

Improving the efficiency of post-translational modification protein analysis based on combining machine learning models and natural language processing.

Tran Thi Xuan

Abstract: This study focuses on improving the efficiency of analyzing protein post-translational modifications (PTMs), a key factor in understanding biological functions and disease mechanisms. The main approach combines machine learning models and natural language processing (NLP) techniques to exploit the features of amino acid sequences as a biological language. The research applies Transformer architecture and deep learning to accurately predict modification sites on a large data scale. Expected results will optimize accuracy compared to traditional methods, while saving experimental costs and time. The project contributes an important technological solution to bioinformatics, supporting drug discovery and the development of personalized medicine.

Keywords: post-translational modification, machine learning, natural language processing, bioinformatics, Transformer model