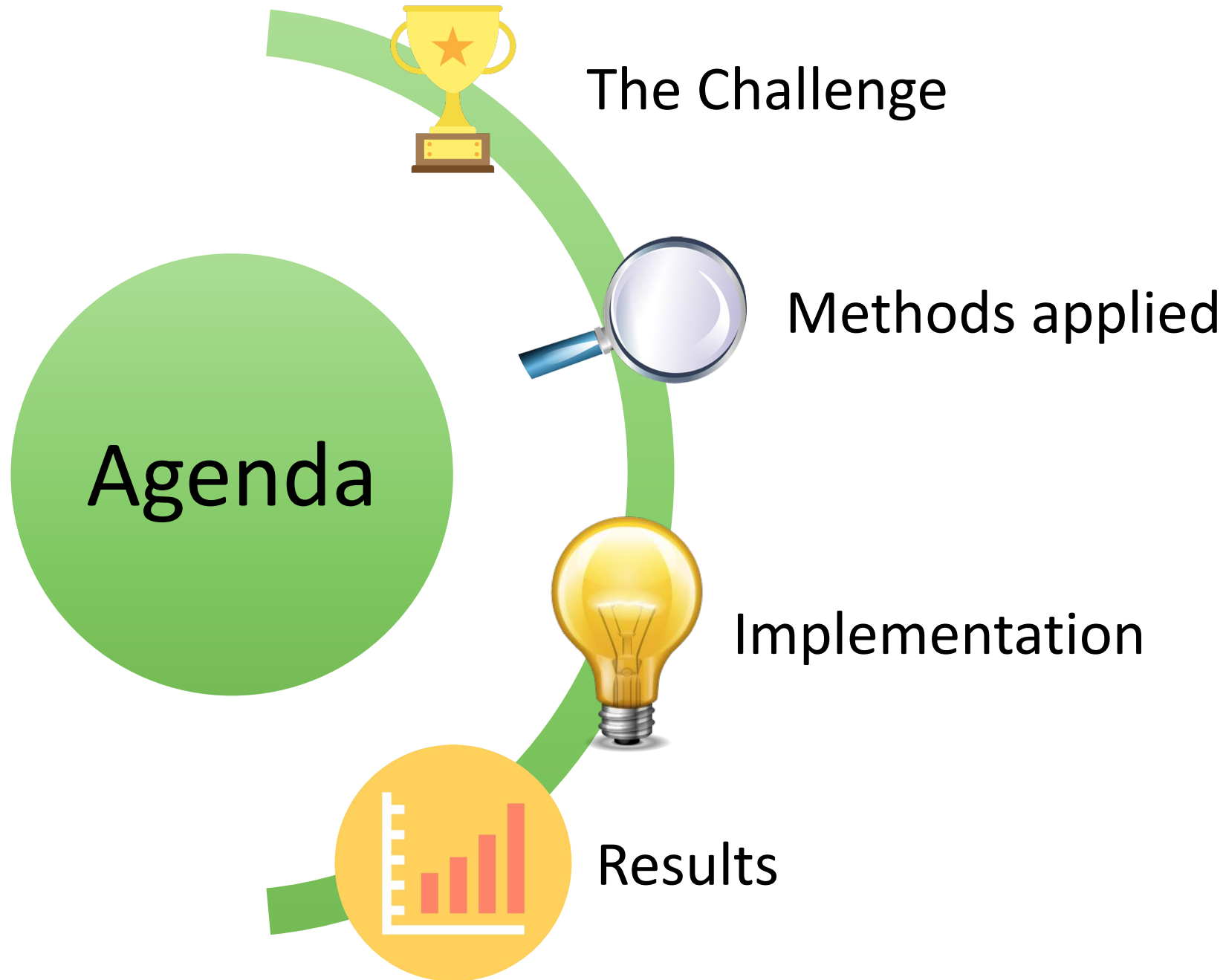


# Microbiome Data Dimensionality Reduction with Autoencoder

Students: Andrea Mihajlovic, Hunor Tot-Bagi

Professor: dr Sanja Brdar

Novi Sad 2020.



# The Challenge



IBD (103)

Inflammatory  
bowel  
disease

UC (86)

Ulcerative  
colitis

CD (17)

Crohn's  
disease

non-IBD (67)



- IBD vs non-IBD

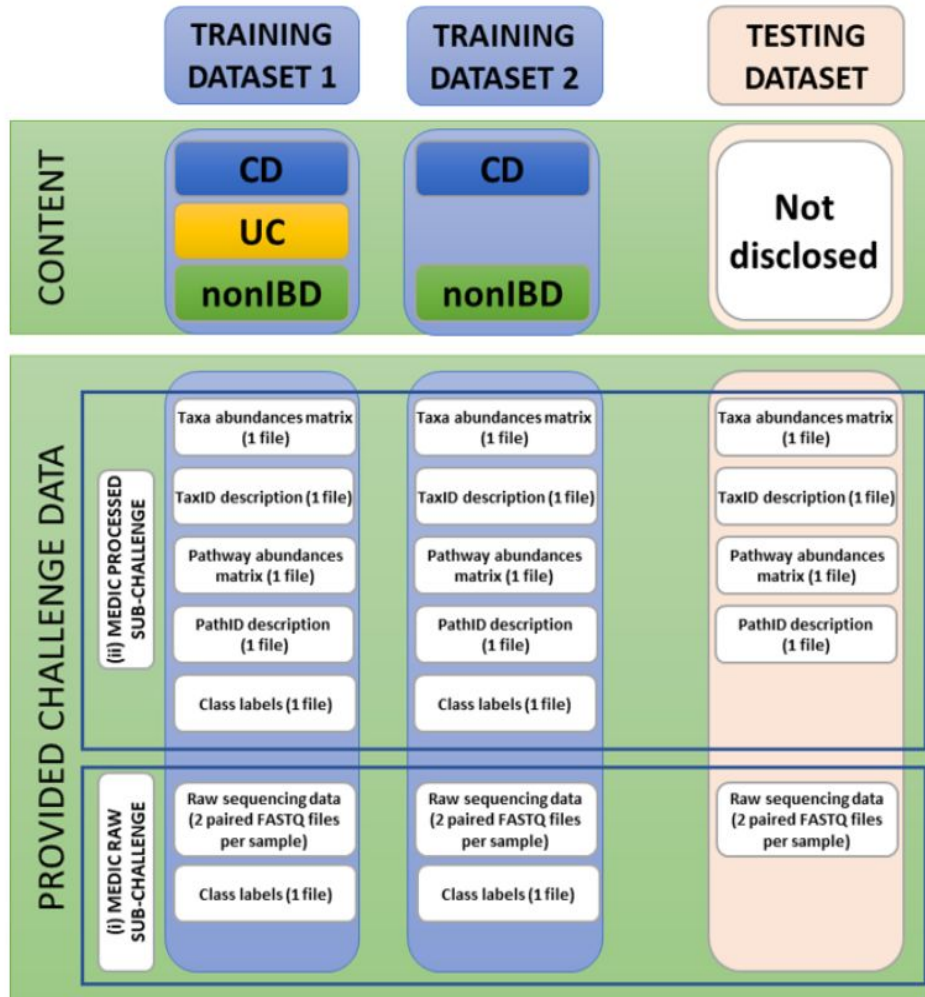
- CD vs non IBD



- UC vs non-IBD

- UC vs CD

# Datasets



	Sample 1	Sample 2	....	Sample m
TaxID 1	10	0.0	....	30
TaxID 2	30	21	....	40
....	....	....	....	....
TaxID n	50	15	....	0.0

Samples names

% of reads from the particular sample assigned to the particular TaxID relative to all reads from this sample assigned at the same taxonomy level

TaxIDs

	Sample 1	Sample 2	....	Sample m
PathID 1	10	0.0	....	12
PathID 2	15	0.2	....	100
....	....	....	....	....
PathID n	252		....	0.0

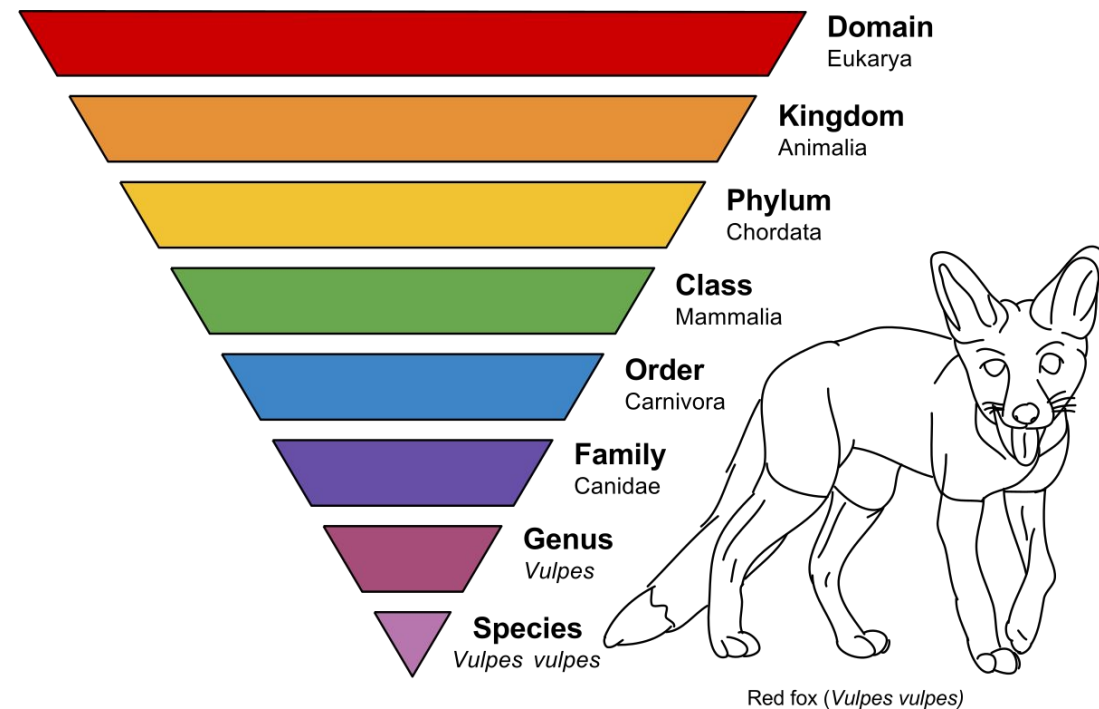
Samples names

Abundance of the particular PathID in the particular sample.

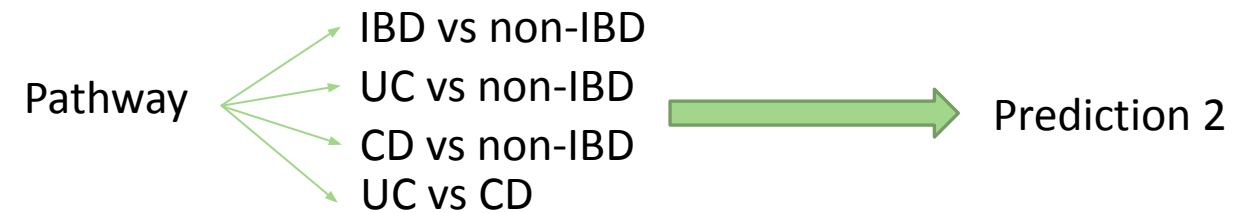
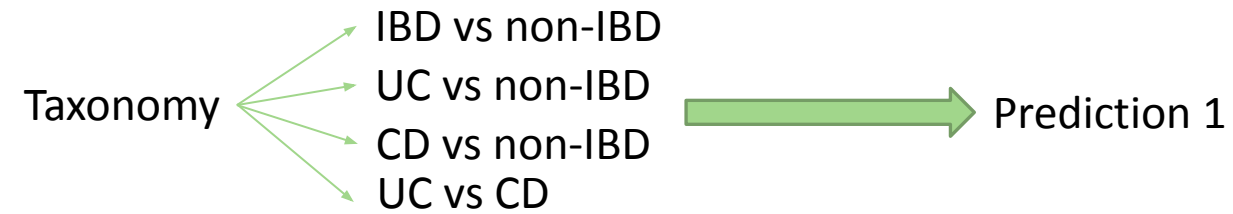
PathIDs

# Taxonomy

- Science of naming
- Organisms are grouped together into taxa
- These groups are given a taxonomic rank



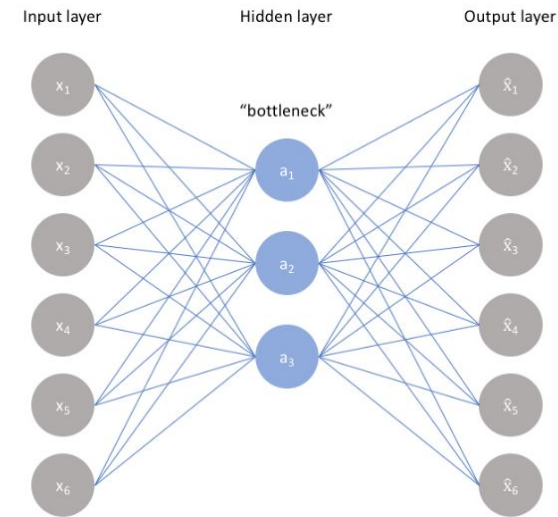
# Pathway and Taxonomy



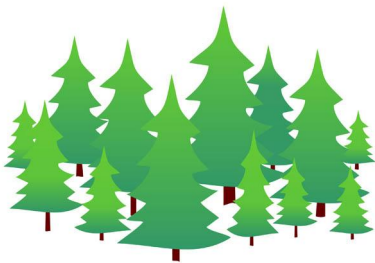


# Methods

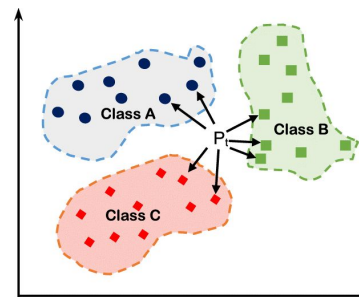
- What is an Autoencoder?
- Classification algorithms



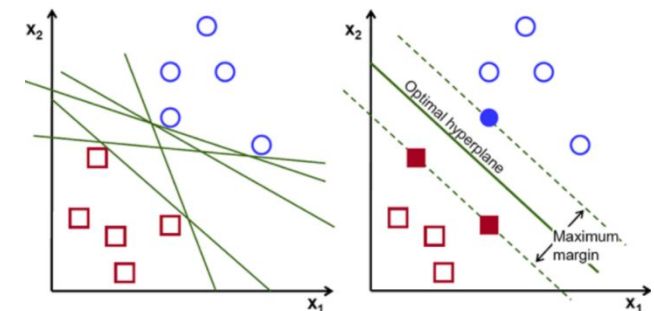
RF



kNN

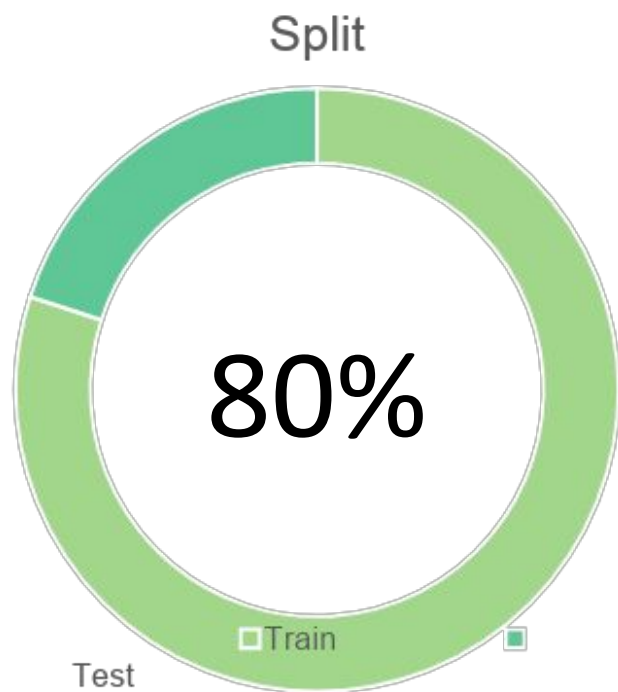


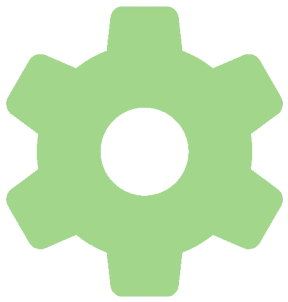
SVM



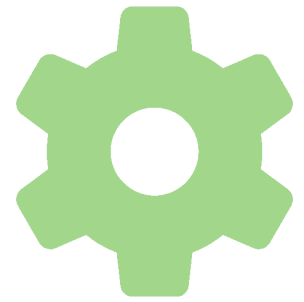


# Implementation





# Parameter



## Activation

- Sigmoid
- ReLU

## Optimizers

- Adam
- Nadam
- Adamax
- Adadelata
- SGD

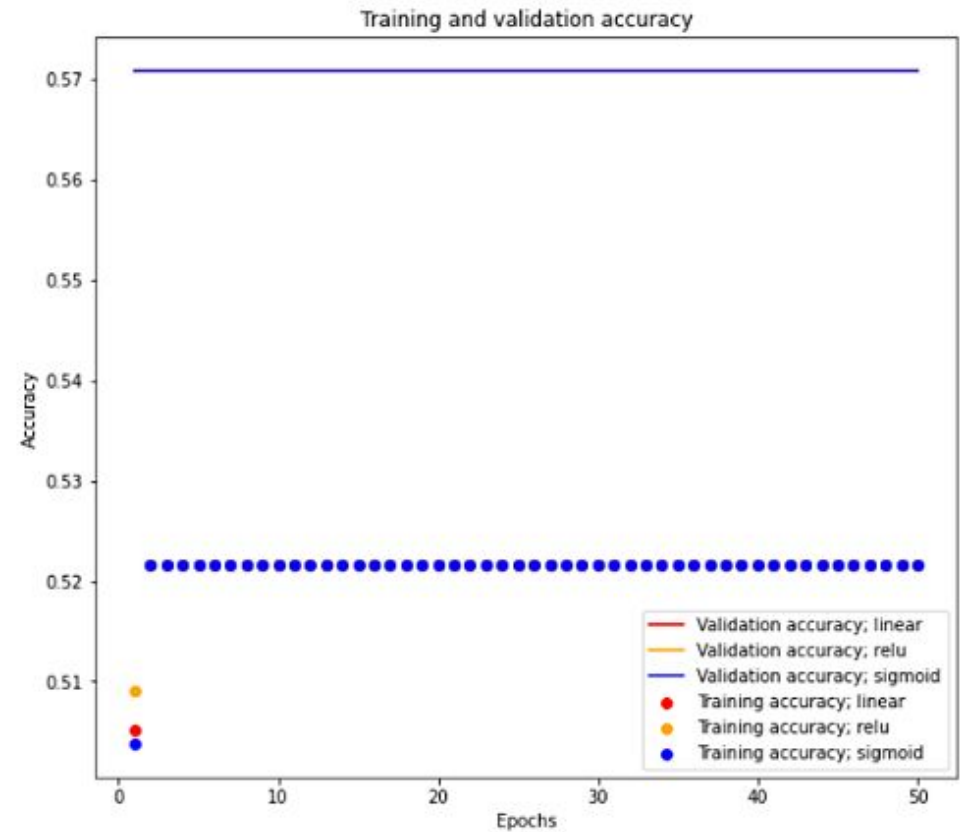
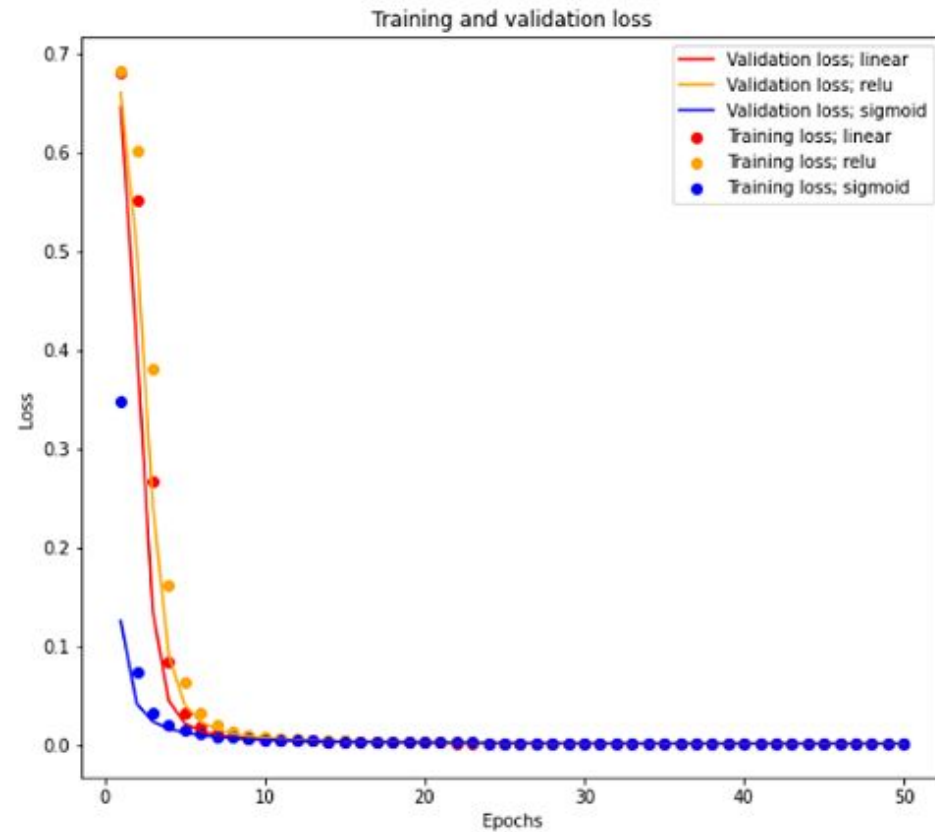
## Hidden units

