17th European Symposium on Computer Aided Process Engineering – ESCAPE17
V. Plesu and P.S. Agachi (Editors)

Evaluation of sunflower collection by genetic variability based on germination and plantlet development parameters using Artificial Neural Networks

Dorina Bratfalean[a], Mircea Vasile Cristea[a], Paul Ş. Agachi[a], Dan Florin Irimie[a], Ahmad Sarrafi[b], Michel Petitprez[b]

[a]*Faculty of Chemistry and Chemical Engineering, Str. Arany Janos 11, RO 400028 Cluj-Napoca, Romania dorinet63@yahoo.com*
[b]*INP-ENSAT, pôle de Biotechnologie végétale Laboratoire de Biotechnologie et Amélioration des Plantese le chemin de Borde Rouge 31326 Castanet-Tolosan, Toulouse, France*

Abstract

The aim of the research work was to study the genetic variability of sunflower (*Heliantus annuus L.*) seeds collection. From 1500 available genotypes a set of 120 recombinant inbred lines coming from different countries were selected to perform the study. The genetic variability consisted in two types of experiments: *in vivo* and *in vitro* cultures. The recombinant inbred lines were classified using Artificial Neuronal Networks (ANNs). A first ANN was designed according to genotype origin and variety. Results confirm the ability of the ANN to predict the genotype origin and variety of the sunflower lines of seeds, based on the germination and plantlet development parameters. A second ANN was successfully designed and tested to classify the category of germination plantlet. Classification was performed, for the test set of data, and the results show a very good accuracy.

Keywords: Sunflower seeds, genetic variability, germination, plantlet development, Artificial Neural Networks.

1. Introduction

Sunflower seeds are considered to be an important oilseed crop due to their rich oil composition and high nutrition quality [1]. They represent an important source of raw materials required for industrial purposes, both for human or animal food and nonfood applications. Sunflower is considered to be a drought tolerant crop as its root has multiple branches and extracts water from depths not reached by other crops. In the future it is expected that sunflower will be grown in more arid areas of the world and this process is predicted to accelerate in the next ten years [2, 3]. This is the reason of searching and developing new methodologies that allow more flexible and better controlled means for classification of the crops' properties.
The field of ANNs has a history of five decades but is still developing rapidly. Founded on an idealized model of the biological neuron, the calculation paradigm of ANN is able to represent information on complex systems [4]. The main benefits of the ANN approach consist in their remarkable ability of classification and generalization [5, 6]. The aim of this study was to evaluate the sunflower collection using ANNs [7]. The new approach for the genetic variability prediction based on Artificial Neural Networks increasingly carries out the task of processing the large amount of scientific information and evaluation of the sunflower collection.

2. Importance of the research work

Germination and plantlet development are main characters for genetically selecting plants, depending on the agronomical character importance. This selection has a large contribution to ameliorative work, in genetic analysis and yield level of cultivars in order to obtain new hybrids. The genetic variability was investigated by two types of experiments: *in vivo* and *in vitro* cultures. Selection of the biological material for ameliorative purposes may be performed by the help of the classification aptitude the ANN owns. The ANN may produce efficient classification for cases when the intrinsic relationship between parameters is complex, as it is the case of sunflower seeds. As a result, for each type of experiment (*in vivo* and *in vitro*) two pairs of ANNs have been designed. The first designed ANNs have been used for predicting the origin and variety of the sunflower collection based on the plantlet developement biological parameters. Irrespective of *in vivo* or *in vitro* culture prediction, the input of ANNs for predicting the origin and variety is the same. The second designed ANNs have been used for classifying the category of plantlet germination. Irrespective of *in vivo* or *in vitro* culture classification, the input of ANNs for classifying the category is the same.

3. ANN Classification

3.1. Methodology

Probabilistic ANNs have been used for classification. They consist in two layers of neurons. The first layer is a radial basis layer and the second one is a competitive layer. The radial basis layer computes distances from the input vector to the training input vectors and produces a vector whose elements measure how close the particular input is to a training input. The competitive layer cumulates these contributions for each class of inputs to produce, as its net output, a vector of probabilities. The second layer uses a compete transfer function for selecting the maximum of these probabilities and generates a vector having as elements a *1* (one) for that class and *0* (zeros) for the other classes. The probabilistic ANN is guaranteed to converge to a Bayesian classifier [7].

