Prediction of protein subcellular localization with oversampling approach and Chou's general PseAAC

Shengli Zhang , Xin Duan

PII: S0022-5193(17)30499-X
DOI: 10.1016/j.jtbi.2017.10.030
Reference: YJTBI 9254

To appear in: *Journal of Theoretical Biology*

Received date: 5 August 2017
Revised date: 29 September 2017
Accepted date: 27 October 2017

Please cite this article as: Shengli Zhang , Xin Duan , Prediction of protein subcellular localization with oversampling approach and Chou's general PseAAC, *Journal of Theoretical Biology* (2017), doi: 10.1016/j.jtbi.2017.10.030

Highlights

- We combined oversampling method with SVM to deal with the protein subcellular localization of unbalanced data sets.
- Results of SVM by Jackknife tests show that oversampling methods have successfully decrease the imbalance of data sets.
- The excellent overall accuracy indicates that the feature representation and selection capture useful information of protein sequence.