

Covid-19 automated diagnosis and risk assessment through Metabolomics and Machine Learning

by Delafiori et al., 2020

Datasets adopted in the paper

There are three folders, one for each model, in the article:

1. Cov19_Control_x_Covid
 - . Diagnosis Model
 - . Control(negative) x Covid-19(positive) patients;
2. Cov19_LowRisk_x_HighRisk
 - . Risk Assessment Model
 - . Covid19-Mild(negative) x Covid19-Severe(positive) patients
3. Cov19_Control_x_LowRisk
 - . Mild Discrimination Model
 - . Control(negative) x Covid19-Mild(positive) patients;

In each folder, there are 2 feature vector CSV files corresponding to:

1. FeatureVectors_<model>_Fitting.csv the fitting dataset used for training, validation and final test.
2. FeatureVectors_<model>_BlindTest.csv the blind test dataset.

In the feature vectors CSV files, the columns correspond to vectors (spectra) and lines to m/z intensity. The 6 first lines are vectors identification information, and the first column corresponds to the m/z labeling.

As example:

```
# vector name Covid19_01021_01 Covid19_01021_02 ... Covid19_00986_10
# individual Covid19_01021 Covid19_01021 ... Covid19_00986
# indiv number 1 1 ... 801
# obsv number 1 2 ... 10
# categ number 5 5 ... 24
# class 0 0 ... 1
200.013 0 0 ... 0
...
200.173 145175 185876 ... 0
...
958.64 8742643 12213819 ... 2247116
...
```

Then **185876** is the intensity of m/z **200.173** for the replicate (obsv) 2 of patient **Covid19_01021**, who is of class negative (0) and belongs to the group of samples labelled as category 5.

The patients are grouped according to their demographics and condition in 26 groups called here categories.

# categ number	# Condition
1	Control
2	Control
3	Control
4	Control
5	Control
6	Control
7	Control
8	Covid-19_Mild
9	Covid-19_Severe
10	Covid-19_Severe
11	Covid-19_Severe
12	Covid-19_Mild
13	Covid-19_Severe
14	Covid-19_Severe

```
15 Covid-19_Severe
16 Covid-19_Mild
17 Covid-19_Mild
18 Covid-19_Severe
19 Covid-19_Severe
20 Covid-19_Severe
21 Covid-19_Severe
22 Covid-19_Severe
23 Covid-19_Severe
24 Covid-19
25 Covid-19
26 Covid-19
```

Important: Feature vectors of blind tests do not have the class assigned (# class). Instead, results obtained with the blind tests should use the categories information.

* * *