

Read-Across-v2.0

Overview: Read-Across-v2.0 is a tool which takes input from two different files for the training and test sets, respectively, and predicts the biological activity using different similarity based functions and also evaluates the quality of predictions in terms of different validation metrics like $q^2_{\text{ext_F1}}$, $q^2_{\text{ext_F2}}$ and rmsep, if the experimental biological activity data for the query compounds are available.

Input file specifications:

- The training and test set input files should have the extension (.xlsx).
- File format: Compound numbers (1st column), Descriptors (Subsequent columns) and Biological Activity (Last Column).

	A	B	C	D	E
1	S.No	X1	X2	Y	
2	1	2.2	5.7	2.32	
3	2	3.7	5.5	3.14	
4	3	6.1	6.1	3.2	
5	4	2.8	4.9	3.61	
6	5	4.2	5.9	2.79	
7					

Sample training set.xlsx

	A	B	C	D	E
1	S.No	X1	X2	Y	
2	6	3.8	4.11	2.11	
3	7	4.7	5.56	3.34	
4	8	5.9	6.1	3.1	
5	9	4.35	4.2	3.81	
6	10	3.98	5.78	2.79	
7					

Sample test set.xlsx

S.no = Compound numbers X1, X2 = Descriptors Y = Biological Activity

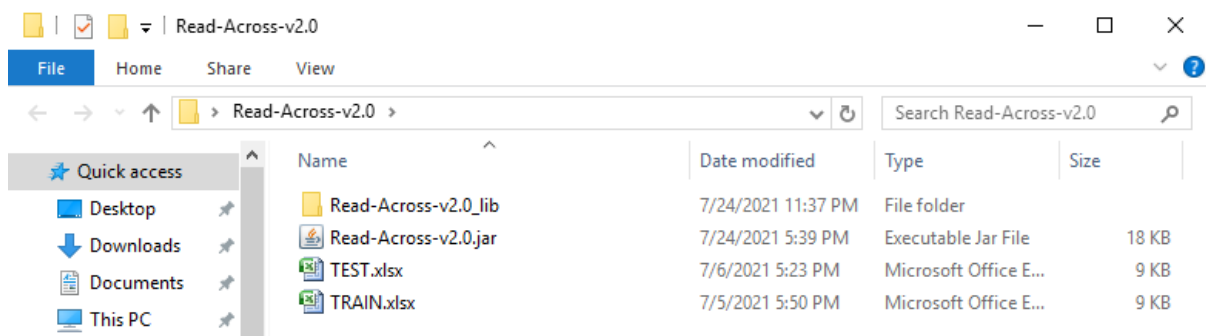
- **NOTE:** For calculation of only the biological activities without evaluating the quality of predictions, put “999” in the 1st observed biological activity value of the test set and any random entry for other compounds.

	A	B	C	D	E
1	S.No	X1	X2	Y	
2	6	3.8	4.11	999	
3	7	4.7	5.56	3.34	
4	8	5.9	6.1	3.1	
5	9	4.35	4.2	3.81	
6	10	3.98	5.78	2.79	
7					

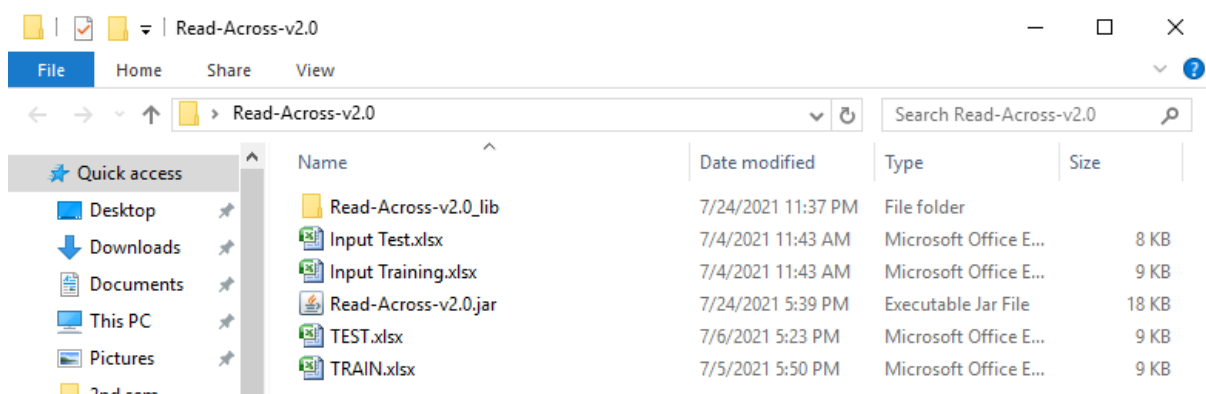
Sample test set.xlsx

To download and run the program:

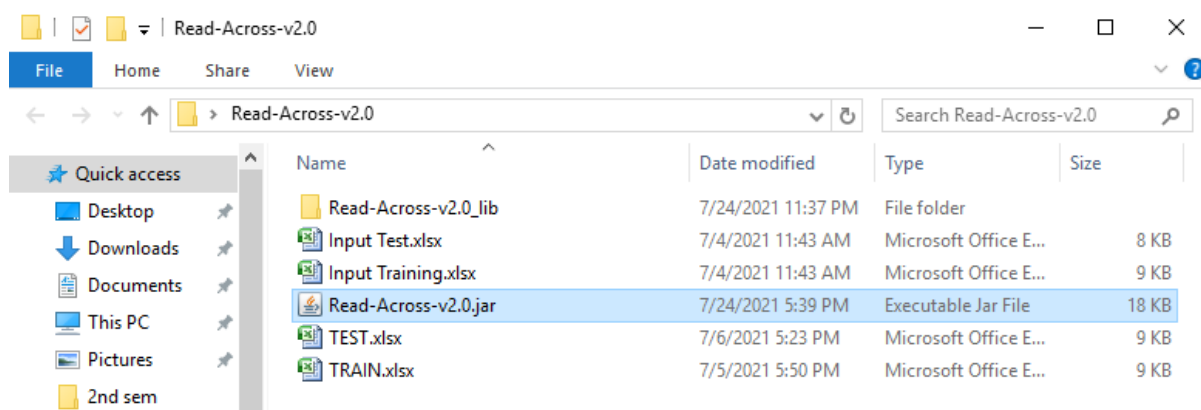
- Download the **.zip** file. Extract it. You will find that it consists of a folder (Read-Across-v2.0) inside which, there is a **.jar** file, a library folder, and two sample input files.



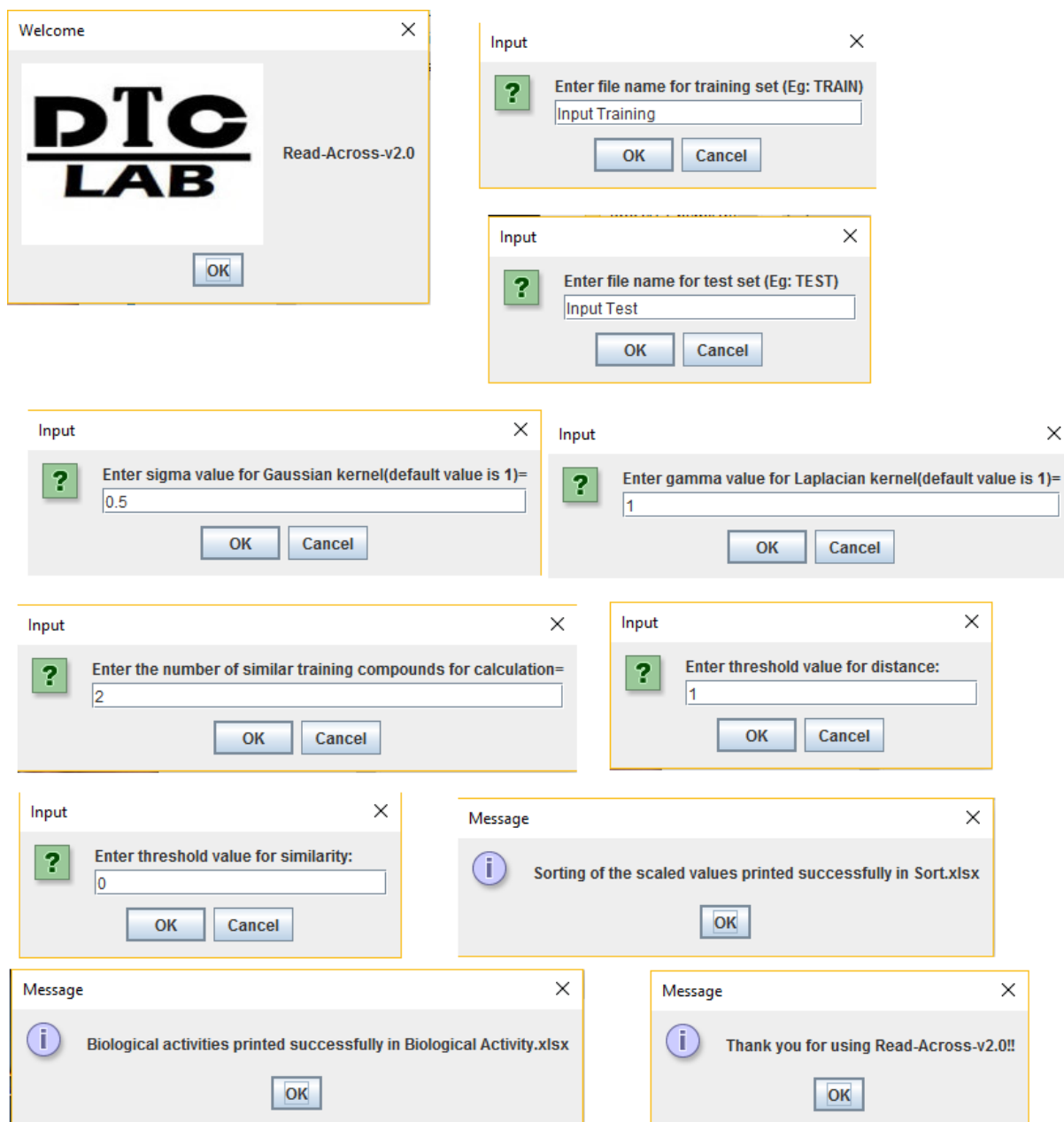
- Place your training and test set files inside this Read-Across-v2.0 folder i.e. the same folder which contains the **.jar** file.



