World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:14, No:12, 2020

Comparison of Different Artificial Intelligence-Based Protein Secondary Structure Prediction Methods

Authors : Jamerson Felipe Pereira Lima, Jeane Cecília Bezerra de Melo

Abstract : The difficulty and cost related to obtaining of protein tertiary structure information through experimental methods, such as X-ray crystallography or NMR spectroscopy, helped raising the development of computational methods to do so. An approach used in these last is prediction of tridimensional structure based in the residue chain, however, this has been proved an NP-hard problem, due to the complexity of this process, explained by the Levinthal paradox. An alternative solution is the prediction of intermediary structures, such as the secondary structure of the protein. Artificial Intelligence methods, such as Bayesian statistics, artificial neural networks (ANN), support vector machines (SVM), among others, were used to predict protein secondary structure. Due to its good results, artificial neural networks have been used as a standard method to predict protein secondary structure. Recent published methods that use this technique, in general, achieved a Q3 accuracy between 75% and 83%, whereas the theoretical accuracy limit for protein prediction is 88%. Alternatively, to achieve better results, support vector machines prediction methods have been developed. The statistical evaluation of methods that use different AI techniques, such as ANNs and SVMs, for example, is not a trivial problem, since different training sets, validation techniques, as well as other variables can influence the behavior of a prediction method. In this study, we propose a prediction method based on artificial neural networks, which is then compared with a selected SVM method. The chosen SVM protein secondary structure prediction method is the one proposed by Huang in his work Extracting Physico chemical Features to Predict Protein Secondary Structure (2013). The developed ANN method has the same training and testing process that was used by Huang to validate his method, which comprises the use of the CB513 protein data set and three-fold cross-validation, so that the comparative analysis of the results can be made comparing directly the statistical results of each method.

Keywords : artificial neural networks, protein secondary structure, protein structure prediction, support vector machines
Conference Title : ICSRD 2020 : International Conference on Scientific Research and Development
Conference Location : Chicago, United States
Conference Dates : December 12-13, 2020