3.2. Experimental arrangement

The experiments have been carried out during 2002-2003 at the BAP-ENSAT Laboratory France. The genetic variability consisted in two types of experiments: *in vivo* and *in vitro* cultures. The first experiment (*in vivo*) was carried out selecting one hundred genotypes, in a ten seeds set for each variety. Seeds were treated with calcium hypochlorite (5%) and rinsed three times in sterile water. They were placed into compost (vermiculite) and maintained in the culture room under the light conditions of 16h light/8h dark cycle, at 25°C temperature. Seeds were observed twice a day in the morning and evening, in order to check germination kinetic and germination rate. After 10 days the plantlets were assisted with nutritional solution for development. Following germination, after 15 days, plantlets were checked by observing a set of biological parameters. The second experiment *(in vitro*) was carried out with a selection of sixty genotypes, in a ten seeds set of each variety. *In vitro* culture seeds were treated with calcium hypochlorite (5%) and rinsed three times in sterile water, and then placed on filter paper in a Petri dish, under the light conditions of 16h light/8h dark cycle, at 25°C temperature. Seeds were observed twice a day in the morning and in the evening in order to check germination kinetics and germination rate. When the germination stopped, they were transferred into the vermiculite. After 15 days, plantlets were checked by observing a new set of biological parameters. Fresh tissue for DNA extraction was sampled for each experiment. All experiments were reiterated in three steps. DNA extraction technique used a kit nucleon (RPN 8510), and dosage of DNA was performed by the spectophotometric technique.

3.3. Result & discussions

For the case of the *in vivo* experiments, the ANN was trained using a set of 77 input data (seeds fresh weight, percentage of germination, hypocotyls length, number of leaves, amount of DNA). A testing set of 77 data was given to the ANN in order to test the ability of the trained ANN for classifying new input data (not seen at training). The classification results are presented in Fig. 1.

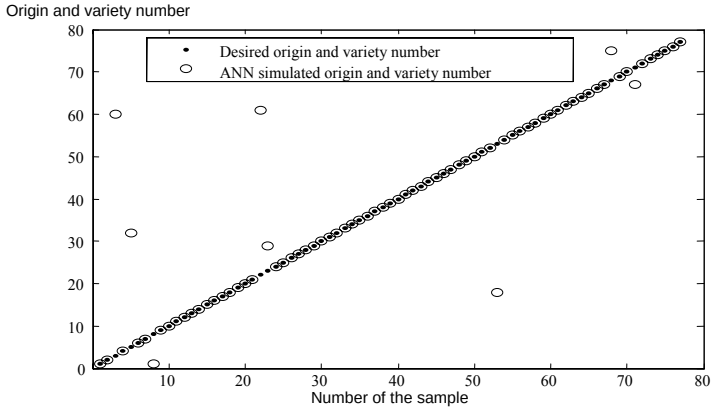

Figure1: Prediction results of the origin and variety for the *in vivo* testing set of data.

Results of Fig. 1 show about 90% accuracy of the trained ANN for predicting the origin and variety number.

A second classification of the *in vivo* experiments was performed by the ANN for dividing the germination behaviour into three groups: group 1 features germination percentage between 0-30%, group 2 features germination percentage between 30-70% and group 3 features germination percentage between 70-100%. The ANN was trained using a set of 77 input data (seeds fresh weight, hypocotyls length, number of leaves, amount of DNA). A testing set of 77 data was given to the ANN in order to test the ability of the trained ANN for classifying new input data. Fig. 2 presents the classification results of the second trained ANN and they reveal good accuracy of the trained ANN for classifying in three classes the germination behaviour, with no error.

For the case of the *in vitro* experiments, the ANN was trained using a set of 51 input data (seeds fresh weight, percentage of germination, hypocotyls length, number of leaves, amount of DNA). A testing set of 51 data was given to the ANN in order to test the ability of the trained ANN for classifying new input data (not seen at training). The classification results, presented in Fig. 3, show about 85% accuracy of the trained ANN for predicting the origin and variety number.

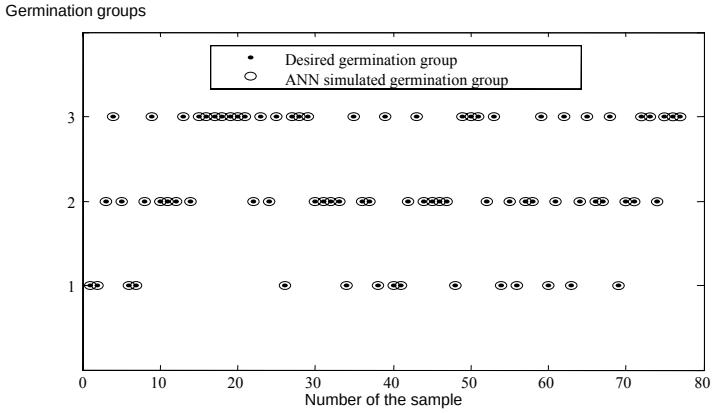

Figure 2: Classification results of germination groups for the *in vivo* testing set of data.

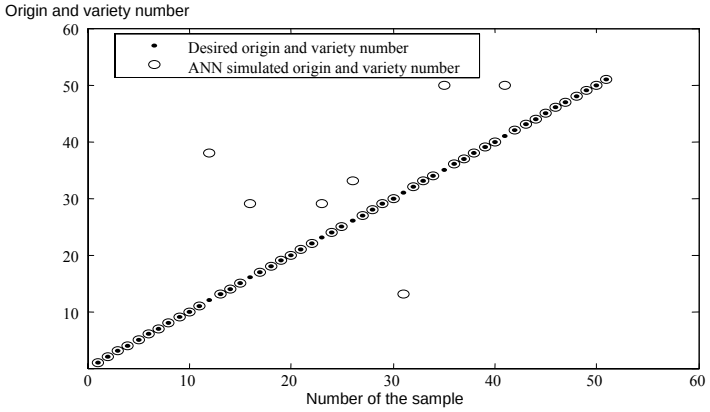

Figure 3: Prediction results of the origin and variety for the *in vitro* testing set of data.

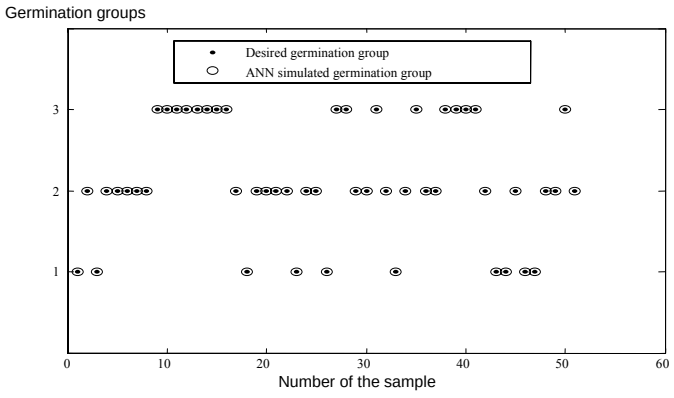

Figure 4: Classification results of germination groups for the *in vitro* testing set of data.

The second classification of the *in vitro* experiments was performed by the ANN for dividing the germination behaviour into the same three groups: 0-30%, 30-70% and 70-100% germination groups. The ANN was trained using a set of 51 input data (seeds fresh weight, hypocotyls length, number of leaves, amount of DNA). A testing set of 51 data was given to the ANN in order to test the ability of the trained ANN for classifying new input data. The classification results, presented in Fig. 4, show again error free classification capability of the designed and trained ANN.

4. Conclusions

Origin and variety prediction of the designed ANNs is good, taking into consideration the limited number of parameters considered as ANN inputs. Further improvement may be obtained by extending the number of input parameters and by carefully filtering the training set of data. Nevertheless, the germination group classification capability of the designed ANNs is very good and this may be very useful in ameliorative applications. Starting from the methodology of designing ANNs, for the prediction and classification of genetic variability of sunflower seeds germination, further classification procedures may be developed for similar applications. Their incentives consist in reducing experiment time and costs, bringing out hidden aspects revealed by the data mining techniques.

Acknowledgements

The authors acknowledge for the support offered by the QLK3-CT-2001-600670 Marie Curie fellowship and by the CEEX 180/2006 MOLPAT-PCa project.

References

1. A.A. Schneiter and J.F. Miller, Crop Science, 21 (1981) 901.
2. J. Roche, A. Essahat, A. Bouniols, M. El Asri, Z. Mouloungui, M. Mondiès and M. Alghoum Helia, 27 (2004) 73.
3. K. Kýlýç¸ I.H. Boyacý and H.K. Kusmenogçl, Journal of Food Engineering, 78 (2007) 897–904.
4. S. Haykin, Neural Networks A Comprehensive Foundation, Macmillan Publishing Company, Englewood Cliffs, NJ, 1994.
5. M.V. Cristea, S. Varvara, L. Muresan and I.C. Popescu, Indian Journal of Chemistry, 42A (2003) 764.
6. V.M. Cristea and I. Batiu, Revue Roumaine de Chimie, 50 (2005) 1009.
7. Matlab, Neural Networks Toolbox, MathWorks, 2006.