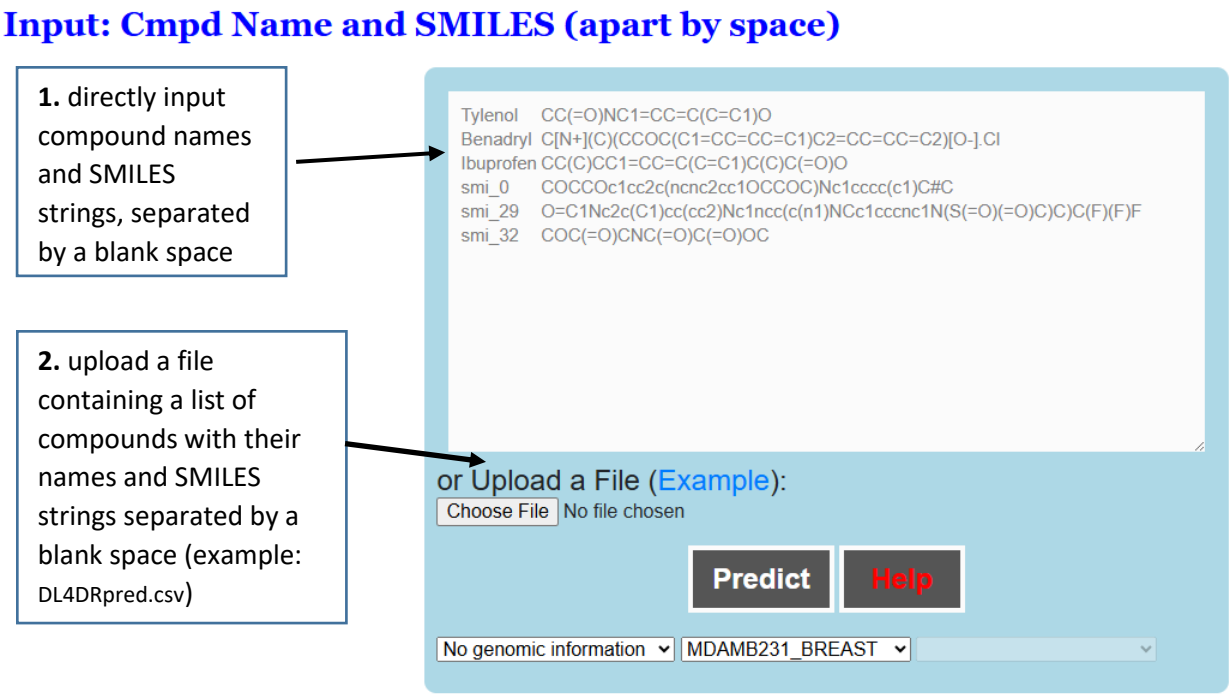
DL4DR Instruction

DL4DR is a user-friendly, easy-to-access web server that can be used to predict the cancer drug response using their SMILES as the only input. Cell genomics data is built-in information in our system. We developed **DL4DR** for the scientific community to screen compounds with good treatment efficacies for >1000 cancer cell line in an ultra-fast manner. The models were constructed based on a novel supervised Octave residual convolutional neural network (ORCNN) method with a new image-based representation scheme (**ChemImages**) to characterize compound structures. The web server is implemented using the Flask framework, with Python 3.8.10, Torch 1.11.0, and Flask 2.4.

1. Input to DL4DR

The input from users is simple and straightforward: just the names and SMILES strings of compounds to be predicted. There are two ways to input them: 1). Direct input in the online form, and 2). Upload a file containing a list of compounds with their names and SMILES strings, as shown below:

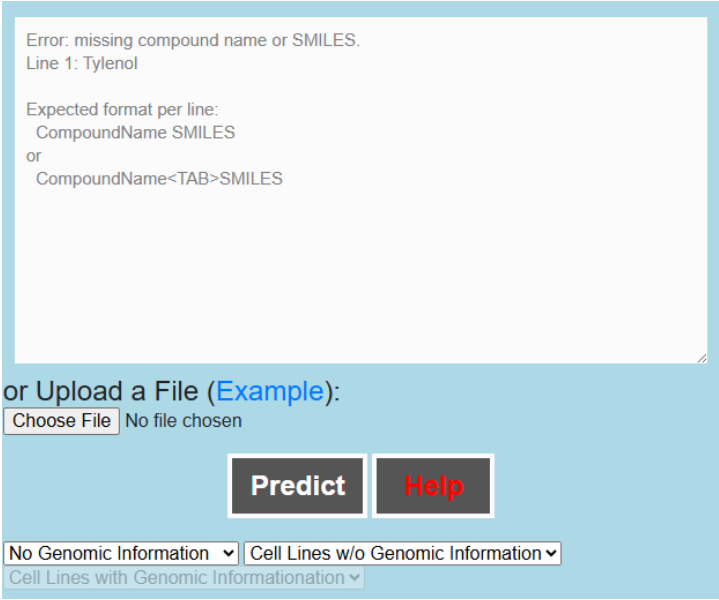


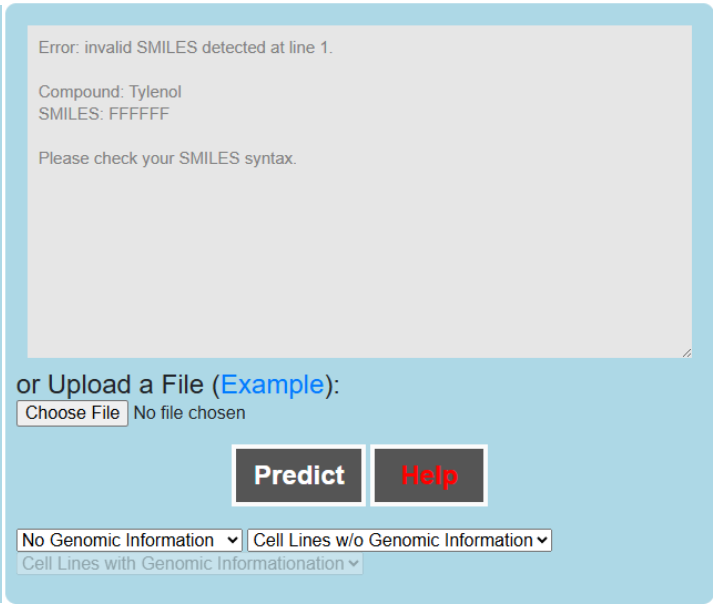
The input compounds must consist of names and SMILES strings separated by a blank space. Otherwise, the input will be considered as “illegal”. If this happens, an error message will show up in the input Form. Here are some examples with incorrect input:

1). No Name or SMILES

Phenol
Oc1ccccc1

An error message will show up to indicate there is either no name or SMILES (for either Phenol or Oc1ccccc1 [treated as a name] here).





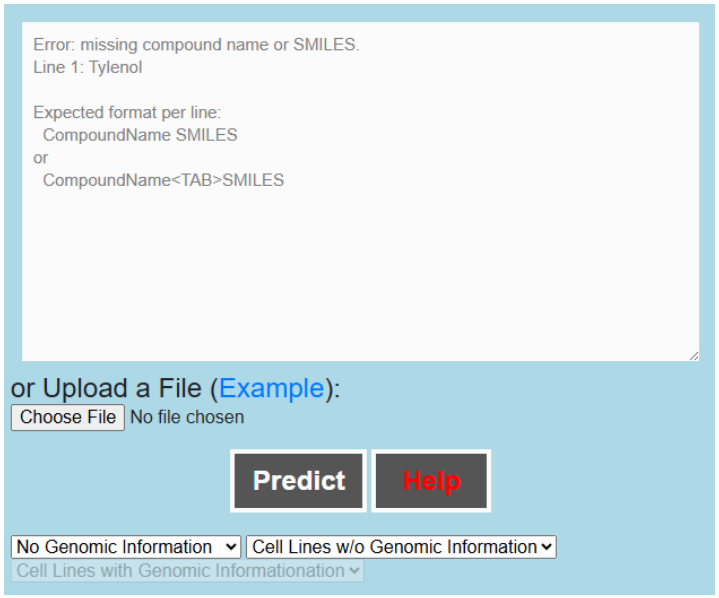
The correct format should be:

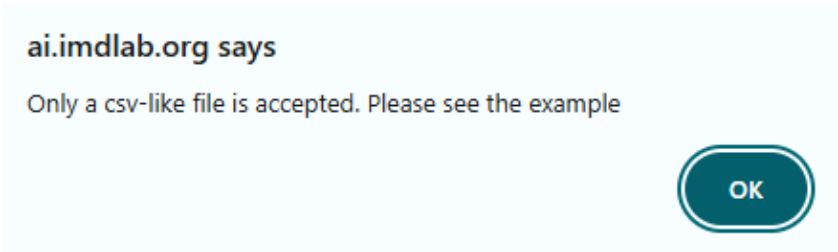
Phenol Oc1ccccc1 or

Phenol Oc1ccccc1

2). The uploading file extension is wrong

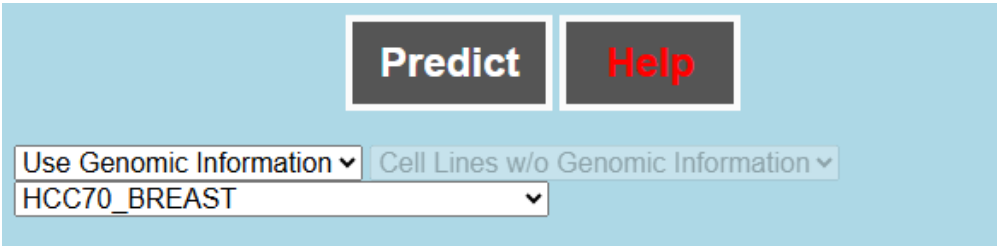
Of note, for safety purposes, the uploaded file must use .csv extension, although the compound name and SMILES are actually separated by a space. Otherwise, the following error will occur.





3). Other Options to Select

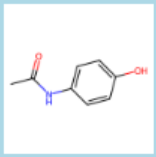
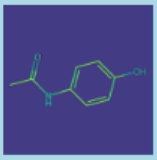
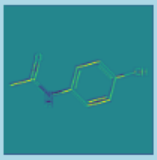
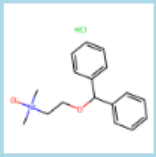
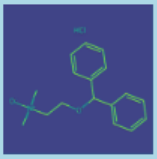
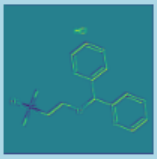
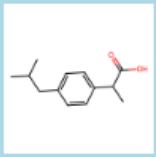
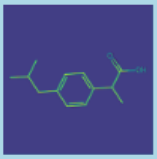
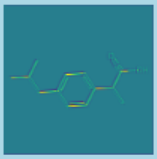
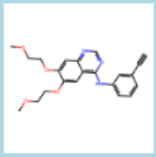
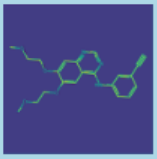
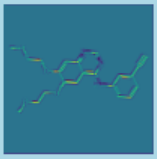
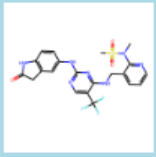
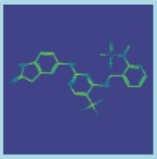
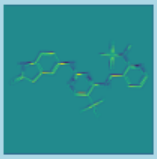
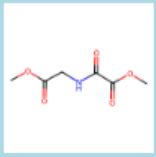
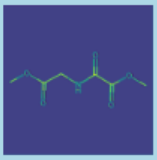
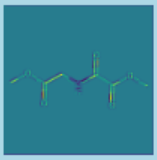
Currently there are two types of models. One type is individual cancer cell line models which were built for only one cell line and based on only chemical structures, like traditional QSAR models. The other type is built based on many cancer cell lines and this model can be used to predict the activities of the same compound for all of those cancer cell lines based on



their genomic information. Therefore, both chemical structures and cell line genomic data were used to build the model. Users have the option to using or not using genomic information to make predictions. The corresponding option will be disabled (greyed out) based on users' selection.

2. Output of DL4DR

The output of **DL4DR** is also easy to understand, with 5 columns in a table. The first column includes the compound names, as provided by users in the input; The 2nd is for predicted bioactivities. The 3rd column is to show the chemical structures of the input compounds. The 4th and 5th columns are for high frequency and low frequency feature maps, respectively. Please refer to our manuscript for more details. The below shows some examples of predictions.

Name	IC_{50}	Chem_Structure	High_Freq_Feature	Low_Freq_Feature
Tylenol	0.35			
Benadryl	0.07			
Ibuprofen	0.94			
smi_0	-0.0			
smi_29	0.29			
smi_32	0.63			

In addition to visualizing the result in a table, we also dynamically provide users with an option to download the data (the Download Data link). This is particularly useful if the input number of compounds is large (e.g., >1,000). The format of the downloadable table is as follow (basically the first two columns of the above table):

```
Tylenol         0.45
Benadryl        0.28
Ibuprofen       0.25
smi_0           0.33
smi_29          0.37
smi_32          0.49
```