# User Manual

The ACP-MLC is a software program written in the Python package Tkinter (version 8.6.12), which can be used for the prediction of anticancer peptides (ACPs). We used two machine learning algorithms in this study to build predicted models. For ACP-MLC, the 1st-level predictor was a random forest-based model for distinguishing ACPs from non-ACPs, and the 2nd-level predictor was a binary relevance-based model for predicting the targeted tissue types. Researchers can input their in-house data with the software and obtain predicted lists. The help document is as below.

1. Open the software

Please decompress the package first, then find the "*ACP-MLC.exe*" file in the unzipped folder, and double-click on it.

2. Start to predict
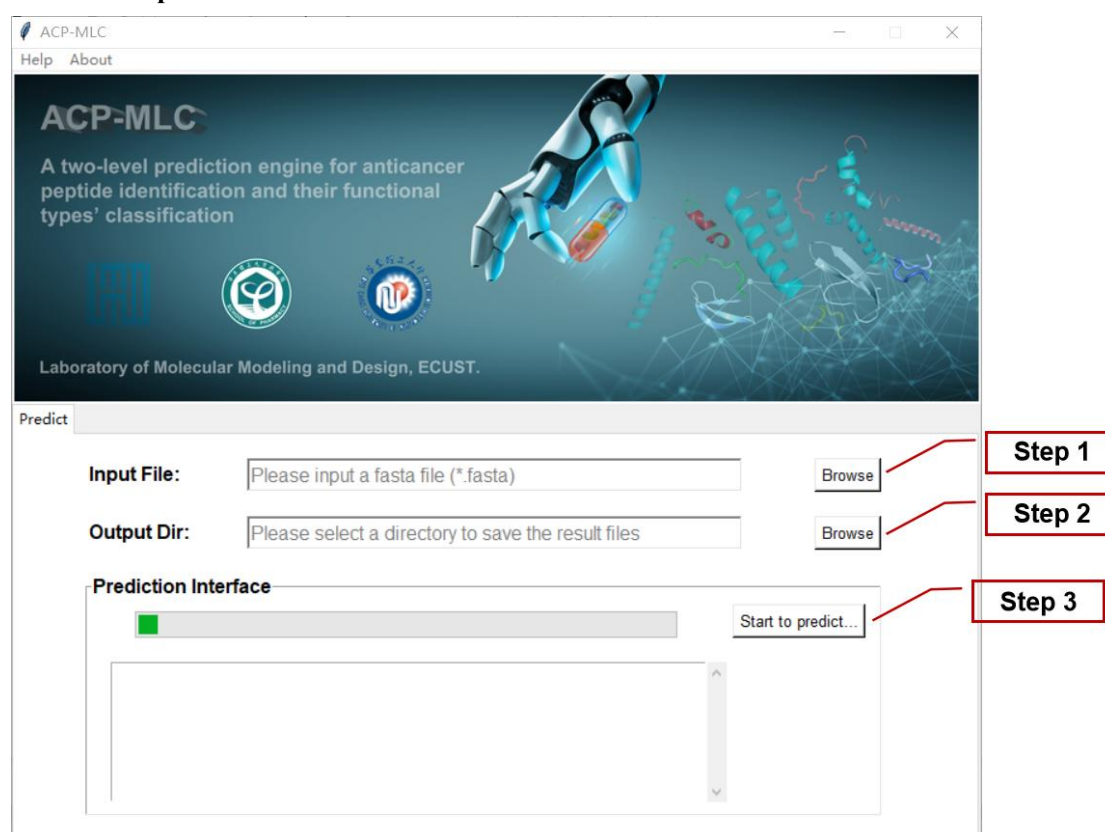
Step 1: Input sequences

Please click the "*Browse*" button to choose the input peptide sequences. Note that the input file should meet the following **requirements**: (i) The accepted input format is FASTA, which

means a sequence should begin with a single-line description with a greater-than (">") symbol at the beginning and followed by lines of sequence data. (ii) The input sequences can only contain natural amino acids.

An error message will appear if the input file does not meet the requirements. Meanwhile, the button "*Browse*" in step 2 and the button "*Start to predict…*" will turn dim and cannot be clicked.

