Pacific Northwest
NATIONAL LABORATORY

PNNL-38307

Refining PeakDecoder Version 2

September 2025

Aivett Bilbao
Ashfiqur Rahman
Andrea Harrison

U.S. DEPARTMENT of ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-38307

Refining PeakDecoder Version 2

September 2025

Aivett Bilbao
Ashfiqur Rahman
Andrea Harrison

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

Novel computational tools for processing multidimensional mass spectrometry (MS) data are necessary to enable deeper and automated detection and quantification of metabolites in complex backgrounds. Multidimensional MS data includes measurements from liquid chromatography (LC) and ion mobility spectrometry (IM) separations, and precursor and fragment ion spectra collected in data-independent acquisition (DIA) mode. PeakDecoder is an artificial intelligence (AI)-based software that enables automated interpretation of this kind of data to identify and quantify individual metabolites in complex mixtures. The goal of this project was to improve and re-implement PeakDecoder in a better suited programming language to enable its commercialization.

Summary

During this project we expanded the training set, implemented PeakDecoder v2 in C#, fine-tuned the model on the larger dataset, benchmarked results, implemented a prototype C# inference module and GUI, and implemented an additional ion mobility instrument vendor format.

Acknowledgments

This research was supported by the Strategic Investment under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

1.0 Introduction

Metabolites are molecules produced and transformed in large networks of cellular processes and biochemical reactions in biological and environmental systems. The sheer diversity of molecular classes and structures constitutes a significant analytical challenge in terms of detection and annotation in complex samples. Multidimensional mass spectrometry (MS) measurements, including liquid chromatography (LC) and ion mobility (IM) separations, and data-independent acquisition (DIA) spectra, capture rich information to better characterize and study biomolecular signatures with increased separation power. LC-IM-MS provides the deepest ID coverage with high confidence and reproducible quantitation with DIA [1], and further distinguishes structurally similar molecules with IM [2]. However, this kind of multidimensional data is complex and cannot be processed with traditional tools. It requires more sophisticated processing algorithms compared to conventional LC-MS platforms. To address this challenge, we developed PeakDecoder, an algorithm that uses a support vector machine to enable metabolite annotation and accurate profiling in multidimensional MS data [3]. We demonstrated the usefulness and capabilities of the instrument platform and PeakDecoder to interpretate metabolite features in synthetic biology-related microbial strain samples, and more recently to identify and quantify metabolites from distinct environmental microbes in a complex background [4]. However, this first version of the algorithm was not fully automated due to software dependencies. Therefore, a new version was needed to enable automated data processing of multidimensional MS data.

2.0 Results

We expanded the training set of images and labels by curating additional examples from multidimensional metabolomics MS data from prior publications and applying data augmentation techniques to simulate varied ion abundances and isomers. We implemented PeakDecoder v2 in C# by converting the PyTorch model to ONNX Runtime, loading pre-trained weights, and running inference on ion images. We fine-tuned on the larger dataset of 1068 images, achieving over 20% more metabolite IDs. Results from both model formats agreed to three decimal places across metrics. We implemented a prototype C# inference module and GUI and implemented the Bruker ion mobility format to support both Agilent and Bruker data.

3.0 References

1. Bilbao, Aivett, Emmanuel Varesio, Jeremy Luban, Caterina Strambio-De-Castillia, Gérard Hopfgartner, Markus Müller, and Frederique Lisacek. "Processing strategies and software solutions for data-independent acquisition in mass spectrometry." Proteomics 15, no. 5-6 (2015): 964-980.
2. Ross, Dylan H., Harsh Bhotika, Xueyun Zheng, Richard D. Smith, Kristin E. Burnum-Johnson, and Aivett Bilbao. "Computational tools and algorithms for ion mobility spectrometry-mass spectrometry." Proteomics (2024): 2200436.
3. Bilbao, Aivett, Nathalie Munoz, Joonhoon Kim, Daniel J. Orton, Yuqian Gao, Kunal Poorey, Kyle R. Pomraning et al. "PeakDecoder enables machine learning-based metabolite annotation and accurate profiling in multidimensional mass spectrometry measurements." Nature Communications 14, no. 1 (2023): 2461. DOI 10.1038/s41467-023-37031-9.
4. Veličković, Marija, Ruonan Wu, Yuqian Gao, Margaret W. Thairu, Dušan Veličković, Nathalie Munoz, Chaevien S. Clendinen et al. "Mapping microhabitats of lignocellulose decomposition by a microbial consortium." Nature Chemical Biology (2024): 1-11. DOI 10.1038/s41589-023-01536-7.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov