2. Error rates at Stage 2.

### Importance & Sensitivity analysis

Instead of the standard separated interpretation about Importance and Sensitivity, a clusterization analysis was provided, every point consisting of two components: Importance and Sensitivity. Thus, the analysis of Importance and Sensitivity finds the relationships between them.

The distribution of predictors at Stage 1 by their Importance and Sensitivity is presented in Fig. 3.

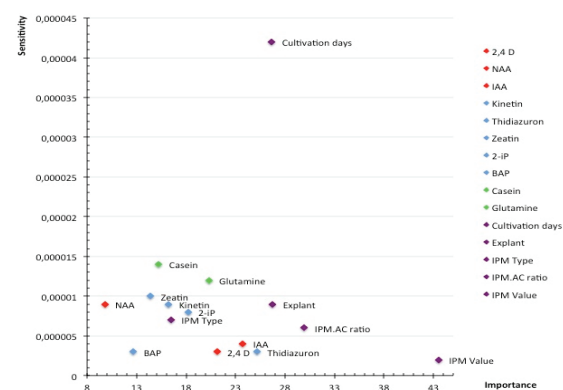


Fig. 3. Scatter plot analysis of Importance and Sensitivity at the first forecasting stage.

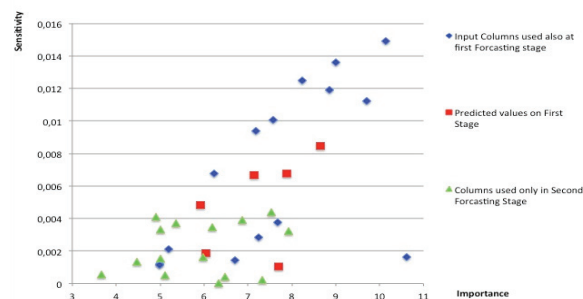


Fig. 4. Scatter plot analysis of Importance and Sensitivity at the second forecasting stage.

The number of cultivation days was detected as a factor with a high level of sensitivity. Therefore, a change from 30 days to 40 days could be fatal and will produce necrotic tissues. Most of the query rows include preferably cultivation days = 30. The IPM Value has the highest level of importance, but the lowest sensitivity, which means that the hypothesis about the weighted value of the media is true. Thus, to consider the effect in the results, a bigger change is needed. The low price of most of the phytohormones proves the observed results. Almost the same holds true for the IPM.AC ratio. The

other group of interest was Casein, Glutamine, and Explant type complex. Casein and Glutamine are commonly used for inducing growth, no matter if it is of shoots or of roots.

The analysis just proved their good level of importance and high level of sensitivity as compared to all the others. It is valid that as the cytokinin is more important, it is less sensitive, which is also true for common supplements like Glutamine and Casein, while for the auxins there is no specificity.

The distribution of predictors at Stage 2 by their Importance and Sensitivity is given in Fig. 4.

The analysis focused on the distribution between Columns used only at the second stage of predicting (G2), predicted values at the first stage (G1) and input columns used also at the first stage (G12). It was defined that the G2 predictors have no more than the average ratio of importance and a low level of sensitivity, which means that there are more important and more sensitive parameters than them. That is why, much different results should not be expected if slight changes are made in their values. On the other hand, it was obtained that the predicted values depend much more on the predicted results and predictors from the first stage, and also they are highly sensitive, which means that, if we find media which produce something else, this will generally affect the predicted result at the second stage.

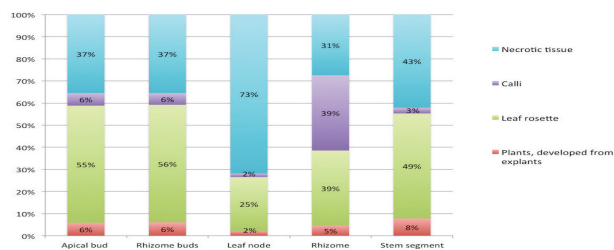


Fig. 5. Percentage proportions of developed plants, leaf rosettes, calli and necrotic tissues from the explants, correlated to the type of the explant.

