A COST-EFFECTIVE METHOD FOR IDENTIFYING NUTRIENT MEDIA COMBINATIONS PRODUCING PLANTS WITH MAXIMUM BIOACTIVE SUBSTANCES

Valeriya Simeonova, Krassimira Tasheva, Georgina Kosturkova,
Dimitar Vassilev

ABSTRACT. The aim is to find a cost-effective method of identifying nutrient media producing identical plants with maximum performance in terms of the bioactive substances contained in them. An adaptation of QSAR is used. The "spatial structure of the chemical component" is replaced with "the multidimensional structure of the nutrient medium" or with the treated day schemes. For each process, a separate forecast is made. All nutrition media produced in silico are based on the ranges of phytonutrient hormones in biotechnological experiments. We found 43 theoretical combinations of media with more than 80% success under conditions of limited resources in the price range of [0–1,5] euro/liter. The obtained results can be used as: a theoretical guideline for determining the optimal nutrient media and combinations; to the study of other medicinal plants in order to establish effective biotechnological schemes for growth and rooting that are also cost-effective; using ANN, taking into account the species and the ecotype.

1. Introduction. Bulgaria is one of the major world exporters of medicinal plants. Bulgarian medicinal plants are among the best in the world. The intensive use of medicinal plants leads to the extinction of some species, the limitation of natural populations and biodiversity. Rhodiola Rosea L. or Golden Root is an endangered species. It is protected by law in Bulgaria and other countries (Great Britain, Finland, Russia and Mongolia). The extract of roots and rhizomes is used in the field of prevention and treatment of socially significant diseases of the cardiovascular and central nervous system. In recent years, the in vitro culture method has been used as one of the advanced biotechnology systems to produce a large number of identical plants for a short period of time that are free of pathogens. Such requirements are posed by the needs and demand of industries such as horticulture, agriculture and forestry. This is particularly useful for species with wide or massive use, or with slow and difficult cultivation under natural conditions.

Therefore, it is important to find a cost-effective way of identifying nutrient media producing identical plants with maximum performance in terms of the bioactive substances contained in them. Here, the concept of "cost-effective" is linked, on the one hand, to the multiple scaling of the time needed for testing, on the other hand, the repeated shrinking of the cost of materials and consumables from real experiments, and by a third party—highest productive culture media. This allows biotech studies to be conducted in "depth", i. e., massive investigation of effect in small differences in concentrations.

The Quantitative Model of Structure-Activity Dependency (QSAR) for analysis is interdisciplinary and uses knowledge in pharmacology, molecular biology, organic and quantum chemistry, analytical methods for structure analysis, mathematical and engineering methods, statistics, informatics, etc. [1]. QSAR is a method based on the hypothesis that similar biological activity is

---

Abbreviations: AC ratio—cytokinin/auxin ratio; AI—Artificial Intelligence; ANN—Artificial Neural Network; BAP—N6-benzylaminopurine; CM—Culture Meduim/a; 2,4-D—2,4-dichlorophenoxyacetic acid; GA$_3$—gibberellic acid; IAA—Indolyl-3-acetic acid; 2-iP—6-(y,y-dimethylallyl amino) purine; IBA—Indole 3-butyric acid; IPM—Initial Plant Medium/a; MC—Media Combination; NAA—a-naphthyl acetic acid; TDZ—thidiazuron; QSAR—Quantity Structure-Activity Relationship.

determined by common structural characteristics. The aim is to quantify the relationship between the chemical components (in our case, the concentrations of phyto-regulators and other components of the nutrient medium) and the biological activity they provoke. Analyses are often graphic and depending on the dimensionality of the data is subject to different two-dimensional or three-dimensional graphic models, and any combination thereof, as projected data are multidimensional. Three-dimensional QSARs are methods used to detect the quantitative link between the spatial structure of the chemical component and its activity [4].

Here we show a successful adaptation of the method in its part of the "spatial structure of the chemical component," which we replace with "the multidimensional structure of the nutrient medium as determined by the different phytohormon concentrations and other quantifiable environmental and experimental conditions" or with the time intervals in which explants are treated with various decontaminating chemicals.

2. Brief information about applied methods and software.

Bioinformatics experiment. The data is produced by conducting biological processes cleaning, germination, propagation and rooting, and collecting the observed results. The data is preprocessed and then represented in training tables, used by ANN. Constructed training tables follow the logic of the applied biological experiments. Each data set serves as basis for the neural network construction, respectively training tables and networks' architectures. The results of the trainings are represented in a single table, which is then analyzed statistically.

Artificial Neural Networks (ANN). Neural Networks (NN) is a subclass of Artificial Intelligence (AI) methods. It is known that AI is the first step of the QSAR methodology. ANN are chosen because of their strong impact in forecasting and classifying data. The method is applied on data with high percentage of uncertainty and imperfection.

EasyNN Plus [3] is a software environment for ANN construction, optimization, execution and visualization. The application uses back-propagation training algorithm. The software works with different type of files

for input/output like: excel [2] workbooks, text files and .CSV files. It is implemented real-time graphic generation and visualization, including for errors estimation. It is possible to import an excel file with queries after the training process is done. All exported results related to this study are further analyzed. The verification mechanism includes statistics and it is based on the average error value less or equal to 0.0001.

The forecasting stages are developed in the context of the process controlling of tissue and roots formation because of their role in *in vitro* reconstruction [5].

The biological experiments. The data used for bioinformatics analysis comes from biological experiments conducted in 2 to 3 iterations. The obtained results are summarized and described in detail [6, 7, 8, 9].

3. Applied methodology. In this paper, we set the focus over the "cost-effective"-ness and "time-effective"-ness of the applied method to identifying nutrient media producing identical plants with maximum performance in terms of the bioactive substances contained in them. That is why it is important to describe what that is mean to be for us.

