The details of the directory and files in <https://github.com/beherasan/MS2Compound/benchmark> are provided below:

**Directories:**

1. *casmi2017\_pos:* MS/MS fragments for all the compounds (solution) of casmi (challenge dataset) predicted using positive model of CFM-id

2. *casmi2017\_neg:* MS/MS fragments for all the compounds (solution) of casmi 2017 (challenge dataset) predicted using negative model of CFM-id

3. *Positive:* Casmi challenge datasets acquired in positive mode

4. *Negative:* Casmi challenge datasets acquired in negative mode

**Files:**

1. *MonoIsotopicMass.txt*: Monoisotopic mass of the solution compounds for the challenges provided by casmi 2017

2. *PrecursorMass.txt*: Precursor m/z of the challenges provided by casmi 2017

3. *ScoreComparison\_Pos.pl*: Perl program for comparison of scores from positively acquired dataset

4. *ScoreComparison\_Neg.pl*: Perl program for comparison of scores from negatively acquired dataset

5. *PositiveScore0.05.txt*: Output from “*ScoreComparison\_Pos.pl”* perl program. This contains assignments of compounds to the positively acquired data. The dot product, *mS-*score, hypergeometric score, and fit-score are provided for each match.

6. *PositiveScoreRank0.05.txt*: Number of compounds assigned correctly for each rank. This contains the following columns, “Score, Rank, Number, CumulativeShared, Shared”.

-- Score: Name of the score, Dot product as DP, *mS-score* as mS, Hypergeometric score as HGS, and fit score as FS.

-- Rank: Rank for true positive hits

-- Number: Number of true positive hits at the given rank

-- CumulativeShared: Total number of false positive hits at the given rank

-- Shared: Total number of true positive hits shared with false positive hits at given rank

7. *NegativeScore0.05.txt*: Output from “*ScoreComparison\_Neg.pl”* perl program. This contains assignments of compounds to the negatively acquired data. The dot product, *mS-*score, hypergeometric score, and fit-score are provided for each match.

8. *NegativeScoreRank0.05.txt*: Number of compounds assigned correctly for each rank. This contains the following columns, “Score, Rank, Number, CumulativeShared, Shared”. The details of column names are same as explained for “*PositiveScoreRank0.05.txt*”.

9. *ScoreCompilation\_Pos4ROC.pl*: Calculation of True positive fraction, and false positive fraction from positively acquired data.

10. *ScoreCompilation\_Neg4ROC.pl*: Calculation of True positive fraction, and false positive fraction from negatively acquired data.

11. *TPR\_FPR\_Pos.txt*: Output from “*ScoreCompilation\_Pos4ROC.pl*” program

12. *TPR\_FPR\_Neg.txt*: Output from “*ScoreCompilation\_Neg4ROC.pl*” program

13. *CalculateAUC.pl*: Perl program to calculate AUC using trapezoidal rule