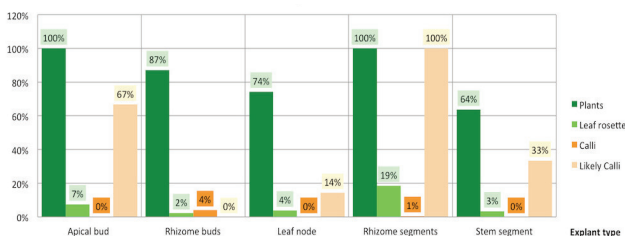


Fig. 6. Percentage of culture media that induce rooting by explant type and response after IPM developing process.

Moreover, some predictors were less important and less sensitive. These predictors are the columns with names: Glutamine, Zeatin, IAA, IPM.AC ratio and IPM.Value. There is direct proportional dependence between the importance and the sensitivity about G2 and G12, i.e. opposite to the dependences at Stage 1. This means that the data from the first stage are more stable than the data from the second stage, which is normal because at the first stage we have more training examples than at the second one.

## Analysis of the IPM, CM and Media Combinations (MC) between IPM and CM

**Explant development in IPM.** A 100 % stacked column chart was used to estimate the percentage proportions of developed plants, leaf rosettes, calli and necrotic tissues from the explant, correlated to the explant type (Fig. 5).

The biotechnological interest is to define nutrient media which can develop plants and leaf rosettes so they could be rooted during the next passage. The leaf node is the less productive one, as the other have a mean of about 6 % for plant growth and 49 % for leaf rosettes growth, i.e. 6 % of all the nutrient media invoke plant growth and 49 % of the nutrient media invoke leaf rosettes growth.

Explant development in MC (IPM; CM). Clusterization was used to estimate the percentage of MCs that induce a rooting process, and a base statistics about the absolute number of MCs that induce a rooting process is provided in Fig. 6. The clusters are based on explant type, and every cluster consists of 4 columns: plants, leaf rosettes, calli and likely calli.

The analysis showed that, to achieve successful rooting, nutrient media which favors rooting plants and leaf rosettes is needed, mostly from apical buds and rhizome segments, as there are also very good possibilities to use explants from the other types. Even if the “Calli Like” percentage is high, the number of nutrient media favoring this might be 5 only. It is not recommended to try to induce a rooting process for calli as far as almost 100 % of them necrotize. The criterion of effectiveness is defined as: successful rooting of plant or leaf rosettes at less costs.

## Analysis of cost effectiveness over biotechnologically effective media

**Cost analysis of the nutrient media over plants (Fig. 7, Fig. 8, Fig. 9).** There are much more MCs of high price level that could support plant growth followed by rooting, as the price levels are 7 euro/L to 8 euro/L, 10 euro/L to 11 euro/L, 14 euro/L to 15 euro/L. The decision made was to take only the MCs with a price range between [0; 6] euro/L. The number of MCs corresponding to this interval is only 22 MCs that meet the criteria (out of 119), or it is about 18 %. No MC for explant type “rhizome segment” meeting our criteria was found.

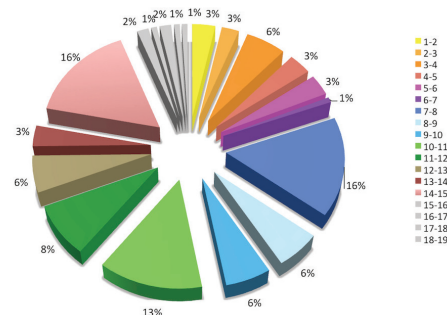
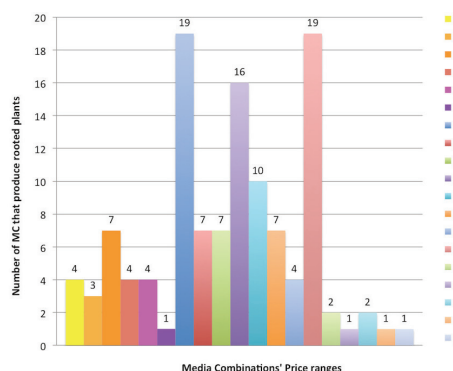
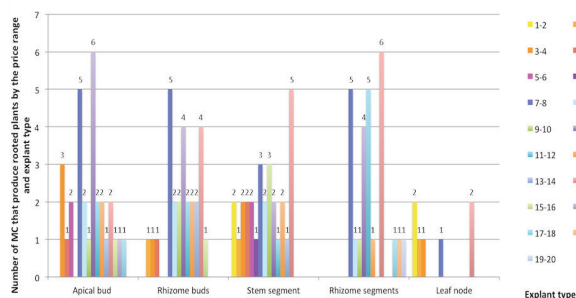


Fig. 7. Percentages of MC that induce developing roots from plants by the price range (euro/L) of media combinations.

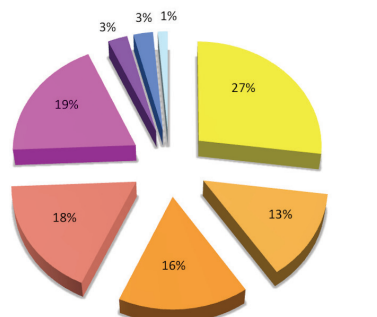
**Cost analysis of the nutrient media over leaf rosettes (Fig. 10, Fig. 11, Fig. 12).** There is much more MCs of low and middle price class that could support leaf rosettes growth and later formation of roots. In the price range [0; 6] euro per liter there are 5 groups with similar percentages, and the average is about 16 %. Positively, the number of MCs within this range is large enough, i.e. 71 media combinations out of 74 meet the criteria, accounting for about 96 % of all the combinations. There is no limitation about the explant type.



**Fig. 8.** Number of MC that produce rooted plants by the price range (euro/L) of media combinations.



**Fig. 9.** Number of MC that produce roots from plants by the explant type and the price range (euro/L) of media combinations.

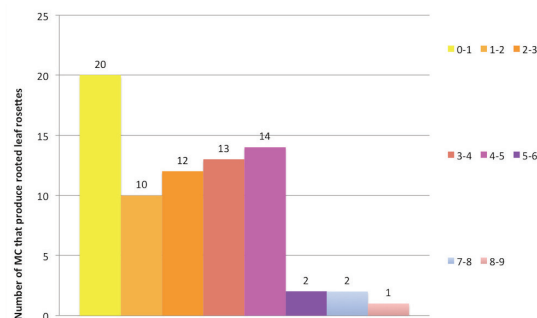


**Fig. 10.** Percentages of MC that induce developing roots from leaf rosettes by the price range (euro/L) of media combinations.

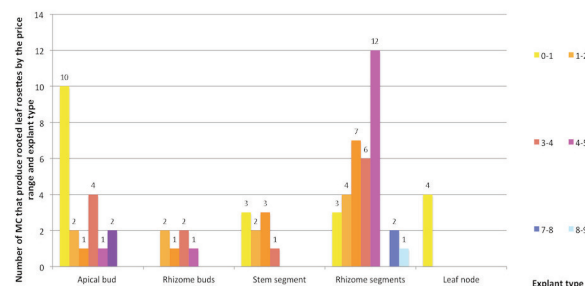
#### Analysis of MCs which meet the criteria of effectiveness

To determine which MCs meet the criteria of effectiveness, we did a filtration by IPM response = {plants (Fig. 13) or leaf rosettes (Fig. 14)} and CM response = rooting. The observed

fields were Explant type; Media Combination's name; Media Combination's Type; Media Combination's Value, where the last three are calculated fields. Data was sorted by ascending by the price of the MC.



**Fig. 11.** Number of MC that produce rooted leaf rosettes by the price range (euro/L) of media combinations.



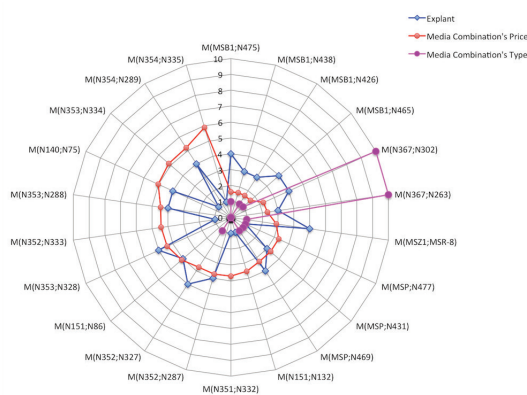
**Fig. 12.** Number of MC that produce roots from leaf rosettes by the explant type and the price range (euro/L) of media combinations.

The obtained results could be discussed in the context that the type of explant has no relation with the final results, as the data were several times filtrated. It was found that:

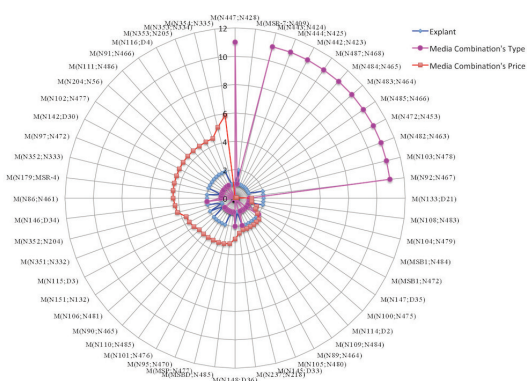
There are MCs that have the same productivity both, for the plants and for the leaf rosettes. Also 5/6 of them have MCs type = 0 (i.e. these media combinations strongly inducing the growth of shoots); and 1/6 of them, type = 1. The MCs have the following coding: M(MSP;N477), M(N151;N132), M(N351;N332), M(N352;N333), M(N353;N334), M(N354;N335). There are 4–5 MCs where media combinations are a variation with MC type =1, with different explants and prices.

These media are within the same price level with minimal changes. This supports our argument that MCs in closed price levels should have similar effects and compounds. There are 13 media combinations of type 0 where the price level for plants is between 4 and 5, but there are indications that such media are mostly exceptions. There are 2 MCs of type 2; i.e. IPM supplies shoot growing, while the CM supports the growing of shoots, roots and calli. The initial response type leaf rosette price level is about 2.80 euro/L. There are 2 MCs of type 10 – a non-standard type of MC, in which IPM provides root growing, while the CM provides shoot growing. This occurs when the explants are leaf nodes; stem segments have initial response “plant”.





**Fig. 13.** Analysis of media combinations which produce plants developing roots in economy effectiveness.



**Fig. 14.** Analysis of media combinations which produce leaf rosettes developing roots in economy effectiveness.

Overall, after using the AQSAR for analysis of our data, it can be emphasized that: 1) It is important in case of limited resources to try to avoid a great number of explants of rhizome segments; 2) Leaf rosettes are the most appropriate type of initial response to the nutrient medium treatment, as there is no limitation about the type of explant and it is quite easy to find MCs meeting our criteria. Rooted leaf rosettes could easily become *in vitro* cultivated “plants” with roots, which are ready for transfer to natural conditions. There are 12 nutrient media of type 11 and total cost under 0.1 euro/L, and 31 nutrient media of type 1, explant type = rhizome segments and total cost in the range of [1.62–4.43] euro/L.

## Conclusions

The main conclusions are that AQSAR under a softcomputing pipeline could be used for modeling of *in vitro* data. The development of such a data analysis model gave us the opportunity to forecast nutrient media never tested in the lab, and also to theoretically derive the optimal conditions, including the quantities of nutrient media constituents for *in vitro* cultivation. The most important results are that we are able to define the most effective theoretical model (both from a biotechnological and economical point of view) to produce *in vitro* cultures ready to be transferred in the soil, meaning grown and rooted plants.

The proposed AQSAR as a representative of regression and classification models wrapped in a soft computing processing flow might be applied to other medicinal plants, especially if their *in vitro* cultivation is very difficult. Creating the neural network could be done with consideration of the species, ecotype, explant type, phytohormones etc. This might contribute to overcoming the problem of the limited resources of the biological experiment.

The provided statistical and clusterization analysis successfully gave an answer to the following questions: “How to find the optimal theoretical parameters of an *in vitro* biotechnological system?”, “How to choose the proper type of material for explant and proper nutrient medium?”, “What are the guidelines for reaching biotechnological effectiveness under economical restrictions?”

The main advantages of AQSAR are:

- Strong relationship to the biotechnology experiments and biotechnology scheme which have to be optimized and at the same time possibility to expand the ANNs with genotype and/or ecotype information and any other classifiers of the data or experiment, or to be applied over other *in vitro* experiments with the same or different plants.
- Trusted and accurate method for predicting – ANN, which is the core of the method; Innovative scatter plot analysis of Importance & Sensitivity of predictors of ANNs; Innovative application of wind rose diagrams which shows the relationships between explant type, MC type and MC Price and gives the final conclusions.
- Weighted value of the nutrient medium, classifiers about CM and MC, supporting the relationship between both biotechnological stages.

- Trusted and accurate method for predicting – ANN, which is the core of the method; Innovative scatter plot analysis of Importance & Sensitivity of predictors of ANNs; Innovative application of wind rose diagrams which shows the relationships between explant type, MC type and MC Price and gives the final conclusions.
- Weighted value of the nutrient medium, classifiers about CM and MC, supporting the relationship between both biotechnological stages.

- Weighted value of the nutrient medium, classifiers about CM and MC, supporting the relationship between both biotechnological stages.

## Acknowledgements

The research was supported by the National Science Fund of Bulgaria—Project for Junior Scientists DMU 03/55.

## REFERENCES

1. **Brown R.P., Gorbarg P.L., Ramazanov Z.** (2002) *Herbal Gram*, **56**, 40-52.
2. **Egelund J., Skjot M., Geshi N., Ulvskov P., Petersen B.L.** (2004) *Plant Physiol.*, **136**(1), 2609-2620.
3. **Li X., Ruan D., van der Wal A.J.** (1998) *Int. J. Intell. Syst.*, **13**(2-3), 287-300.
4. **Murashige T., Skoog F.** (1962) *Physiol. Plant.*, **15**(3), 473-497.
5. **Nikolova N.** (2001) Ph.D. Thesis, Sofia. (In Bulgarian)
6. **Platikanov S., Evstatieva L.** (2008) *Econ. Bot.*, **62**(4), 621-627.
7. **Su Y.H., Liu Y.B., Zhang X.S.** (2011) *Mol. Plant*, **4**(4), 616-625.
8. **Tasheva K, Kosturkova G.** (2010) *Cent. Eur. J. Biol.*, **5**(6), 853-863.

2. **Egelund J., Skjot M., Geshi N., Ulvskov P., Petersen B.L.** (2004) *Plant Physiol.*, **136**(1), 2609-2620.
3. **Li X., Ruan D., van der Wal A.J.** (1998) *Int. J. Intell. Syst.*, **13**(2-3), 287-300.
4. **Murashige T., Skoog F.** (1962) *Physiol. Plant.*, **15**(3), 473-497.
5. **Nikolova N.** (2001) Ph.D. Thesis, Sofia. (In Bulgarian)
6. **Platikanov S., Evstatieva L.** (2008) *Econ. Bot.*, **62**(4), 621-627.
7. **Su Y.H., Liu Y.B., Zhang X.S.** (2011) *Mol. Plant*, **4**(4), 616-625.
8. **Tasheva K, Kosturkova G.** (2010) *Cent. Eur. J. Biol.*, **5**(6), 853-863.

3. **Li X., Ruan D., van der Wal A.J.** (1998) *Int. J. Intell. Syst.*, 13(2-3), 287-300.

4. **Murashige T., Skoog F.** (1962) *Physiol. Plant.*, **15**(3), 473-497.
5. **Nikolova N.** (2001) Ph.D. Thesis, Sofia. (In Bulgarian)
6. **Platikanov S., Evstatieva L.** (2008) *Econ. Bot.*, **62**(4), 621-627.
7. **Su Y.H., Liu Y.B., Zhang X.S.** (2011) *Mol. Plant*, **4**(4), 616-625.
8. **Tasheva K, Kosturkova G.** (2010) *Cent. Eur. J. Biol.*, **5**(6), 853-863.

- 5. Nikolova N.** (2001) Ph.D. Thesis, Sofia. (In Bulgarian)

6. Platikanov S., Evstatieva L. (2008) Econ. Bot., **62**(4), 621-627.

7. Su Y.H., Liu Y.B., Zhang X.S. (2011) Mol. Plant. 4(4). 616-625.

8. **Tasheva K, Kosturkova G.** (2010) Cent. Eur. J. Biol., **5**(6), 853-863.