The concept of "cost-effective" is used in terms of:

- Artificial tests can be done using high range of examined materials (often limited source in real world). Artificial tests make it possible because of the assuming that all artificial values are possible in nature, so they could be used for making queries. The artificial values are only those we really can find in nature and are actual possible characteristics.

- Artificial tests with a high range of combinations from different consumables' concentrations. They provide:

  o *Once*: identifying the low-cost consumables' combinations, producing identical plants with maximum performance in terms of the bioactive substances contained in them.

  o *Second*: a chance to make *in vitro* experiments only with the best from Once.

The concept of "time-effective" is used in terms of:

- Artificial tests with a high range of combinations from different consumables' concentrations need in way less time than providing real *in vitro* experiments.

- The chance to make *in vitro* experiment only with the best results from *Once* shorten the time to discover the best performance in biotechnological aspect.

This allows biotech studies to be conducted in "depth", i.e. massive investigation of effect in small differences in concentrations.

The need to adapt QSAR is provoked by several reasons:

- It is used originally in drug discovery.

- In drug discovery, the 3D model represents the molecular structure.

- With appropriate adaptations could be used for describing any relationship.

- In the case of *in vitro* experiments for bioactive substance:

  o it describes the relation: process success ~ medium & time period & material

  o the 3D model represents cluster formations with combinations of media that assures in less time highly production of *in vitro* plants with minimum cost. We called 3D because of the three dimensions of the analysis.

  o the "spatial structure of the chemical component" is replaced with "the multidimensional structure of the nutrient medium as determined by the different phytohormon concentrations and other quantifiable environmental and experimental conditions" or with the time intervals in which explants are treated with various decontaminating chemicals.

We follow four steps to maintain the data (see Fig. 1):
1) Data Exploration;
2) Data Preparation;
3) ANN Production (based on adapted QSAR);
4) Analysis of the forecasted results.

```
┌─────────────────────────────────────────┐
│            Data Exploration              │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│            Data Preparation              │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│   ANN Production (based on Adapted QSAR) │
└─────────────────────────────────────────┘
                    ↓
┌─────────────────────────────────────────┐
│      Analysis of the forecasted results  │
└─────────────────────────────────────────┘
```
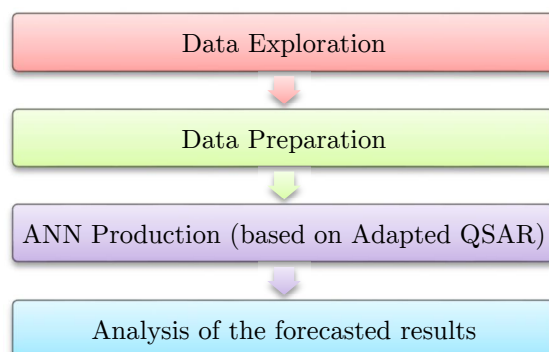
Fig. 1. The algorithm of the methodology

Step 1: Data Exploration. It is one of the preconditions to know the data if it is need to derive information. Data knowledge gives the possibilities to prepare and/or to preprocess the data if needed. We found that:

- The initial data comes aggregated, described with standard deviation.
- There are 4 processes: cleaning, germination, propagation, rooting.
- Material can be:
  - seeds—fresh or old
  - Explants—buds, adventive buds, root buds, stem segment, root segment
- There are Price lists for:
  - cleaners
  - medium nutrients and phytohormones
- Treatment days.

Step 2: Data Preparation. Data preparation has the aim of identifying which characteristics should be included in the training table (named as input neurons) and how many of them will be preprocessed (disaggregated, aggregated, included in new artificial characteristics— calculated by some of the main, etc.). Here are the processes that are used to obtain the necessary input data:

- Disaggregation process is needed, because often data comes already aggregated. For this process, all is needed to know is:
    - generators for normal distributed data
    - the mean
    - standard deviation
    - the results
- Identifying the initial input neurons for the ANN: all supporting information about current process is used
- Generate the calculated input neurons: supports forecasting and conclusions:
    - Medium_value $= \sum_{k=1}^{n} \binom{n}{k} p_k q_k$, where: p_k is the price and q_k is the quantity of the kth phytohormone
    - Medium_coefficient: calculates the auxin/cytokinin ratio, based on phytohormone classification
    - Medium_type: based on the medium_coefficient there are:
    - Type 0 with medium_coefficient in [1;24]—stimulates growth and development of the plant
    - Type 1 with medium_coefficient in [0;0.5]—stimulates root formation
    - Type 2 with medium_coefficient in (0.5;1)—stimulates as well as growth, development and root formation and caluses
    - G.Price.Days = GD*medium_value/10, where GD is germination days. This field compensate the lack of data due to the germination days was excluded at the stage of optimizing the training table
    - MC_Price = G.Price.Days + Rooting price
    - + a lot of other calculation fields supporting those we finally used for training such as: the phytohormone classification as auxin to cytokinin, price list, price calculation by unit, vlookup tables.

Step 3: ANN Production (based on Adapted QSAR). This step includes most of the processes needed to be done to obtain trained tables, which could be queried (see Figure 2). The final one is to produce forecasted results, which are analyzed at Step 4.
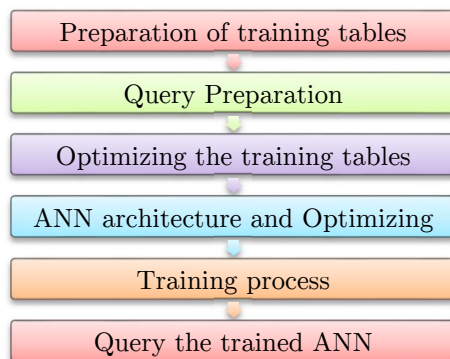
Preparation of training tables
Query Preparation
Optimizing the training tables
ANN architecture and Optimizing
Training process
Query the trained ANN

Fig. 2. ANN Production (based on Adapted QSAR)—Processes

Preparation of training tables. For each biotechnological process, there is a separate training table, as for the process of cleaning they are two—one for the seeds and one for the explants. The aggregated training table is shown in Table 1.

