# CAMDA Contest Challenge

# Data Analysis Procedure and Results: CMap Drug Safety Challenge

## 1. Data sets:

This challenge we the Connectivity Map (CMap) data generated by Broad Institute. This is the Affymetrix-based CMap dataset, referred to as “build 02,”. For this analysis participants received the gene expression dataset generated from two different cell lines (MCF7 and PC3). MCF7 cell line is the breast cancer cell line and PC3 cell line is prostate cancer cell line. Data from each cell line is divided into a training set and a validation set. CAMDA provided 2 sets of datasets: “training set” and “validation set”. Training set containsclinical DILI results as training labels for 190 drugs. For reporting the results and analysis, please use these exact names. You need to specify following details in the report:

1. For the “training set”, if the entire set was used, please state “the training set was not altered”. If only partial data from the “training set” used, please provide the complete list for the “training set”.
2. For the “validation set”, it should be used in the entirety.

## 2. Data Pre-processing:

Please describe the normalization and pre-processing methods used for the data Pre-processing.

## 3. Feature selection and cross-validation:

* If a feature section method used to narrow down a small set of genes for the model development, please provide description for the feature selection procedure used.
* If the feature selection and cross-validation is integrated, please describe as well.

### *4. Classification method:*

Please provide the information about classification method used for this analysis. Along with the parameter description. Detailed description of the classification method.

Please provide the “training set” results with these parameters:

To report Accuracy, Sensitivity, Specificity and Matthews correlation coefficient (MCC) please use following equation based on 2X2 confusion matrix

**Accuracy:**

 (1)

**Sensitivity:**

 (2)

**Specificity:**

 (3)

**Matthews correlation coefficient (MCC):**

 (4)

Where,

TP = True positive, TN = True negative, FP = False Positive and FN = False Negative

### *5. Prediction of the “Validation Set”:*

Please fill the form below with the prediction results from each compound from the **“Validation Set”.**

|  |  |  |
| --- | --- | --- |
| **Compound No. from Validation List** | **MCF7** | **PC3** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |