**Operational Instruction For MPwave Algorithm**

This algorithm has been developed to distinguish the Moonlighting Protein (MPS) from non-Moonlighting Protein (non-MPS). We have used four different python classes to implement and evaluate the model by 5-fold cross-validation and independent dataset tests. We have provided two practical datasets used in the article, and users could easily reproduce the result.

**Essential python libraries for proposed algorithm :**

1. numpy

2. threading

3. sklearn

4. pandas

5. scipy.special

6. csv

**Dataset 1 :**

* Moonlighting proteins\_Dataset1.fasta
* non\_Moonlighting Proteins\_dataset\_1\_2.fasta

**Dataset 2 :**

* Moonlighting proteins\_Dataset2.fasta
* non\_Moonlighting Proteins\_dataset\_1\_2.fasta

**Python files list (in order)**

For reproducing the results, the following python's files should be run in a row.

* integrate\_test\_train\_data.py
* transform\_Protein\_sequence\_to\_6D\_Signal.py
* feature\_generation\_from\_all\_wavelet\_filter\_banks.py
* Eight\_ML\_Method\_Classification\_and\_evaluation.py

**For reproducing and test software :**

* Copy the MPs and non-MPs to the same place as the projects.
* Copy the full name of MPs and non-MPs data file in "integrate\_test\_train\_data.py" as "f1" and "f2" input parameters. for example :

f1 = f1 = open(**"Moonlighting proteins\_Dataset1.fasta"** , **'r'**)

f2 = open(**"non\_Moonlighting Proteins\_dataset\_1\_2.fasta"** , **'r'**)

* Run " integrate\_test\_train\_data.py." After running this script, three folders would be generated. the folder names are as below
  + Training\_data: This folder would use to store the results of 6D signals during the running of classes. in the first run, " Train\_Data.txt" would generate which includes all the MPs and non-MPs sequences, and in the next step, "Feature\_Results\_for\_Train.txt" would generate which contains the 6D Signals which uses for DWT analysis
  + Filter\_banks\_feature\_vectors: This folder uses to save the DWT output.
  + Result: All the evaluation results would be saved in this folder.
* Run " transform\_Protein\_sequence\_to\_6D\_Signal.py," which leads to generating the 6-D signals. All the data will be saved in a text file by the name of " Feature\_Results\_for\_Train.txt" in the "Training\_data" folder. This file will use for the next step of feature extraction by DWT.
* Run " feature\_generation\_from\_all\_wavelet\_filter\_banks.py " to generate 31 different filter banks' features. each of which filter banks would be saved in a separate file in the " Filter\_banks\_feature\_vectors" folder. The filter bank's type could be recognized by the name of the text file for instant " sym8Coeffs\_Avrage\_features\_for\_Training\_Data.txt" indicates that this file belongs to sym8 filter bank.
* Run Eight\_ML\_Method\_Classification\_and\_evaluation.py in order to reproduce the results and evaluate the model with independent dataset test and 5- fold cross-validation test. A more precise description is available in the material and methods of the article. All the results would be stored in the result folder.