Columns [2–6] define the way each input neuron participates in training table formation for each process (Cleanings, Germination, Propagation, Rooting). It can be: I (Input)—input data (neuron); O (Output)—forecasted values; E (Excluded)—eliminated because of the lack of different values.

Using computational fields in the training table are of great importance, because they are weighted variables used as features to the culture medium. A classification of the medium for the process of dissolving in vitro cultures was performed according to the AC coefficient (Medium_koeficient). As known high *cytokinin/auxin ratio* causes rapid growth while the high *auxin/cytokinin ratio* stimulates the formation of roots [5]. Therefore, the following criteria for the classification of nutrients are defined:

- The interval [0; 0.5]—stimulates roots formation, the medium is Type 1.
- The interval (0.5; 1)—stimulates as well as the growth and development of the plant, and the formation of roots and calluses, the medium is Type 2.

Table 1. Training table structure

Column name	Seeds Cleaning	Explant Cleaning	Germination	Propagation	Rooting	Data Type	Field Type
Plant Material	x	I	x	x	x	Text	Empirical
Ache 5%	I	I	x	x	x	Integer	Empirical
Bleach 2,2%	I	I	x	x	x	Integer	Empirical
Ethyl Alcochol (C2H5OH) 70	I	I	x	x	x	Integer	Empirical
HgCl2 (0,2%)	I	I	x	x	x	Integer	Empirical
Peroxyd (2,2) (3%)	I	I	x	x	x	Integer	Empirical
Seeds	O	x	x	x	x	Integer	Empirical
Period	I	x	x	x	x	Integer	Empirical
Cleaned	O	O	x	x	x	Real	Empirical
Development	O	O	x	x	x	Real	Empirical
GA3	x	x	I	E	I	Real	Empirical
Sucrose	x	x	I	E	I	Real	Empirical
Agar	x	x	I	E	I	Real	Empirical
Days	x	x	I	x	x	Integer	Empirical
VegetationPersentage	x	x	O	x	x	Real	Empirical
High	x	x	O	x	x	Real	Empirical
Explant Type	x	x	x	I	x	Text	Empirical
Zeatin	x	x	x	I	x	Real	Empirical
BAP	x	x	x	I	x	Real	Empirical
Kinetin	x	x	x	I	x	Real	Empirical
2-iP	x	x	x	I	x	Real	Empirical
IOK	x	x	x	I	I	Real	Empirical
ANO	x	x	x	I	I	Real	Empirical
Glutamine	x	x	x	I	x	Real	Empirical
Cazeine	x	x	x	I	x	Real	Empirical
2,4 D	x	x	x	E	x	Real	Empirical
TDZ	x	x	x	E	x	Real	Empirical
IMK	x	x	x	E	I	Real	Empirical
Medium_Value	x	x	x	I	x	Real	Calculated
Medium_koeficient	x	x	x	I	x	Real	Calculated
Medium_Type	x	x	x	I	x	Integer	Calculated
Medium_Type_test	x	x	x	I	x	Integer	Calculated
New_Shoots	x	x	x	O	x	Real	Empirical
New_Shoots1	x	x	x	O	x	Real	Empirical
shoots high	x	x	x	O	x	Real	Empirical
G.GA3	x	x	x	x	I	Real	Empirical
G.Sucrose	x	x	x	x	E	Real	Empirical
G.Agar	x	x	x	x	E	Real	Empirical
G.Days	x	x	x	x	E	Integer	Empirical
Rooting	x	x	x	x	O	Real	Empirical
G.Price.Days	x	x	x	x	I	Real	Calculated
MC Price	x	x	x	x	I	Real	Calculated
Pro Index	x	x	x	x	E	Real	Empirical

- The interval [1; 24]—stimulates growth and development of the plant, the medium is Type 0.

As weighted values, we also use the cost of the medium for the stages of the propagation and rooting processes (Medium_Value, MC_Price), and in the second process it is special that the cost of the germination (G.Price.Days) is multiplied by the number of days for germination and divided into 10. Thus, a dependence is provided between the germination days and the success of the rooting, as initially, germination days were filtered (excluded) from training due to lack of sufficiently diverse data.

Initially, the breeding index (Pro_Index) that was created at the end of the process was added to the rooting process data. After multiple training and validation tests, it was concluded that this column of information should be excluded from the learning process, because no data on the breeding medium are available to form its values, or more specifically, they are unusable in the training process due to the fact that the same environment is used that prevents the existence of any variety—a necessary condition for certain data to participate in the process.

Query preparation. It is well known that if a trained table needs to be queried, the queries need to include the information for the input neurons. This is why query tables were easy to create as different combinations of phytohormones' quantities with all supporting calculated fields, included as input neurons. It is important to say that if biotechnology experiment explores the upper and down values of quantities, the training and queries will give more accurate results.

Optimizing the training tables. Easy NN Plus provides optimizations on the training tables mainly in two directions: 1) Filters columns with same values; 2) Provides analysis of importance and sensitivity. Regarding the data in this case, this stage includes:

- For each neural network, there is an analysis of importance.

- For each neural network, there is an analysis of sensitivity.

- A joint analysis is made in form of scatter plot (see Figure 3):
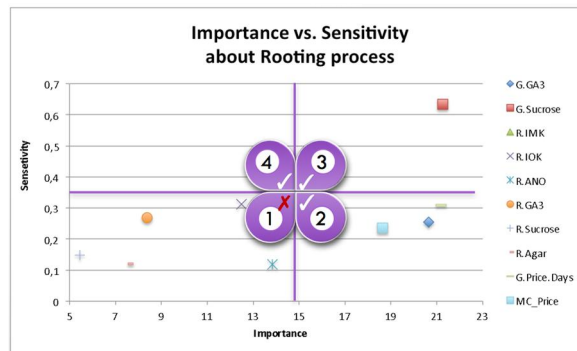
    o   X: Importance

Fig. 3. Example of a joint analysis of Importance and Sensitivity

- o Y: Sensitivity
- o Input neurons are classified in 4 categories: we prefer high importance + high sensitivity, but do not exclude those with low sensitivity.

The categories are:

Quadrant 1:

- the most undesired situation
- Result is slightly or even not reflected by
- There is a need of huge changes to make a difference in the result

    Quadrant 2:

- Better than (1 and 4) and worse than 3
- Result is reflected by
- There is a need of huge changes to make a difference in the result

    Quadrant 3:

- the most desired situation
- Result is totally reflected by
- Little changes make a huge difference in the result

    Quadrant 4:

- Better than 1 and worse than 3

- Result is slightly reflected by

- Little changes make a huge difference in the result

ANN architecture and Optimizing. Easy NN Plus gives the opportunity to set manually all the initial parameters of the training process, architecture of the hidden layers, stop conditions and error rates. Nevertheless, there are several optimizing controls, so the platform itself can find the most appropriate as:

- Parameters:
  - o　Learning rate
  - o　Momentum

- Architecture
  - o　The number of hidden layers
  - o　The number of neurons in hidden layers

Table 2. The optimized parameters and architecture structure of AINN

Parameters of AINN	S. C	Expl.C	G	P	R
Learning cycles	1335	219372	114454	1528	44673
Training error	0,0001	0,00009	0,0001	0,000098	0,000007
Input columns	6	6	4	13	9
Output columns	3	2	2	3	1
Serial columns	0	0	0	0	0
Excluded columns	0	0	0	6	4
Training example rows	12	19	24	15	27
Validating example rows	0	0	7	4	5
Querying example rows	200	512	357	1030	200
Excluded example rows	0	0	0	0	0
Duplicated example rows	0	0	0	0	0
Input nodes connected	7	6	4	13	10
Hidden layer 1 nodes	8	6	6	12	8
Hidden layer 2 nodes	4	6	6	6	6
Hidden layer 3 nodes	0	7	7	0	7
Output nodes	3	2	2	3	1
Serial input nodes	0	0	0	0	0
Serial output nodes	0	0	0	0	0

Learning rate	0,6	0,6	0,3902	0,4676	0,8859
Momentum	0,8	0,8	0,8709	0,8718	0,5472
Target error	0,0001	0,0001	0,0001	0,0001	0,0001
Validating error	N/A	N/A	0,078231	0,036244	0,006815
Validating example rows	N/A	N/A	100%	100%	100%
Adapting	No	No	Yes	Yes	Yes

In Table 2 each column (S.C., Expl.C, G, P, R) represents each one of the five ANNs. The maximum estimated error level is under 0.00001 in accordance with the predefined rules. There are two approaches that can be applied with the limited data collections:

1. Automated simulation that multiplies similar training examples (this option is provided and used in EasyNN). Could lead to over trained results, but in the context of the research it is desired situation. As compensation to the limited data collection it is not provided the validation step of the training procedure.

2. Using the standard deviation and the number of media, can be produced as much examples as we need. This research is the next step. It should provide more accurate analysis.

Step 4: Analysis on the forecasted results. As about the methodology, one of the main questions is how many analyses do we have on the forecasted results? The answer is shown in Figure 4; the numbers of the analyses are in the circles:

Fig. 4. How many analyses do we have on the forecasted results?

Stage 1: Cleaning

Analysis (1.1): number of cleaning schemas that assures the best development results by type of seeds (old and fresh)

Analysis (1.2): distribution of 1.1 by total cleaning time and choosing those with best development progress in short term cleaning schemas

Analysis (1.3): define the characteristics of the most time-effective cleaning schemas.

The same three analyses were provided for explants too.

Stage 2: Germination

Analysis (2.1): number of media for the best germination rate by culture period

Analysis (2.2): distribute the best results from (2.1) by price range and filter the most cost-effective ones

Analysis (2.3): define the characteristics of the most cost-effective media

Stage 3: Propagation

Analysis (3.1): number of media for the best Propagation index, highest shoots' high and lowest price range

Observed: the propagation index is proportional to the shoot's high

Remark: Could be done with radar graphics

Stage 4: Rooting

The rooting training table includes media parameters of germination process, following the biotechnological process, but also following the criteria for cost-effectiveness:

- germination_percentage $> 60\%$

- Germination_medium_price < 1

- Cultivation days ≥ 20

Analysis (4.1): number of media combinations (i.e. germination + rooting medium) for the best rooting rate by cultivation days.

Analysis (4.2): distribute the best results from (4.1) by price range and filter the most cost-effective ones.

Analysis (4.3): define the characteristics of the most cost-effective media combinations.

4. Results and discussion.

4.1. ANN related analysis. *Analysis of Significance and Sensitivity of Data*—the Scatter Plot graphic method was used instead of standard column graphics. This allowed us to analyze the relationship between Significance and Sensitivity. The main tendencies are:

(a) Seeds' Cleaning: Type of seeds—with a high degree of significance, but major changes are needed to make a difference in the forecast (Figure 5).
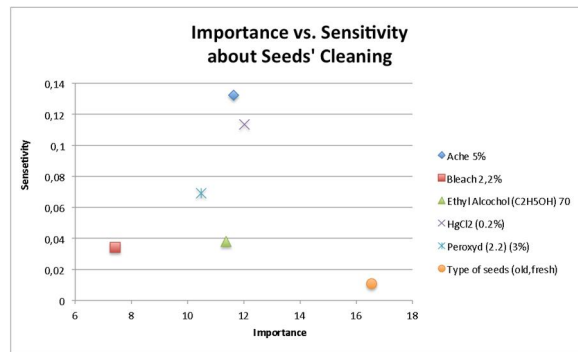
Fig. 5. I/S analysis about Seeds' Cleaning

(b) Explants' Cleaning—The type of the explant is highly predictive of the outcome of the forecast. It has both a high degree of significance and a high degree of sensitivity, which means that even minimal changes would significantly affect the outcome (Figure 6).
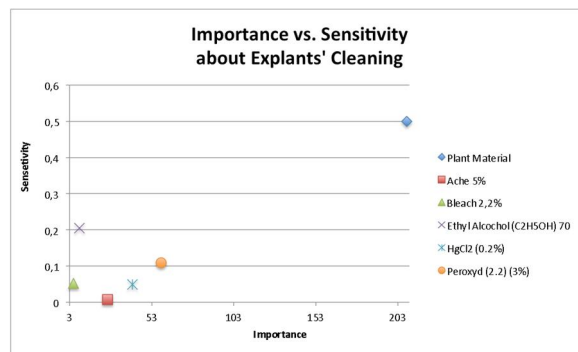
Fig. 6. I/S analysis about Explants' Cleaning

(c) Germination process—GA3 has similar characteristics as well as the type of explant in the process of cleaning of explants (Fig. 7).
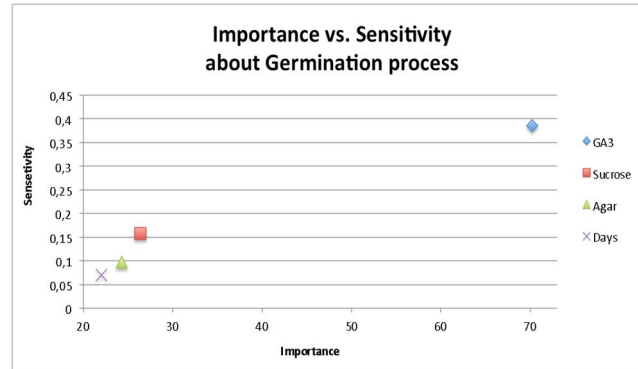
Fig. 7. I/S analysis about Germination process

(d) Propagation—Medium_Koeficient (i. e., type of the nutrient medium) and BAP, have a strong impact on the predicted results (Figure 8).
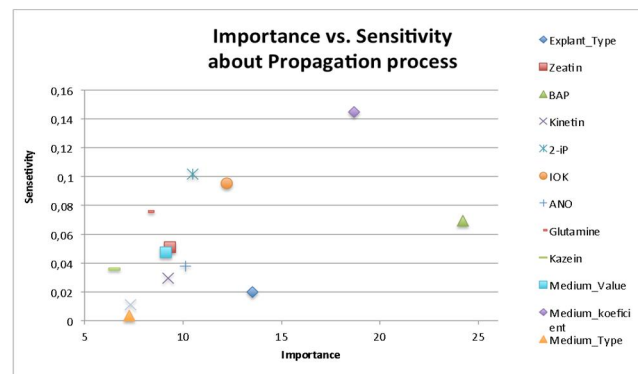
Fig. 8. I/S analysis about Propagation process

(e) Rooting—Sucrose as a growth factor is likely to have a significant impact on predicted results. Similar to the type of explant in the process of cleaning of explants (Figure 9).
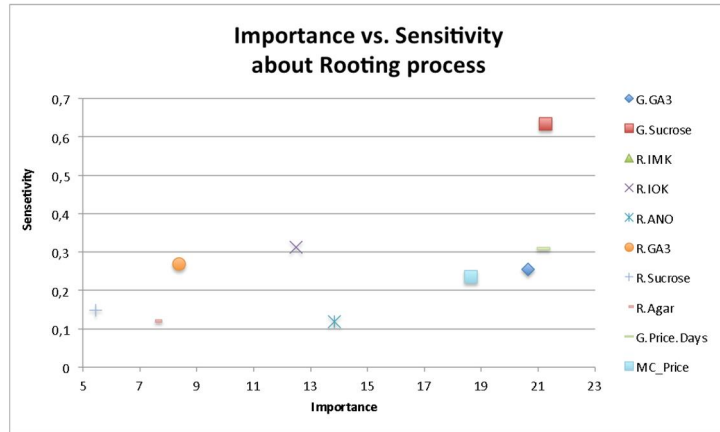
Fig. 9. I/S analysis about Rooting process

4.2. Analysis of the forecasted results.

Stage 1: Cleaning

Stage 1.A: Forecasting the results of applying different cleaning schemes on seeds

The forecasting shows between 99 and 100 % cleaning success, which does not depend on the applied scheme of treatments.

In this situation, it is important what we have received for the development of the seed germination process and here the results are not unambiguous.

High levels of seed development after the cleaning process are achieved exclusively in seed of the same year (101 test environments of 212) and with only 5 media for seeds one year old. An average good result for old seeds is achieved by 36 cleaning schemes at development levels of [8-15)% (see Figure 10).

In the **analysis of old seeds** (see Figure 11), three groups of schemes according to the linear trend can be distinguished:

- Long cleaning with low developmental effect
- Increasing the total number of minutes of treatment leads to increased development indicators

- Short-term cleaning accompanied by high results of subsequent development. It is this group of schemes that is of interest from a biotechnological point of view.
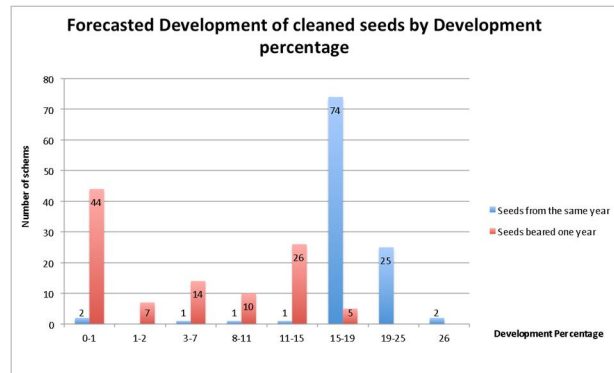
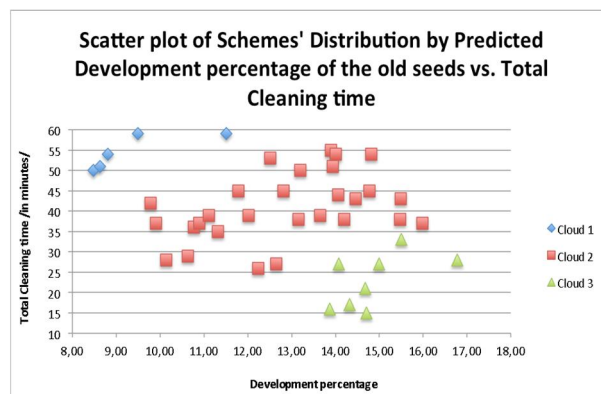Fig. 10. Forecasted Development of cleaned seeds by Development percentage

Fig. 11. Old seeds' Distribution of schemes

Fresh seed analysis. 8 clusters with grouped cleaning schemes are observed, of which 4 are of interest (Fig. 12). The maximum total treatment time (TTT) is 33 minutes, and the minimum development rate (DR) is 19.

Cluster 6 schemes (Cloud 6 in the figure, 13 schemas in total) predominate, with TTT in the interval [20; 30] minutes and DR is in the range [19; 21, 47].
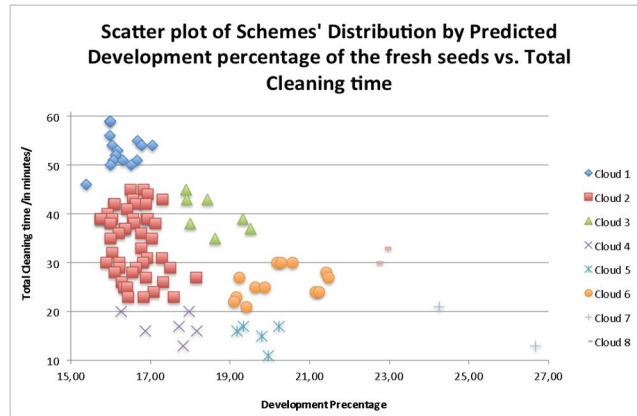
Fig. 12. Fresh seeds' Distribution of schemes

Stage 1.B: Forecasting the results of applying different cleaning schemes to different explants

Out of a total of 512 cleaning schemes, 417 have a very low% of subsequent explant development (between 0 and 24), including all schemes affecting adventive buds, root and stem segments, 1/3 of buds, and 79% root buds. There are a total of 26 schema types of explant buds and root buds that fall within the [24-90] % explant development interval. The result, which is of direct interest, is the interval [90-100] % subsequent development, which is distributed among the explants buds and root buds in a ratio of 64:5 (see Figure 13).
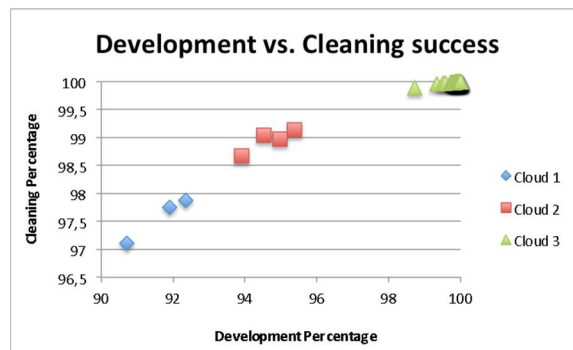
Fig. 13. Analysis of success for Cleaning and Development processes

For all 69 schemes, approximately 99-100% of the cleaning success rate is typical. From the distribution (see Figure 14), it is clear that three groups are distinguished. They are distributed almost evenly in time intervals, concentrating a larger number of cleaning schemes:
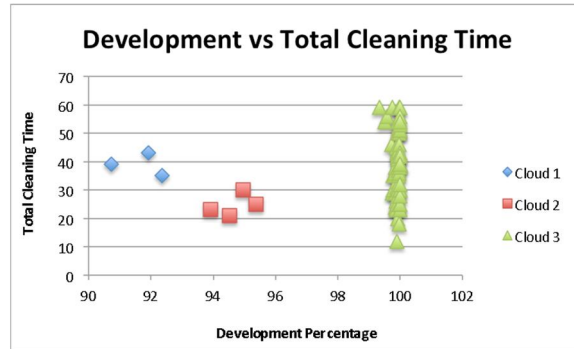

Fig. 14. Relationship between Total cleaning time and Development success

[23-30] total cleaning time /18 schemes in total / 4 schemes in Cloud 2/ 14 schemes in Cloud 3;

[35-39] total cleaning time /14 schemes in total/ 2 schemes in Cloud 1/ 12 schemes in Cloud 3;

[42-43] total cleaning time / 7 schemes in total/ 1 scheme in Cloud 1/ 6 schemes of Cloud 3.

As we are interested in short timed schemes we are not interested in Total cleaning time greater than 43.

The development process is directly related to the quality of cleaning. Experiment data also includes information on cases where the cleaning scheme has led to the destruction of the explants' development capability. Therefore, when referring to qualitative cleaning it is meant that it must be simultaneously with high rates of microorganism cleansing and preserving the vitality of the explant. Therefore, the development of the explant as measured in percent is an important consequence.

Detecting a correlation between very short treatment with ACE and the formation of Cloud 1 and Cloud 2; increased peroxide treatment time, and increased bleaching time. In the last two cases this means over 60% of the time

for implementation of the scheme for cleaning. It can be said that schemes where cleaning agents are below the critical minimum or above the critical maximum will cause deviations from the general trend and will not be effective.