- 24
- 100
- 150
- 200
- 500

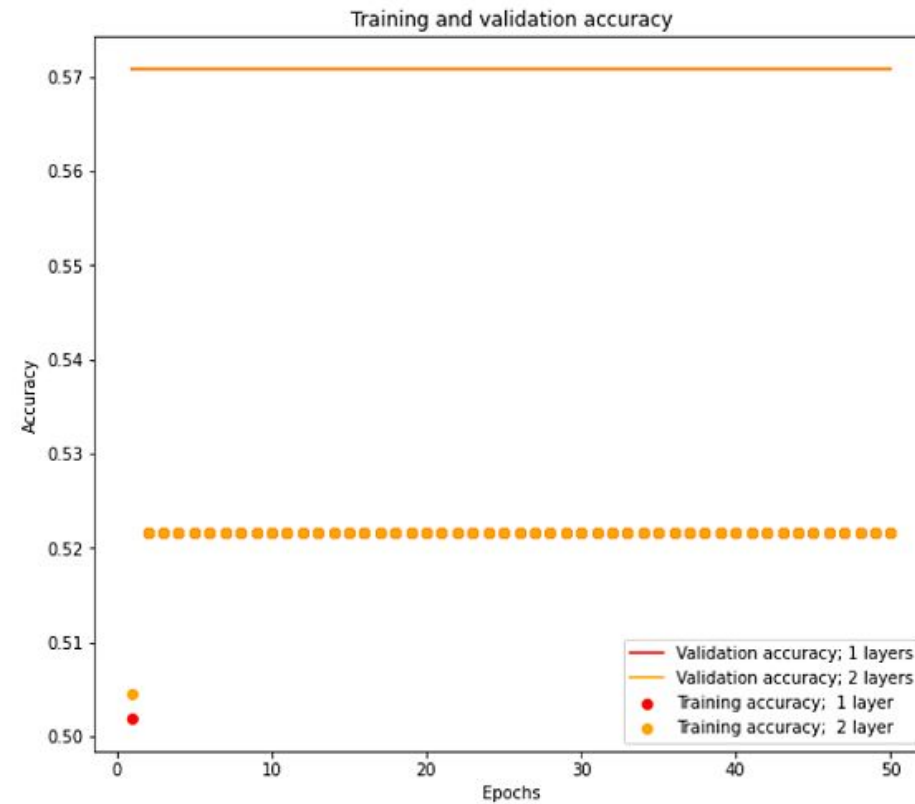
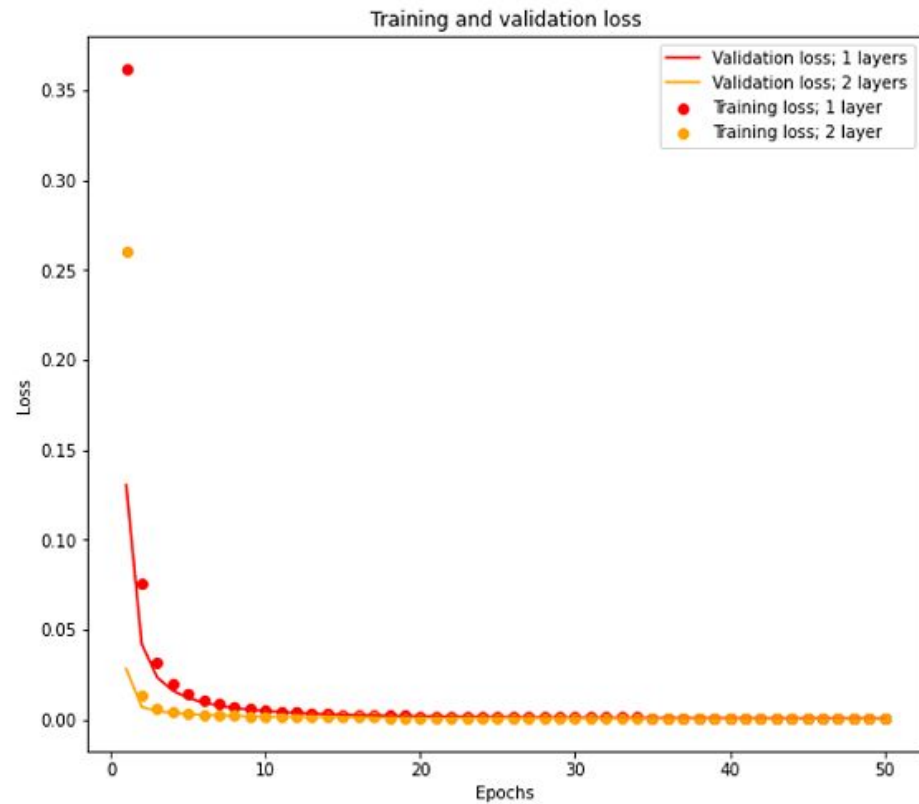
## Epochs

- 50

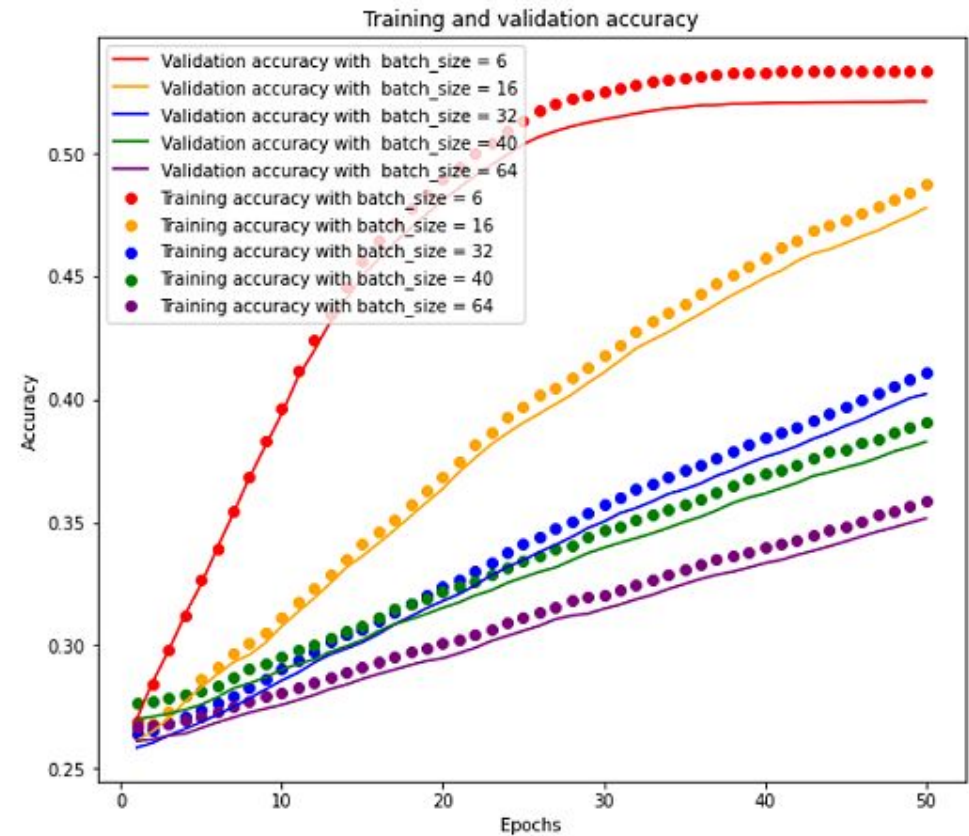
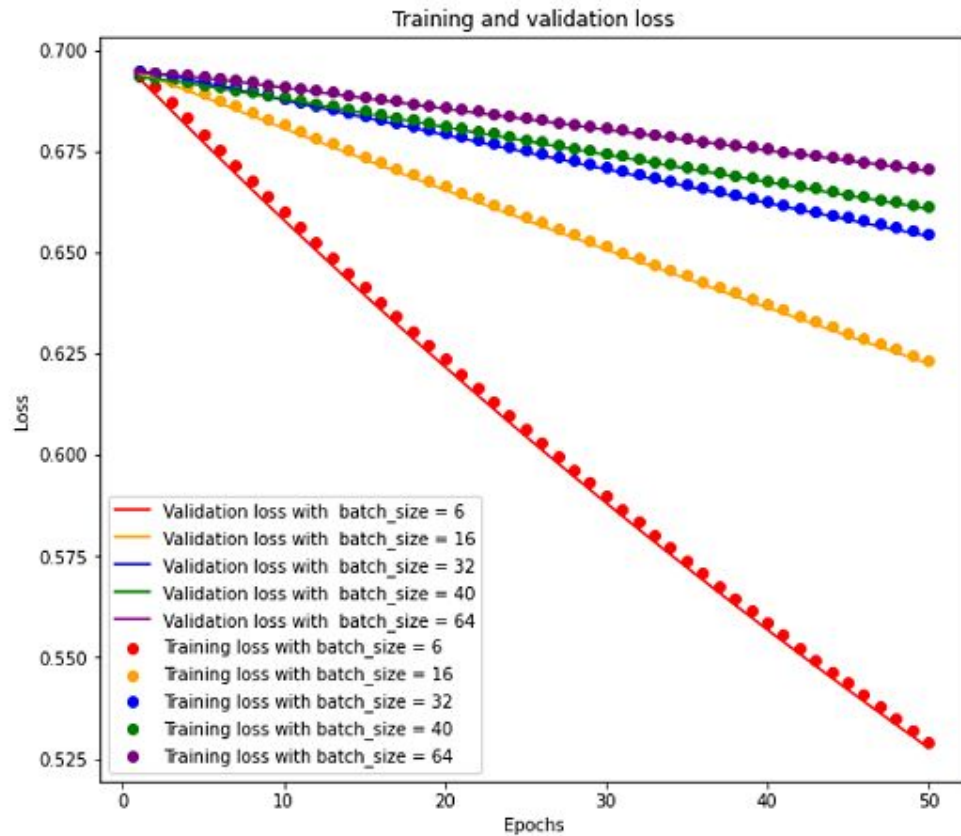
# Results – Activation