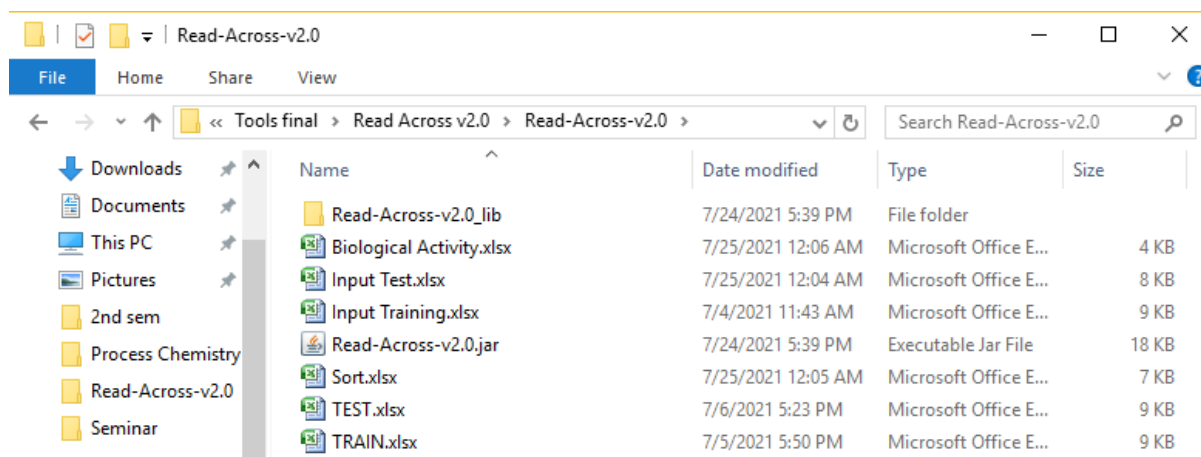
- Next, double click on the **Read-Across-v2.0.jar** file and the program will be executed.



- Enter the required data in the dialog boxes that appear. The user needs to enter file names for training and test sets, some constant values (sigma and gamma, suggested value being 1), the number of close training set compounds (which can range from 2-10), threshold values for distance (suggested value being 0.5; 1 in case of no threshold) and similarity (suggested value in the range of 0-0.05; 0 in case of no threshold).



- Sorting of the similarity measures will be automatically printed in a newly generated file namely **Sort.xlsx**, and the biological activities along with the validation metrics will be automatically printed in a newly generated file namely **Biological Activity.xlsx**. The user can see both these files generated in the same folder (Read-Across-v2.0).



Tool Developed by



In collaboration with



For further query, please contact

Prof. Kunal Roy

Drug Theoretics and Cheminformatics (DTC) Laboratory

Department of Pharmaceutical Technology

Jadavpur University

Kolkata 700 032

India

Email: kunal.roy@jadavpuruniversity.in

