Subject: Thesis Defense: Sydney Monserrate, MS Biology
Date: Thursday, April 4, 2024 at 2:55:06 PM Eastern Daylight Time
From: SSB Faculty List on behalf of Diane St. Germain
To: SSB-FACULTY-LIST-L@LISTSERV.GMU.EDU

Thesis Defense Announcement
To: The George Mason University Community

Candidate: Sydney Monserrate

Program: M.S. in Biology

Date: Friday April 19, 2024

Time: 1:00 PM Eastern Time (US and Canada)

Join Zoom Meeting
https://gmu.zoom.us/j/92306845722?pwd=aWRjTm04b2tlbWplRE50d0p6K1d5Zz09

Meeting ID: 923 0684 5722
Passcode: 893310

One tap mobile
+12678310333,,92306845722#,,,,*893310# US (Philadelphia)
+13017158592,,92306845722#,,,,*893310# US (Washington DC)
Dial by your location
+1 267 831 0333 US (Philadelphia)
+1 301 715 8592 US (Washington DC)

Find your local number: https://gmu.zoom.us/u/adTzQWMz5m
Join by SIP 92306845722@zoomcrc.com

Committee Chair: Dr. Christopher Lockhart
Committee Members: Dr. Iosif Vaisman, Dr. Ancha Baranova

Title: “Predicting Alzheimer’s Disease from miRNA Sequence and Expression Data with Machine Learning”

Abstract:
In the United States, approximately 6.5 million people have been diagnosed with Alzheimer's Disease (AD). This cognitive disorder mainly affects those who are 65 years and older. The use of exosomal and circulating miRNA data is relatively new to biological studies but has become a focal point due to miRNA availability in bodily fluids and potential use in disease diagnostics, including AD screening. The purpose of our study was to investigate the utility of machine learning to predict AD-associated outcomes with miRNA sequence and expression data. Machine learning was performed leveraging the Orange Data Mining platform, which allowed us to quickly prototype various machine learning models and assess their performance numerically and graphically. To utilize miRNA sequence data, we employed a k-mer bag of words model to quantify subsequences within miRNAs and predict if miRNAs interact with proteins involved in AD pathways. We found that a Random Forest model provides the best predictions with an

accuracy of 0.772 and an area under the receiver operating characteristic (AUROC) of 0.813. Interestingly, out all k-mers, those that are rich in purines are the most predictive of miRNA association with AD. As a second modelling effort, we analyzed a previously published dataset that measured miRNA expression in AD or healthy patients. A Random Forest model produced an accuracy of 0.786 and AUROC of 0.862 approximately reproducing the published results. We explored if the probability for miRNA to be associated with AD-related pathways can be used as additional selection criteria for miRNA expression profile analyses, and we discuss broader applications of our machine learning models in AD diagnostics.

###