Deep Learning Based Metabolite Annotation

Hoi Yan Katharine Chau[1]
Hongyu Ao[1], Xinran Zhang[1], Shijinqiu Gao[1], Rency Varghese[1] and Habtom Ressom[1]
[1] Georgetown University

p.p1 {margin: 0.0px 0.0px 0.0px 0.0px; font: 13.0px 'Helvetica Neue'}

Metabolite annotation is a major bottleneck in untargeted metabolomics studies by liquid chromatography coupled with mass spectrometry (LC-MS). This is in part due to the limited publicly available spectral libraries, which consist of tandem mass spectrometry (MS/MS) data acquired from just a fraction of known compounds. Machine learning and deep learning methods provide the opportunity to predict molecular fingerprints based on MS/MS data. The predicted molecular fingerprints can then be used to help rank candidate metabolite IDs obtained based on predicted formula or measured precursor m/z of the unknown metabolite. This approach is particularly useful to help annotate metabolites whose corresponding MS/MS spectra cannot be matched with those in spectral libraries. We previously reported application of a convolutional neural network (CNN) for molecular fingerprint prediction using MS/MS spectra obtained from the MoNA repository and NIST 20. In this paper, we investigate high-dimensional representation of the spectral data and molecular fingerprints to improve accuracy in molecular fingerprint prediction.