Stage 2: Germination

The analysis of prognostic data for the germination process was carried out in the following directions:

1. Analysis of the distribution of the media depending on the days of cultivation and the theoretical (predicted) germination rate (see Figure 15): the following general trends are observed—1) in a 15-day culture period, the number of media increases with increasing the percentage of germination until the 60% germination threshold is reached, then a reverse trend begins; 2) in the 20-day culture period, observations are similar, but with the following
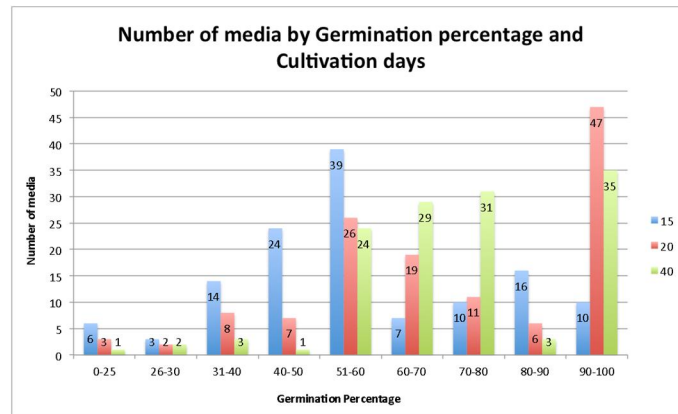

Fig. 15. Media distribution by Germination percentage and Cultivation days

anomaly: a sharp increase in the number of media at a germination rate of 90-100% is noted; 3) in the 40-days of cultivation, stable low levels of media were observed which maintained up to 50% germination followed by a sharp increase in the number of media corresponding to a higher percentage of germination; the following anomaly is observed—at a germination rate of 80-90, the number of environments guaranteeing these results drops sharply. The mediums that suggest a 20- and 40-day cultivation period, with> 60% germination success, are those that focus on further research.
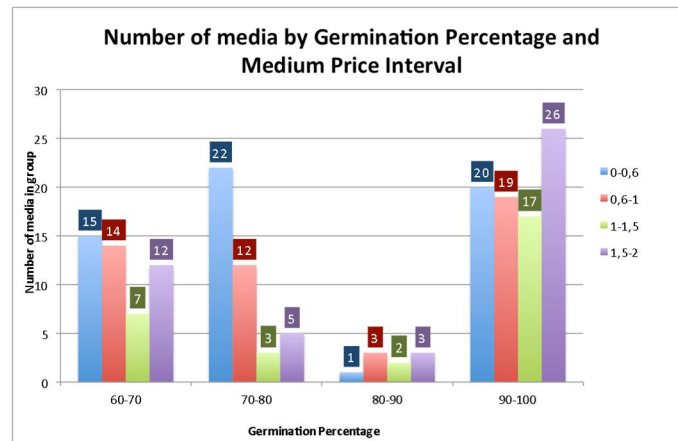
Fig. 16. Media distribution by Germination Percentage and Medium Price Range

2. Price range of the preferred media—analysis of the distribution of the germination medium shows that many of them are found in the lowest price class, 58 media out of total 181, about 0.6 euro/liter (see Figure 16). With respect to their distribution in relation to the percentage of germination, the media are characterized by uniformity, except for the interval [80-90], which is an abnormality of all the analyzed attributes. Therefore, the environments with the following parameters will be determined as acceptable:

- Germination_Percentage > 60 %

- Germination_Medium_Price < 1

- Cultivation days ≥ 20

    These media have the following characteristics:

- GA3 [3;77]—This factor has a uniform distribution

- Sucrose [0;30]—in an analysis of the effect of this factor on the germination process it was found that at levels ≥ 18, 39 media out of 71 would achieve a germination rate of ≥ 90%.

- Agar [0;6]—mostly with values 3 and 4

- Cultivation days [20,40]—prevailing media with a 40-day germination period (62 environments out of a total of 108)

Stage 3: Propagation

The propagation process has been reviewed for different types of explants.

In the apical bud's type of explant, smaller variations in the size of germinated seeds are observed, unlike those of the stem segment. In both cases, large variations have been observed at low propagation coefficient. Stability is observed at a breeding index of about 2; in cost of the medium around 0; at seed germ size: between 5 and 6 for apical buds and about 3 for stem segment. The trend for both types of explants is that with an increase in the breeding index, the cost of the nutrient medium decreases, as for apical buds the best results are for Pro_Index in [3; 4] and Medium_Value in [0; 2,5]; as for stem segment the best results are for Pro_Index in [3,8; 4] and Medium_Value in [0; 1,8]. The main conclusion that can be drawn at this stage of the study is that it is preferable for breeding to place explants from the stem segment in breeding media at a cost of between [0; 1,8].

The next step of the analysis will determine what these environments are characterized by:

- General parameters: Sucrose = 30, Agar = 7, IMC, TDZ, 2.4 D = 0. Exceptions are only MSD, MSD1 and MSD4

- Typical are the low concentrations of zeatin, BAP, kinetin, 2-ip, IEC, as significant variations in concentrations are observed in GA3, Glutamine and Casein.

The group of mediums zero concentrations of BAP, kinetin, 2-ip, ANS, GA3, Glutamine, in Casein = 1000 and low concentrations of zeatin and IEC also gives good results. This means that experimentation has the potential to substitute one plant hormone for another, or, in the presence of small quantities, to achieve a qualitative result, both biological and resource-constrained. All environments are type 0, i.e. "shoot apical meristem formation medium" type.

