Nicole North,[+] Abigail Enders,[+] Morgan Cable,[^] Heather Allen[+]
Ohio State University Department of Chemistry & Biochemistry,[+] NASA Jet Propulsion Laboratory[^]

Introduction

The aqueous plume of Enceladus, one of the moons of Saturn, has the potential to contain signs of life. This plume has been sampled in previous NASA missions, namely the Cassini mission by mass spectrometry. Evaluation of returned data has proven challenging due to the instrumental limitations of mass range and resolution. AI based approaches show promise in increasing knowledge of the organics on Enceladus.
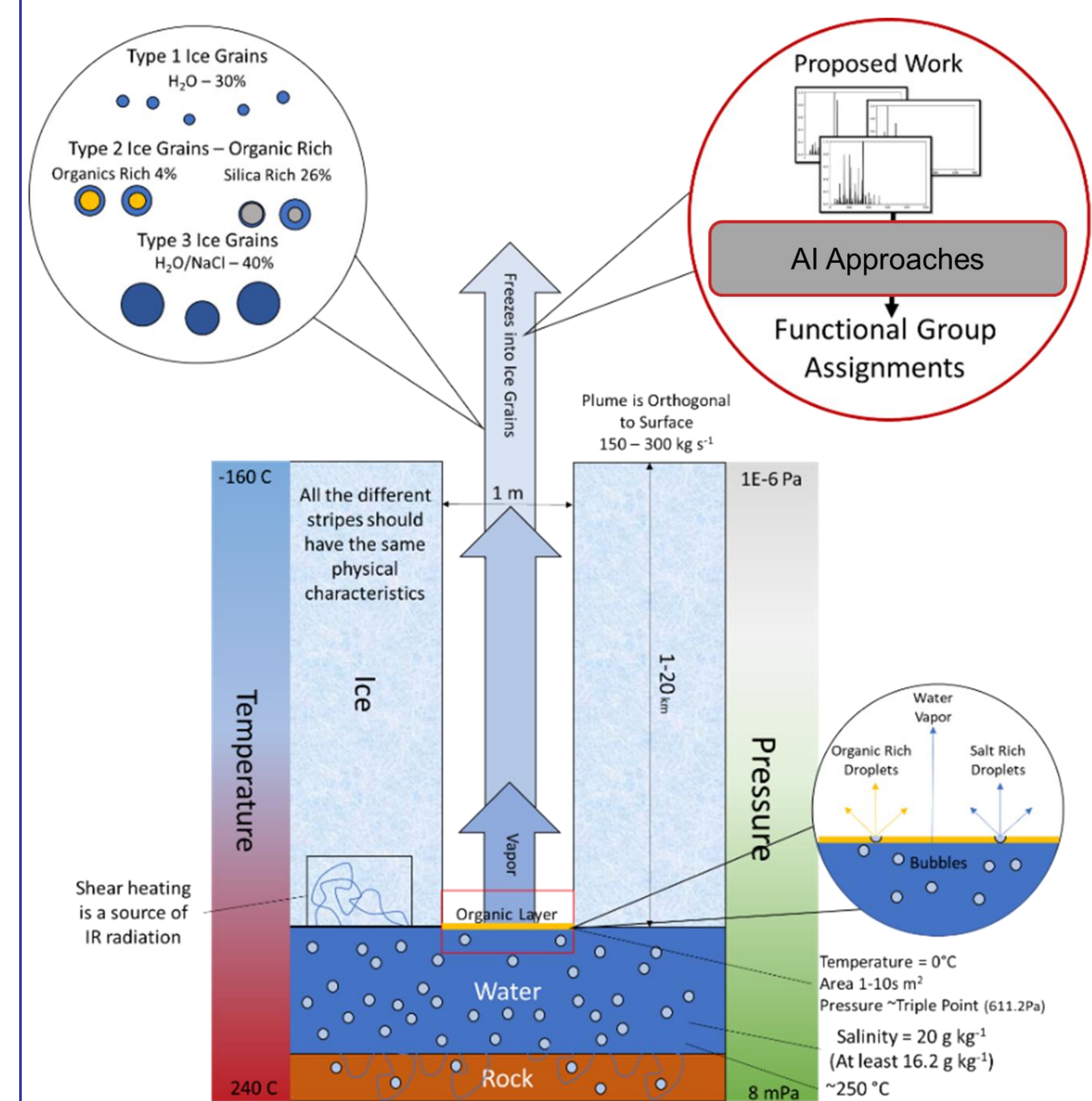

Diagram depicting the different physical and chemical systems that have been proposed to be occurring on Enceladus.[1-3]

One of the challenges of processing data from these interstellar missions is that mass spectrometry fragmentation methods are not well established or studied when compared to more common laboratory mass spectrometry techniques. Beyond only assigning functional groups to molecules, It is important to develop methodology to determine how AI models are making their assignments so that the patterns may inform our understanding of the fragmentation technique.

1.) Postberg, F.; Schmidt, J.; Hillier, J.; Kempf, S.; Srama, R. A Salt-Water Reservoir as the Source of a Compositionally Stratified Plume on Enceladus. *Nature* 2011, 474 (7353), 620–622. https://doi.org/10.1038/nature10175. 2.) Klenner, F.; Postberg, F.; Hillier, J.; Khawaja, N.; Reviol, R.; Stolz, F.; Cable, M. L.; Abel, B.; Nölle, L. Analog Experiments for the Identification of Trace Biosignatures in Ice Grains from Extraterrestrial Ocean Worlds. *Astrobiology* 2020, 20 (2), 179–189. https://doi.org/10.1089/ast.2019.2065. 3.) Klenner, F.; Postberg, F.; Hillier, J.; Khawaja, N.; Cable, M. L.; Abel, B.; Kempf, S.; Glein, C. R.; Lunine, J. I.; Hodyss, R.; Reviol, R.; Stolz, F. Discriminating Abiotic and Biotic Fingerprints of Amino Acids and Fatty Acids in Ice Grains Relevant to Ocean Worlds. *Astrobiology* 2020, 20 (10), 1168–1184. https://doi.org/10.1089/ast.2019.2188. 4.) Enders, A. A.; North, N. M.; Fensore, C. M.; Velez-Alvarez, J.; Allen, H. C. Functional Group Identification for FTIR Spectra Using Image-Based Machine Learning Models. Anal. Chem. 2021, 93 (28), 9711–9718. https://doi.org/10.1021/acs.analchem.1c00867.

Methodology
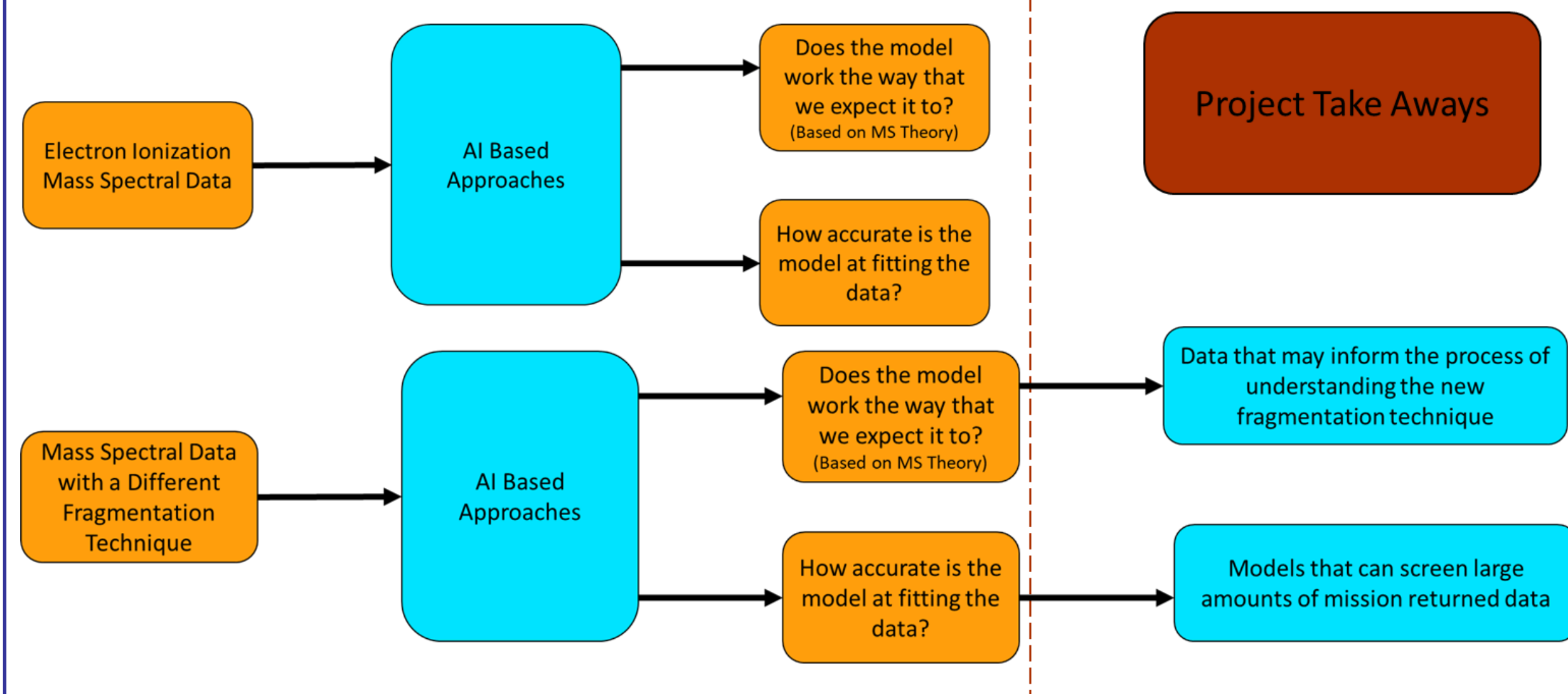
Flow charts depicting how the results of the Electron Ionization (EI) trained models may transfer over to our understanding of different fragmentation techniques.

Over 20,000 election ionization mass spectra were collected from the United States' National Institute of Standards and Technology website available through web scraping the various CAS number pages.
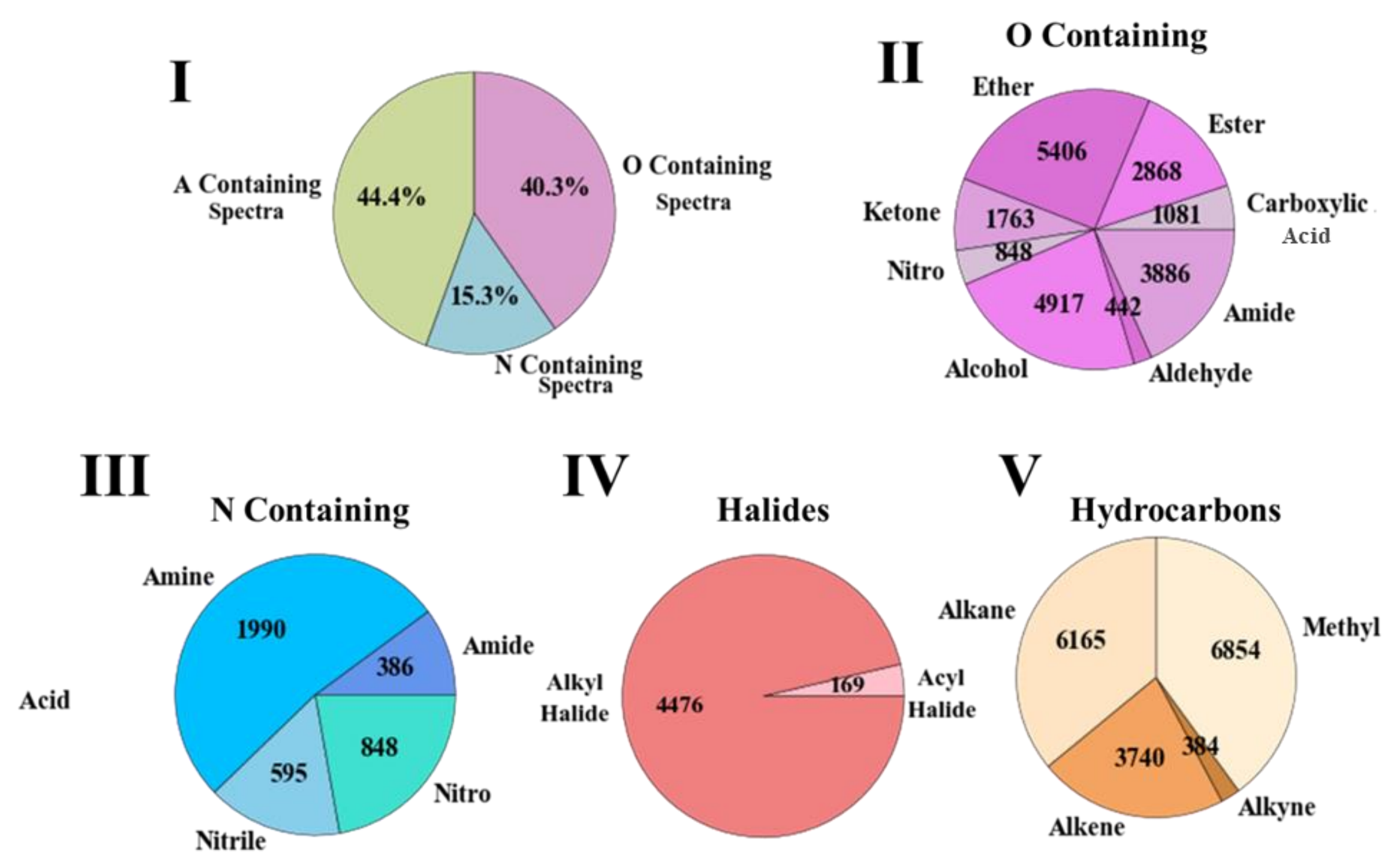

Distributions of available mass spectra from NIST included in this study, organized by functional group classification.

The data were divided into classes to look at two different levels of classification, specific functional groups (ether, nitro, etc.) and generalized functional groups (aromatic, nitrogen-containing, oxygen containing). This was done so that the generalizability and the specificity of the models could both be explored. Both approaches are applicable to screening large amounts of mass spectral data for analysis.
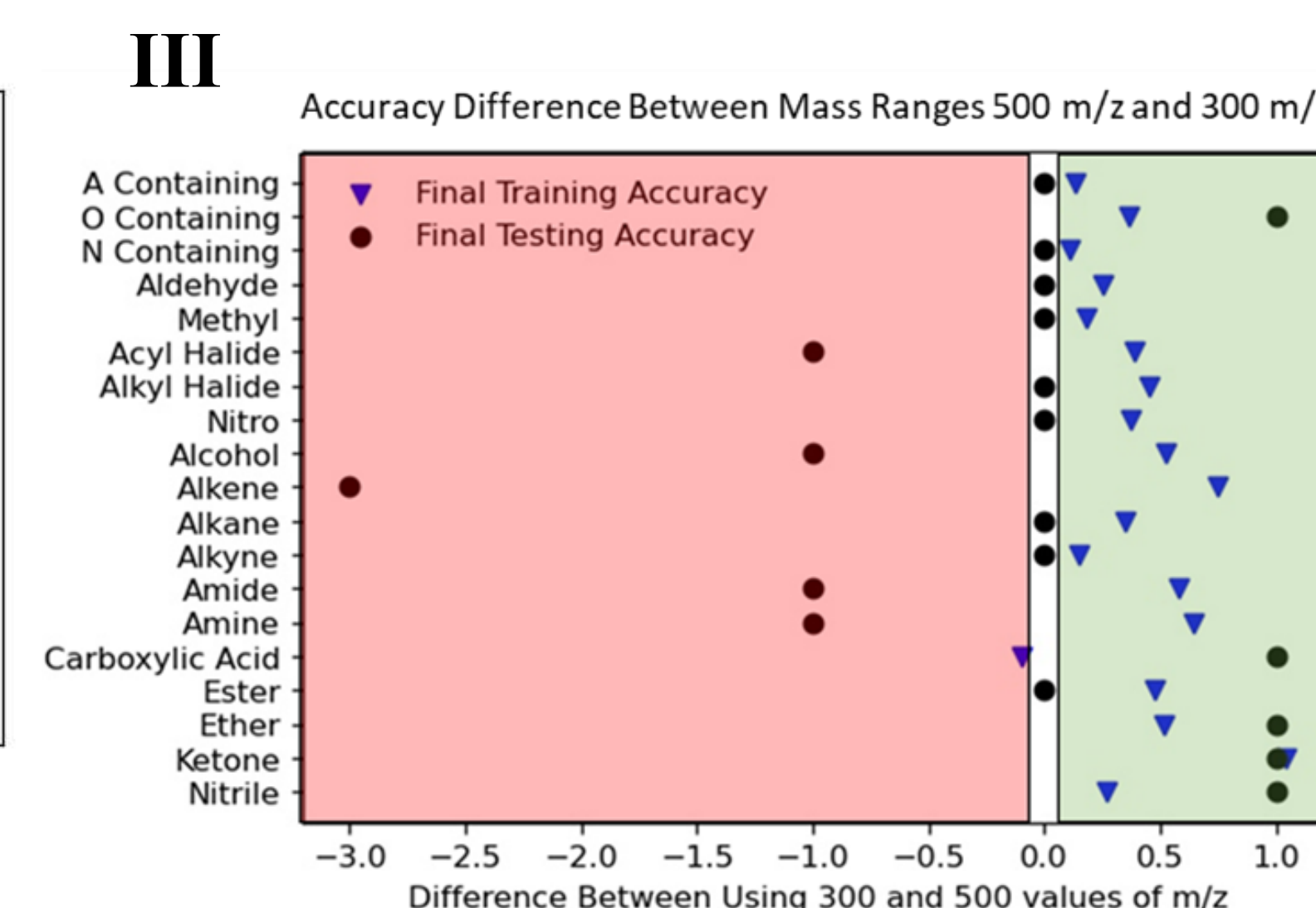
Results

Transfer Learning vs Logistic Regression
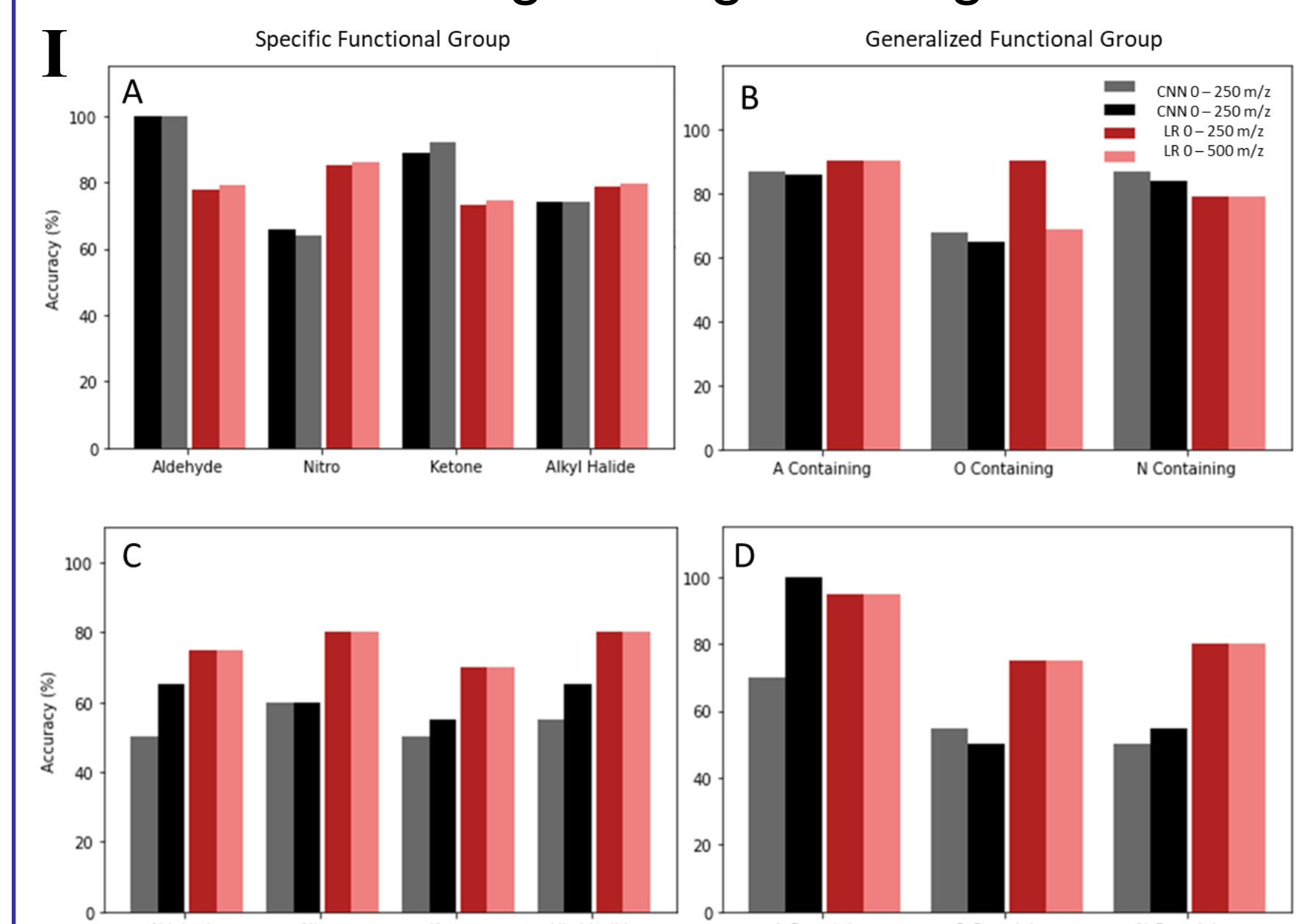
Model Results by Functional Group
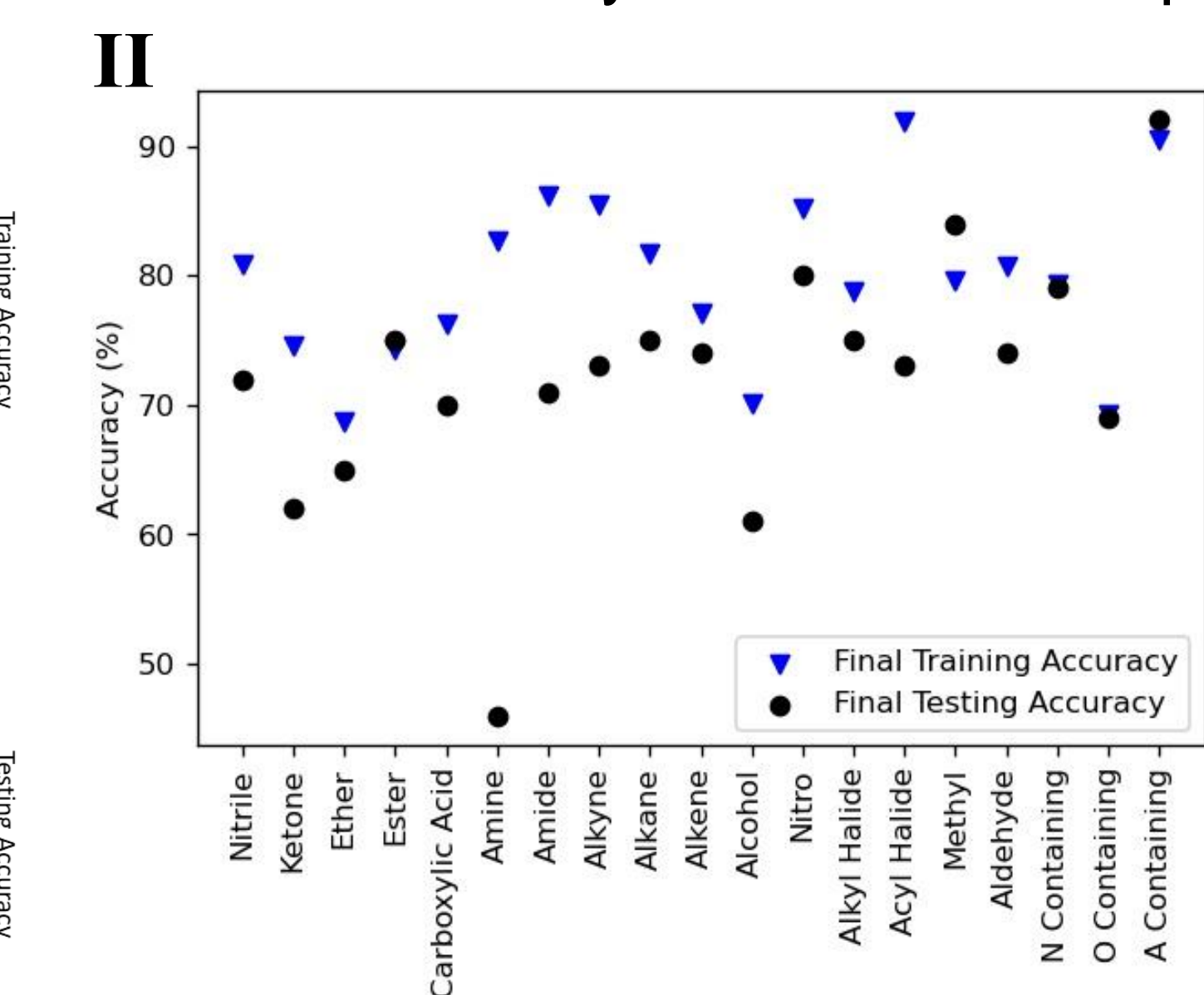
Effects of Feature Size

