**====== Instruction regarding datasets ======**

1. We used two benchmark datasets. One dataset is called train set and the other is known as test set. For fair comparison with existing methods, we employed both train and test datasets which were used by past studies [3-7](#_ENREF_3). These datasets are discussed in detail in section 2.1.
2. The 1224 positive samples of train set were downloaded from DrugBank database (<https://www.drugbank.ca/> ) [1](#_ENREF_1) .
3. The 1319 negative samples of the train set were collected from Swiss-Prot database (<https://academic.oup.com/nar/article/24/1/21/2360095>) [2](#_ENREF_2) .
4. The 224 positive and 237 negative samples of the test set were collected from DrugBank (<https://www.drugbank.ca/> ) [1](#_ENREF_1) .

**Citations:**

1 Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* **46**, D1074-D1082 (2018).

2 Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic acids research* **24**, 21-25 (1996).