Stage 4: Rooting

In the training routing table created for the rooting process, the parameters of germinating media are also involved based on the in vitro experiment procedure. Therefore, it was necessary to choose the germination

culture media to form the query rows to the trained network. The distributions are shown on Figure 17 and Figure 18. They were determined by setting the following criteria:

- Germination_Percentage >60 %
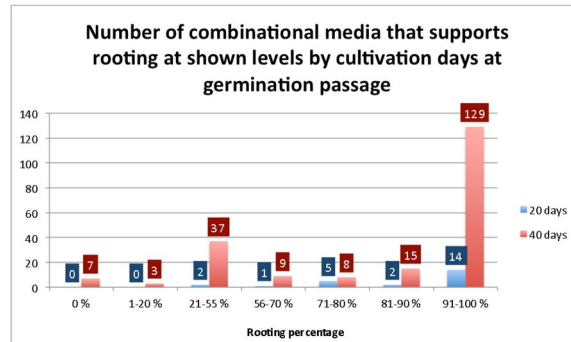- Germination_Medium_Price <1
- Cultivation days ≥ 20

Fig. 17. Media combination analysis

When examining the theoretical predictive results, two types of comparison were made: 1) examination of the dependence: the number of combined media ~ (the cultivation days of stage germination and rooting percentage); 2) after getting the results from 1) it was clear that the maximum number of media that produce high rooting results is assured with percentage
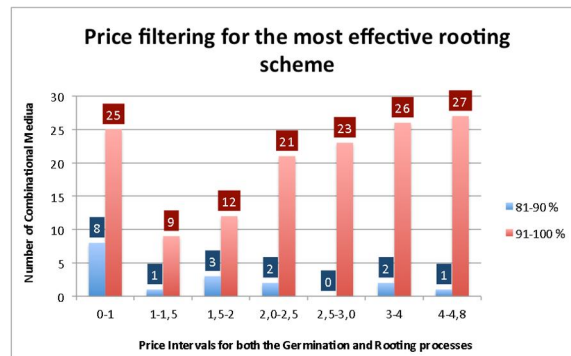

Fig. 18. Cost-effective analysis of Media combinations

level of rooting above 80%. It was estimated a price distribution for the set of media. The analysis shows that out of a total of 160 combinations of media with more than 80% success under limited resources conditions, only 43 of them are in the price range of [0-1,5].

These 43 media combinations have the following general characteristics:

- Days germination = 40

- ANO (for the rooting medium) almost always = 0

- IOK (for the rooting medium) varies between [0; 0.3], with zero concentrations prevailing

- IMK (for the rooting medium) varies between [0; 2], with uniform distribution

- GA3 (for the rooting medium) varies between [0; 1], with zero concentrations prevailing

- Sucrose (for the rooting medium) varies between [0; 20], as a predominant concentration of 10 units

- Agar (for the rooting medium) varies between [4; 8], with uniform distribution

- GA3 (for the germination medium) varies between [3; 22], as a predominant concentration of 10 units

- Sucrose (for the germination medium) varies between [3; 27], as a predominant concentration of above 18 units

5. Conclusions. The presented methodology and results are original and unique, as there is no known combination of biotechnological and bioinformatics methods in the study of Rhodiola Rosea plants that is used to obtain simultaneously cost-effective, time-effective and bio-result-effective procedures. The results can be treated as a theoretical guideline for determining the optimal nutrient media and combinations thereof so that they are effective both from a biotechnological point of view and from an economic point of view. The approach can also be applied to the study of other medicinal plants in order to establish effective biotechnological schemes for growth and rooting that are also cost-effective. Possible perspectives are the

creation of neural networks, taking into account the species and the ecotype, in order to build a more comprehensive system based on already available experimental data.

REFERENCES

[1] Egelund J., M. Skjøt, N. Geshi, P. Ulvskov, B. L. Petersen. A Complementary Bioinformatics Approach to Identify Potential Plant Cell Wall Glycosyltransferase-Encoding Genes. *Plant Physiol.*, **136** (2004), No 1, 2609–2620.

[2] Microsoft. MS Excel. `https://products.office.com/en-us/excel`, 5 November 2018.

[3] Neural Planner Software. EasyNN Plus. `http://www.npsnn.com`, 5 November 2018.

[4] Nikolova N. Computer modeling of chemical structures. PhD Thesis, Technical University, Sofia, 2002. (in Bulgarian)

[5] Su Y. H., Y. B. Liu, X. S. Zhang. Auxin-cytokinin interaction regulates meristem development. *Mol. Plant.*, **4** (2011), No 4, 616–625.

[6] Tasheva K., G. Kosturkova. Rhodiola rosea L. in vitro cultures peculiarities. In: Proceedings of the 3rd International Symposium "New Researches in Biotechnology", Bucharest, Romania, 2010, 103–111.

[7] Tasheva K., G. Kosturkova. Bulgarian golden root in vitro cultures for micropropagation and reintroduction. *Cent. Eur. J. Biol.*, **5** (2010), No 6, 853–863.

[8] TASHEVA K., G. KOSTURKOVA. Establishment of callus cultures of Rhodiola rosea Bulgarian ecotype. *Acta Horticulturae*, **955** (2012), 129–135.

[9] TASHEVA K., G. KOSTURKOVA. The role of biotechnology for conservation and biologically active substances production of Rhodiola rosea: Endangered medicinal species. *Sci. World J.*, **2012**, Article ID 274942, 1–13.

Valeriya Simeonova
e-mail: simeonova@fmi.uni-sofia.bg
Dimitar Vassilev
e-mail: dimitar.vassilev@fmi.uni-sofia.bg
Faculty of Mathematics and Informatics
St. Kliment Ohridski University of Sofia
5, James Bourchier Blvd
1164 Sofia, Bulgaria

Tasheva Krassimira
e-mail: krasitasheva@yahoo.com
Kosturkova Georgina
e-mail: gkosturkova@yahoo.com
Institute of Plant Physiology and Genetics
Bulgarian Academy of Science
21, Acad. G. Bonchev Str.
1113 Sofia, Bulgaria