Note: Due to the encoding strategy used in the 2nd-level model, the input sequence length prefers to be less than 38. If the length is greater than 38, ACP-MLC can only predict whether the query is an ACPs or not, but cannot predict its functional types. In this case, if you want to predict targeted tissue types, please use a sequence with a length longer than 38 amino acids.
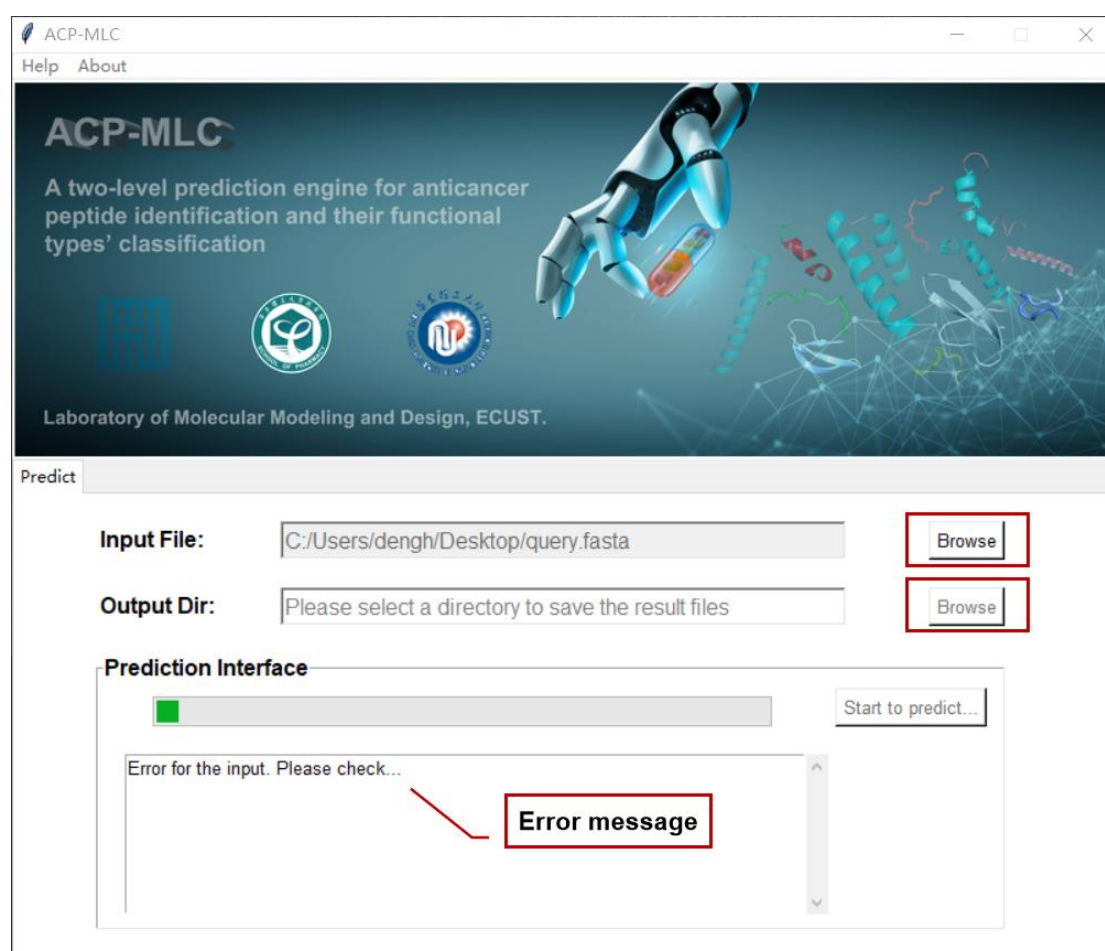

Step 2: Set the output directory

Please choose the directory for saving the predicted results by clicking the "*Browse*" button. If you forget this setting, a warning message of "*The save path of results is not found!*" will appear when you begin the next step.

