# LbiPlantUbi

**LbiPlantUbi:** the prediction of plant ubiquitination sites based on a hybrid deep learning model and word embedding in NLP. The model automatically learns features directly from raw protein sequences and integrates prediction into a single deep learning architecture.

# Requirement

* Keras
* Numpy
* Sklearn
* Pandas

**Dataset**

This process yielded 7000 protein fragments from the plant data subset, with 3500 positive and 3500 negative fragments selected randomly.

* The independent dataset: 1500 sequences were chosen at random from the total fragments (750 positive and 750 negative)
* The training set: 5500 fragments (2750 positive and 2750 negative).

**Train model:** We use Google colab pro buid this model

* Cross validation: CV\_LBiPlantUbi.ipynb
* Train model: Model\_LBiPlantUbi.ipynb.

Model saved and named: LBiPlantUbi.h5. You can use LBiPlantUbi.h5 in the Model folder for predict and independent test.

* Independent test and predict: LBiPlantUBi\_ID.ipynb

# Contact

Please feel free to contact us if you need any help: [nvnui@ictu.edu.vn](mailto:nvnui@ictu.edu.vn)