

Comparison of Existing Predictor and Development of Computational Method for S- Palmitoylation Site Identification in Arabidopsis Thaliana

Authors : Ayesha Sanjana Kawser Parsha

Abstract : S-acylation is an irreversible bond in which cysteine residues are linked to fatty acids palmitate (74%) or stearate (22%), either at the COOH or NH₂ terminal, via a thioester linkage. There are several experimental methods that can be used to identify the S-palmitoylation site; however, since they require a lot of time, computational methods are becoming increasingly necessary. There aren't many predictors, however, that can locate S- palmitoylation sites in Arabidopsis Thaliana with sufficient accuracy. This research is based on the importance of building a better prediction tool. To identify the type of machine learning algorithm that predicts this site more accurately for the experimental dataset, several prediction tools were examined in this research, including the GPS PALM 6.0, pCysMod, GPS LIPID 1.0, CSS PALM 4.0, and NBA PALM. These analyses were conducted by constructing the receiver operating characteristics plot and the area under the curve score. An AI-driven deep learning-based prediction tool has been developed utilizing the analysis and three sequence-based input data, such as the amino acid composition, binary encoding profile, and autocorrelation features. The model was developed using five layers, two activation functions, associated parameters, and hyperparameters. The model was built using various combinations of features, and after training and validation, it performed better when all the features were present while using the experimental dataset for 8 and 10-fold cross-validations. While testing the model with unseen and new data, such as the GPS PALM 6.0 plant and pCysMod mouse, the model performed better, and the area under the curve score was near 1. It can be demonstrated that this model outperforms the prior tools in predicting the S- palmitoylation site in the experimental data set by comparing the area under curve score of 10-fold cross-validation of the new model with the established tools' area under curve score with their respective training sets. The objective of this study is to develop a prediction tool for Arabidopsis Thaliana that is more accurate than current tools, as measured by the area under the curve score. Plant food production and immunological treatment targets can both be managed by utilizing this method to forecast S- palmitoylation sites.

Keywords : S- palmitoylation, ROC PLOT, area under the curve, cross- validation score

Conference Title : ICBCBBE 2023 : International Conference on Bioinformatics, Computational Biology and Biomedical Engineering

Conference Location : Zurich, Switzerland

Conference Dates : September 11-12, 2023