DEPARTMENT OF CHEMICAL ENGINEERING

Molecular, Environment, Energy and Health Data Science Graduate Capstone





Join us for the Molecular, Environment, Energy and Health Data Science Graduate Capstone presentations. FREE FOOD! Projects include:

Machine Learning for AFM image Correction

Christina Yang, Thom Snoeren, Haoqing Zhang, Lejla Biberic, (Gregory Moore)

Collecting high-quality AFM images can be a daunting task, even for an experienced scientist. Navigating all the different kinds of noise and artifacts requires a highly trained eye, but it does not have to be this way. We have built a trained machinelearning model that loads, processes and plots the data for you, and then removes unwanted imaging artifacts, giving you publication-quality AFM images in a matter of seconds. Future work on this software will include real-time feedback, so scan settings can be adjusted on the fly.

XenoFind: a de novo Consensus and Detection Pipeline for Xenonucleic Acid (XNA) Containing Aptamers

Jayson Sumabat, Nicholas Lai, Yujia Huang, Sebastian Peck

XenoFind is a Python program designed to find XNA positions in aptamer datasets. The novelty of XenoFind is its use of consensus-level features, which mitigate per-read nanopore sequencing errors by aggregating reads and employing features that characterize the entire group. Users have the option to apply preexisting models or generate new models using datasets with known XNA positions. XenoFind subsequently generates accurate consensus sequences from unknown aptamer datasets with ligated barcodes. Finally, it leverages these consensus sequences and the selected models to pinpoint regions of XNAs within the sequences in common bioinformatic file formats.

LLMao: Retrieval Augmented Generation (RAG) using Large Language Models (LLMs) for Toxicology

Ali Mahmoud, Ashley Fenton, Jhanvi Rana, Kenny Lam, Omkar Chavan

With the surge in popularity LLMs such as ChatGPT, the accuracy of their responses to complex scientific questions often falls short. By leveraging adverse outcome pathway (AOP) data to augment the model's knowledge, LLMao offers high precision when answering domain specific toxicological questions. With robust evaluation methods, connections to multiple AOP databases, and a userfriendly interface, LLMao can democratize to scientific data and access comprehension.

Organization and Analysis of Battery Data from King County Metro Hybrid-Electric Bus Fleet

Phil Romero, Elena Toups, Babita Giri, Chris Gould

Public transportation agencies are switching to electric and will need better tools for understanding battery health and performance. Onboard battery management systems collect large amounts of information, but extracting important metrics to inform maintenance is challenging. We developed software that makes sense of the mess, by sorting raw data into convenient folders and data structures. Our software then leverages statistical analysis to reveal important features of a battery module's performance. This work paves the way for predicting a battery module's remaining life and time to failure using supervised machine learning.

Thursday, June 6 4:30-6:20 PM WRF Data Science Studio Physics and Astronomy Tower, 6th floor (<u>map</u>)

Sound Minds: An Auditory Stimulus Package for EEG Data Acquisition from Patients in Unresponsive States

Annika Philomin, Jacob Cavon, Khanh Ha, Roni Weissman, (Anika Gupta, Arielle Hancko)

Electroencephalogram (EEG) readouts from unresponsive patients exposed to specific stimuli have significant prognostic value. However, the lack of a standardized tool for administering stimuli and collecting EEG responses makes it difficult to compare datasets across studies. Our auditory stimulation administration software is a user-friendly, standardized tool designed for physicians to conduct and streamline their stimulus-EEG readout studies.

ProspecPy: Automatic Fourier Transform Infrared Spectroscopy (FTIR) Spectrum Analysis

Baide Xue (Eric), Annatu Amadu Somah, Anshul Tambay, Anuradha Ramachandran, Elizabeth Phillips, (Katelyn C Massie)

Be faster and more consistent in your research. ProspecPy is a Python package that automatically analyzes FTIR spectra and replaces the previous 1.5hr/reading point-and-click-manual-analysis.

ProspecPy's process features 7 operations including: water vapor removal, range of interest selection, after range selection plot, second derivative and plot, anchor point selection, peak fitting of baselinesubtracted data, and subtracted baseline plot. ProspecPy also offers an intuitive graphical interface to help users adjust the anchor point selection, reducing human error.