# CBiLSuccSite

**CBiLSuccSite:** A novel approach for the prediction of succinylation sites based on a hybrid deep learning model and word embedding features of proteins. 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**

In order to prepare training and testing datasets for the work, the dataset of 2322 protein sequences has been randomly separated into two sets:

* Training set: 2192 protein sequences (4755 positive, 4755 Negative).
* Testing set: 124 proteins sequences (254 positive, 254 negative).

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

* Cross validation: CV\_CBILSuccSite.ipynb
* Train model: Train\_CBILSuccSite.ipynb.

Model saved and named: Suci\_CNN\_BiLSTM.h5. You can use Suci\_CNN\_BiLSTM.h5 in the Model folder for predict and independent test.

* Independent test and predict: CBILSuccSite\_predict.ipynb

# Contact

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