

Research on some classification methods and their application in classifying Sumoylated Protein data

Silivong Tony

Abstract: This study investigates several classification methods and their application to the classification of protein SUMOylation data. SUMOylation is a critical post-translational modification involved in various cellular processes, and its accurate identification is essential for understanding disease mechanisms. The research evaluates the performance of multiple machine learning algorithms, including Support Vector Machines, Random Forests, and deep learning models, on benchmark datasets of SUMOylation sites. Feature extraction techniques, such as amino acid composition, position-specific scoring matrices, and evolutionary information, are employed to represent protein sequences. Comparative analysis demonstrates that ensemble methods and deep learning architectures achieve superior predictive accuracy and robustness compared to traditional classifiers. The findings highlight the importance of feature selection and model optimization in handling imbalanced biological data. This work provides a comprehensive framework for computational SUMOylation site prediction, offering a valuable tool for proteomics research and facilitating the discovery of potential therapeutic targets.

Keywords: data classification, Protein Sumoylation, classification method, application