Studying the relationships between nutrients in pistachio leaves and its yield using hybrid GA-ANN model-based feature selection

Behrooz Pourmohammadali[a], Mohammad Hassan Salehi[a], Seyed Javad Hosseinifard[b,*], Isa Esfandiarpour Boroujeni[c], Hossein Shirani[c]

[a] Department of Soil Science, Shahrekord University, P.O. Box 115, Shahrekord, Iran
[b] Pistachio Research Center, Horticultural Sciences Research Institute, Agricultural Research Education and Extension Organization (AREEO), Rafsanjan, Iran
[c] Department of Soil Science, Vali-e-Asr University of Rafsanjan, P.O. Box 518, Rafsanjan, Iran

ABSTRACT

Sustainable and reliable management requires special attention to factors affecting crop yield. In the present study, a hybrid model of genetic algorithm and artificial neural network (GA-ANN) was employed to recognize the importance of nutrients in pistachio yield. One hundred seventy-five points in different pistachio orchards of Rafsanjan and Anar regions, Kerman province, the southeast of Iran, were identified and selected for leaf sampling and yield measurement. The concentration of phosphorus (P), potassium (K), iron (Fe), zinc (Zn), copper (Cu), manganese (Mn), calcium (Ca) and magnesium (Mg) was determined. The hybrid GA-ANN model was implemented in MATLAB software, after statistical analysis and multivariate regression modeling. The results showed that the correlation and linear multiple regression analysis could not justify the variations of pistachio yield in relation to leaves' nutrients concentration. The lowest error of the hybrid GA-ANN model was observed by five features including concentrations of K, Mg, Fe, Zn and Cu. Sensitivity analysis of ANN indicated that the highest relative importance for predicting pistachio yield was related to Cu (34.6%), K (28.2%) and Fe (26.1%). The GA-ANN model was able to solve complex and multi-dimensional problems. The accurate and careful interpretation of the results, obtained from this approach can provide a good insight for optimum farm management planning.

1. Introduction

Pistachio is one of the most important strategic and non-petroleum export products of Iran. Generally speaking, maintaining Iran's share in the world pistachio market largely depends on the quantitative and qualitative development of its production. According to FAO (2014), Iran with more than 50% of the world's pistachio production, has been placed at the first rank. However, in terms of pistachio yield per unit area, Iran is positioned after the countries such as China, the United States and Greece. Environmental stresses and improper management approaches can be considered as most important reasons of facing with insufficient pistachio yield and increasing the cost of its production in Iran. Thus, both researchers and farmers have focused on optimum management as a key strategy to achieve acceptable yield per unit area and efficient and justifiable soil and water resources' usage. Sustainable and reliable management requires special attention to the factors affecting crop yield. Despite suitable weather conditions for the growth of pistachio trees in Rafsanjan region, factors such as high soil pH, salinity

and the soil calcium carbonate cause nutritional problems and as a result crop yield reduction in pistachio orchards (Eskandari et al., 2014). However, certain nutrient availability in the soils is not assumed as the only factor affecting its uptake by the plant. Other factors such as soil moisture and plant diseases may significantly affect nutrients' uptake. Nutrient concentration in a plant is not only affected by its availability in the soil, but also by many other factors including other nutrients' concentration, nutrients' competition and mobility in the plant. Therefore, leaves' analysis reflects both soil nutrients' availability and plant's nutritional status (Beede et al., 2005).

Interactions often cannot be perfectly explained by conventional linear methods due to dynamism, complexity, uncertainty and multi-dimensionality of agricultural production. So, robust and supervised multivariate data analysis techniques are required to determine the importance of each particular variable. In this regard, feature selection can be used as a process which chooses an optimal subset of features according to a certain criterion (Liu and Motoda, 2012). Feature selection gives a deeper view of a prediction driving force and also