Development of Models for Discovery of Novel Proteins Associated with Photosynthesis using Machine Learning Approaches

Monendra Grover[1], U.B. Angadi[2] and Naveen Sharma[3]

[1,2,3]*Centre for Agricultural Bioinformatics, ICAR-IASRI, Pusa Campus, New Delhi -110012*

Abstract

Photosynthesis is the most important process in the biosphere and is responsible for its sustenance. With the increase in number of genome sequencing projects a lot of novel genes are expected to be discovered. The existing methods based on sequence similarity may miss some novel and important genes coding for proteins associated with Photosynthesis. In this work we report the development of models for discovery of novel proteins associated with photosynthesis using machine learning approaches based on post translational modification patterns and physico-chemical properties. In this study 45 variables were used for classification. As is evident from results out of 45, 41 variables used are important for classification. Rest of the four variables namely GRAVY index, Tyrosine nitration type B, Half life and S-Nitrosylation type C were not found to significantly contribute to the classification.

International Conference Agriculture, Food Engineering and Environmental Sciences- Sustainable Approaches