Step 3: Star to predict

Please click the "*Start to predict…*" button to launch the prediction using your input sequences. This process requires several seconds. The progress bar will prompt the running status. When the prediction is complete, a warning message of "*Prediction finished!*" will appear.
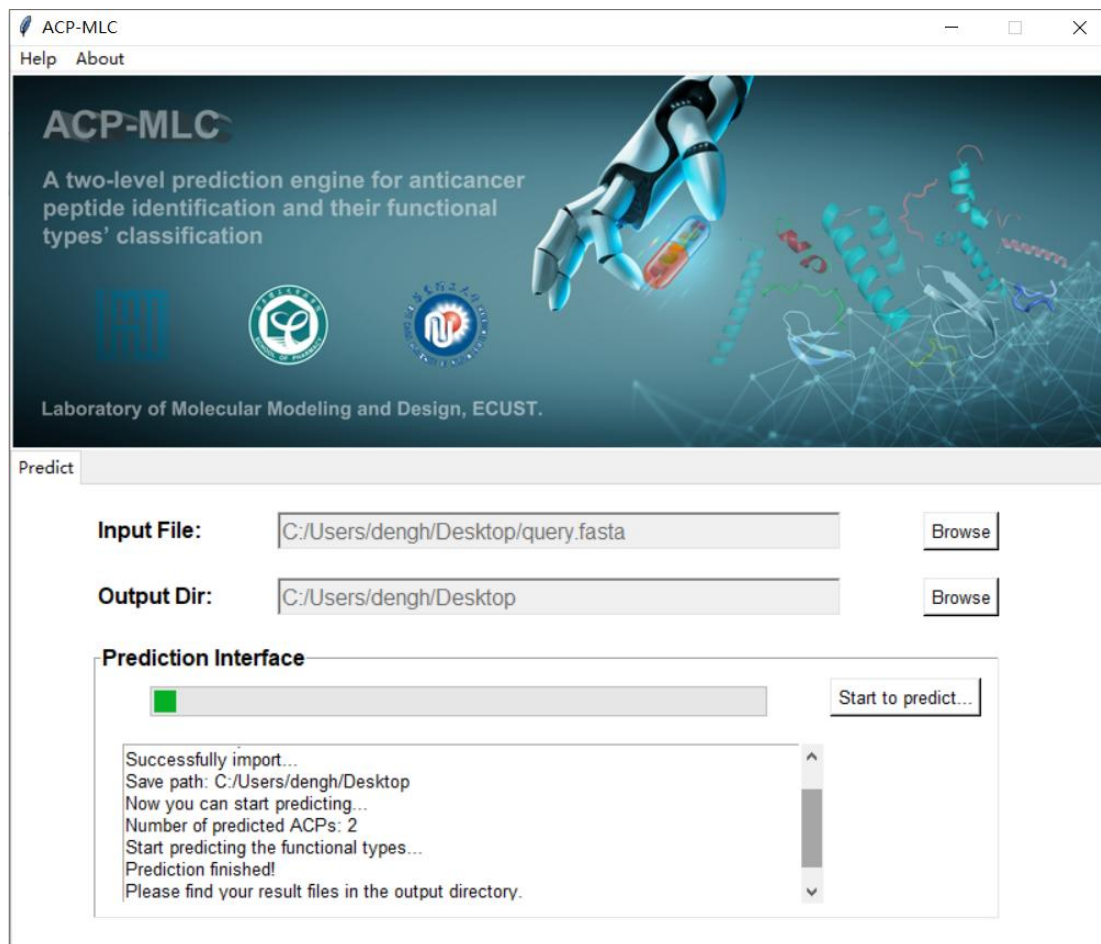
3. Example Results

Please find your results in the directory you set in step 2. Results from the 1st-level and the 2nd-level model will be stored separately in two files named after time. We provide both the predicted results and the predicted probabilities.

Note: As mentioned in step 1, only one result file predicted by the 1st-level model will be provided if the inputs contain sequences with lengths less than 38. Besides, only those predicted as ACPs will be fed into the 2nd-level model, and thus be presented the functional types.

Here is an example of outputs by ACP-MLC.

a) results from the 1st-level model

A	B	C	D
	ACP or non-ACP	Probability	
0	ACP	0.9625	
1	ACP	0.9916667	
2	non-ACP	0.4375	
3	non-ACP	0.2208333	

b) results from the 2nd-level model

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Colon	Breast	Cervix	Skin	Lung	Prostate	Blood	Colon_prob	Breast_prob	Cervix_prob	Skin_prob	Lung_prob	Prostate_prob	Blood_prob
0	1	1	0	1	1	1	1	0.915050319	0.924382586	0.052157636	0.95344027	0.970221764	0.959370893	0.964814986
1	1	1	1	1	1	1	0	0.905166447	0.929158308	0.975401895	0.958479972	0.985298671	0.949942182	0.061946722

For result (a), column A is the index of all input sequences, column B denotes the predicted results, and column C is the predicted probabilities. For result (b), column A is the index of sequences predicted as ACPs. Columns B-H represent the predicted results for each cancer type, where "1" denotes the peptide predicted to target the corresponding tissue type.

4. Developer Information

Maintainer: M.S. Ding, Dr. Deng, and Prof. Tang.

Contact US: ytang234@ecust.edu.cn

Laboratory of Molecular Modeling and Design