- Logistic Regression was more successful than transfer learning with a computer vision convolutional neural network at fitting the mass spec data. This was unexpected based on our previous work with IR data in which transfer learning preformed well.[4] **(I)**
- The variation in model accuracy varied more between functional groups than with adjusting for the number of features or training parameters. Confirming that some functional groups are harder to fit. **(II)**
- Increasing the utilized mass range led to a trend in increasing final training accuracy while reducing final testing accuracy. This suggests that with larger feature sizes overfitting is a potential risk. **(III)**

Summary

In determining how the trained models made their assignments, we looked at which masses were the most correlated with the models displaying high accuracies. These masses were related to common rules that we are taught when interpreting mass spectra without being explicitly taught. These included the odd nitrogen rule when looking for N-containing molecules. **(Left)**

The models were also tested on a molecule of interest, an amino acid, Tryptophan to showcase how these models might be used to screen mass spectral data. The presence of the functional group was correctly identified for 13 out of the 20 models. **(Right)**

These combined results show promise in improving how we screen large mass spectral datasets and interpret the patterns found in mass spectral data.

Ongoing Work

Future directions of this work include testing this training methodology on mass spectra taken on terrestrial instruments that utilize hypervelocity ice grain fragmentation, the fragmentation technique done aboard the Cassini mission, as well as further developing our methodology for identifying impactful peaks for model fitting.