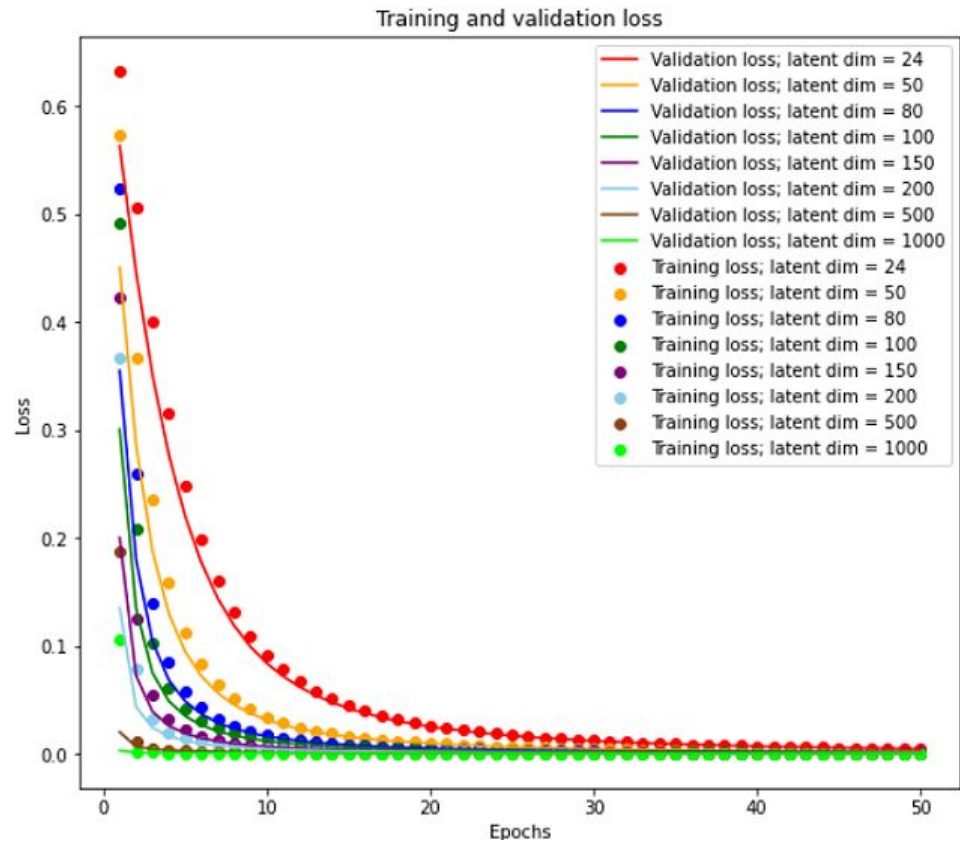
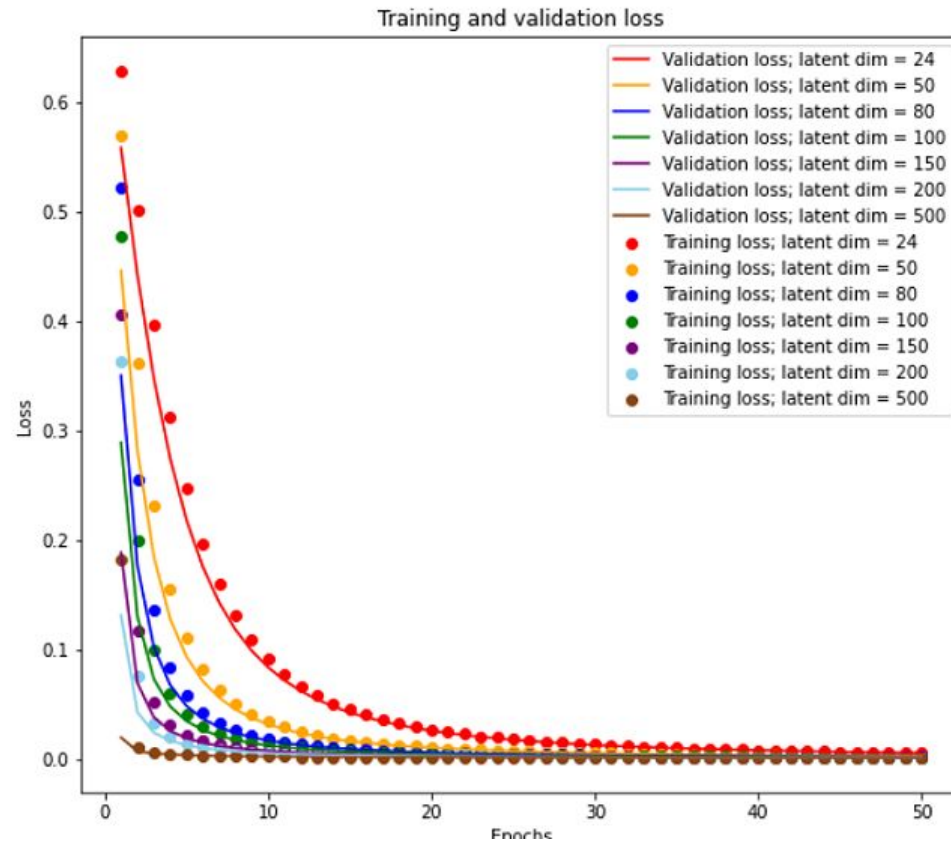
# Results – Number of



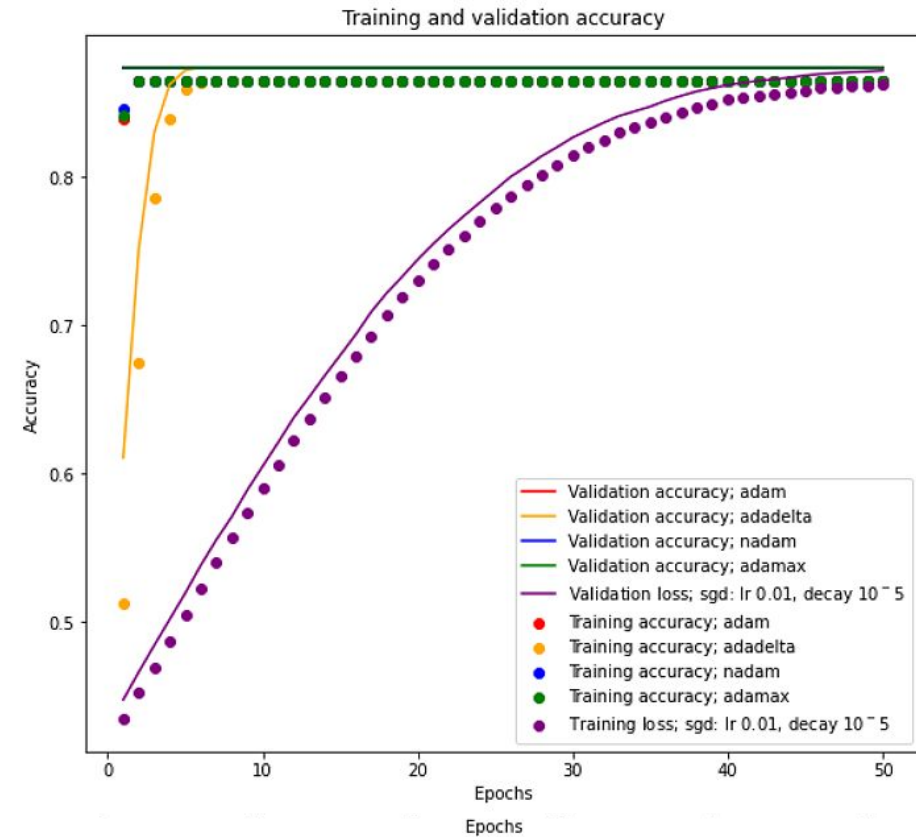
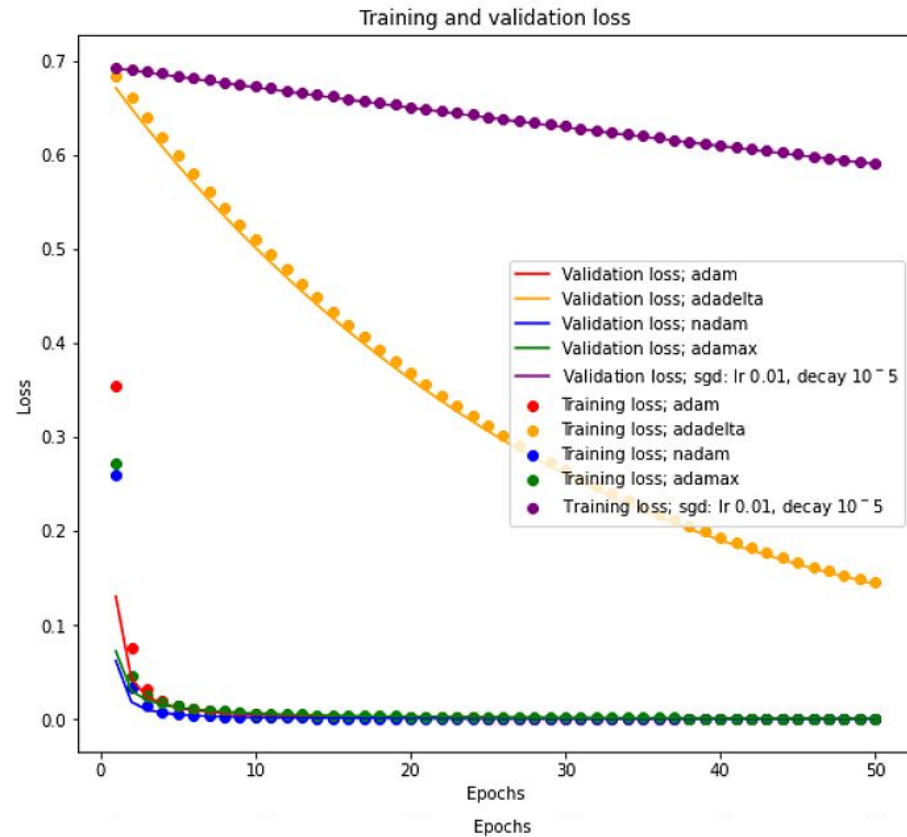
# Results – Batch



# Results – Latent dimension

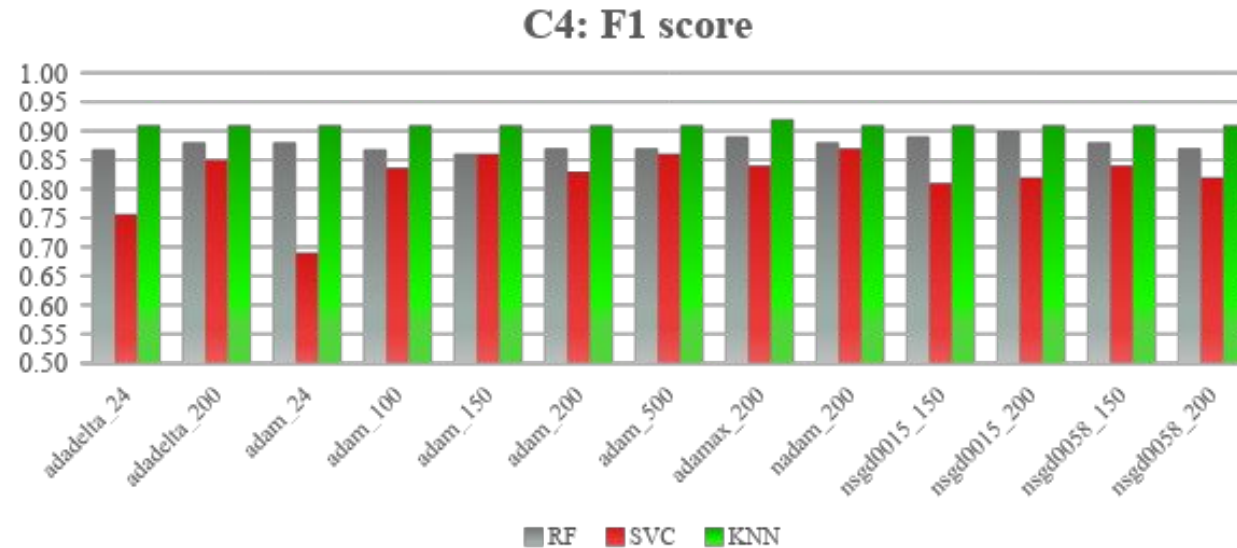
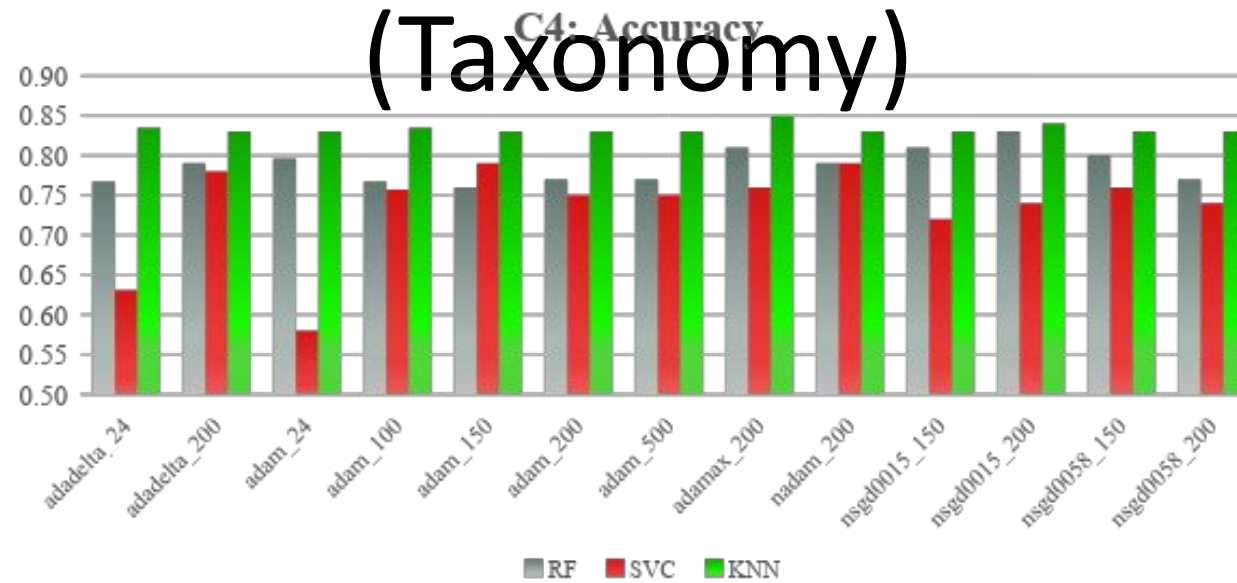


# Results –

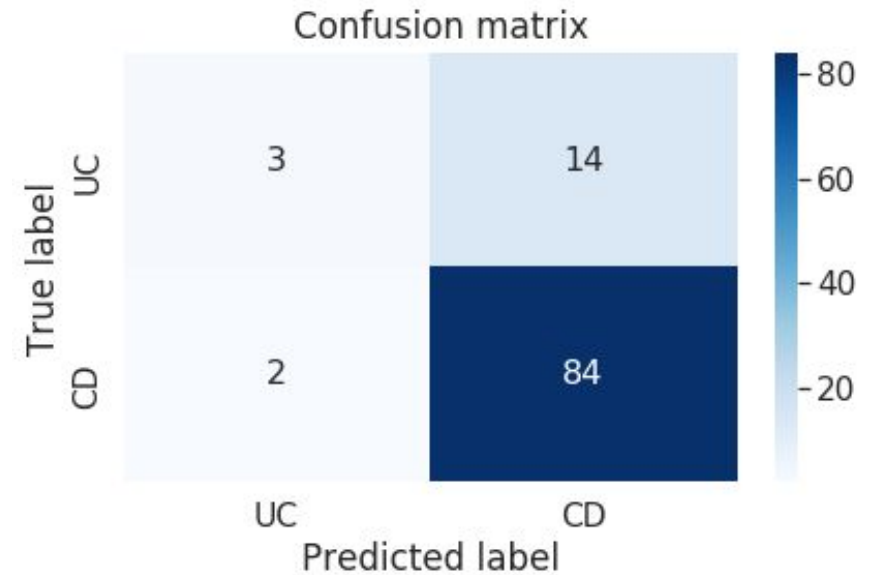
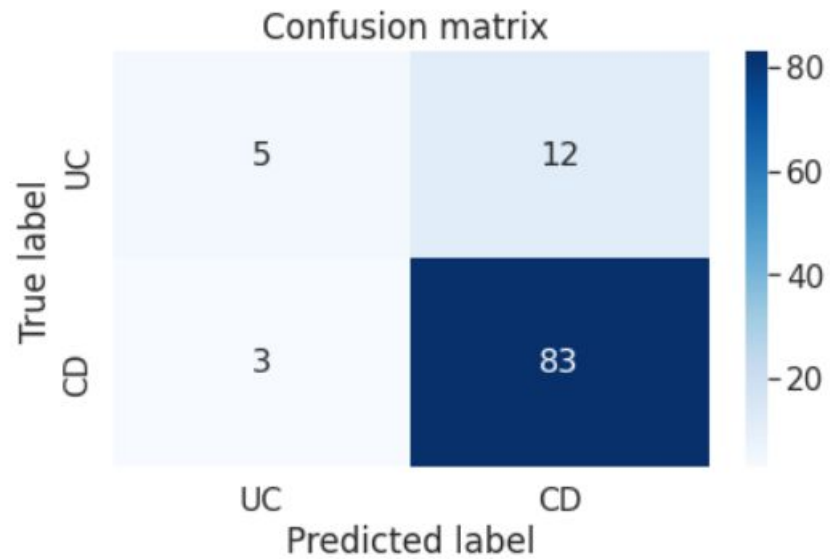




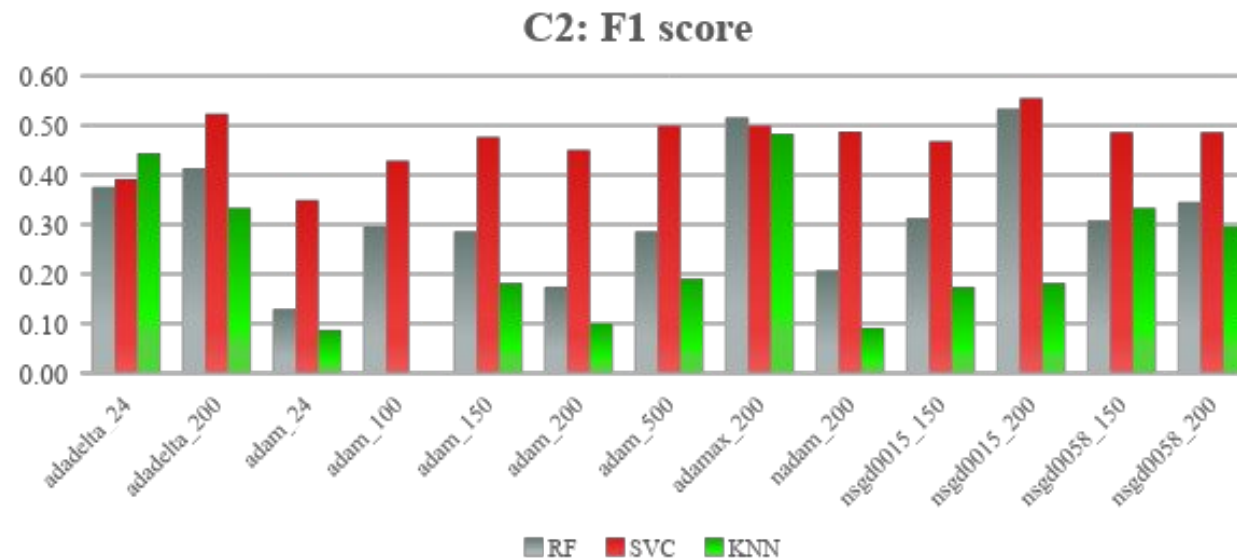
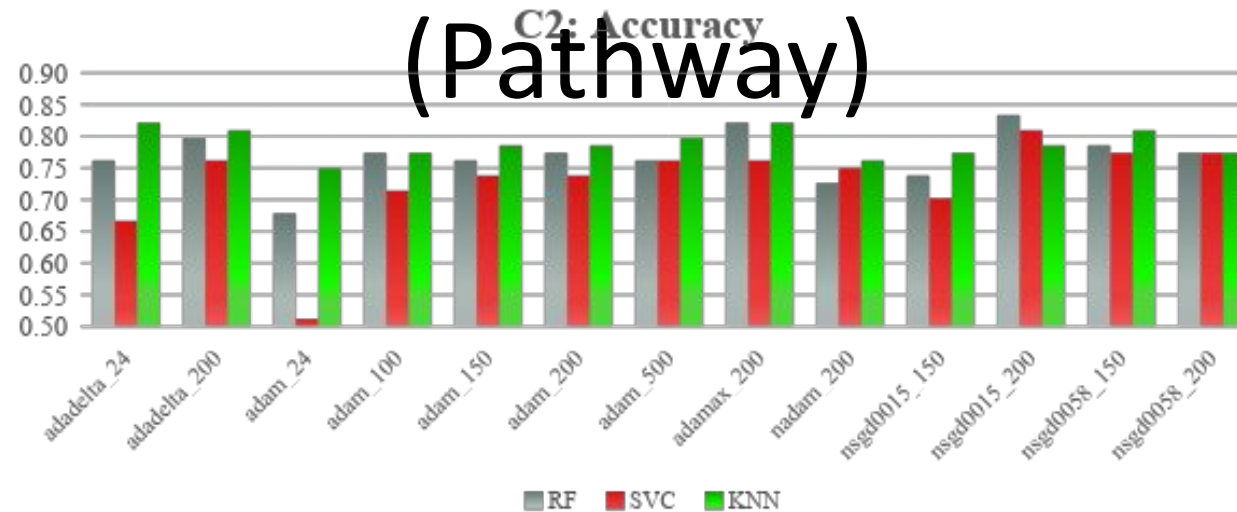
# Results – Classification (Taxonomy)



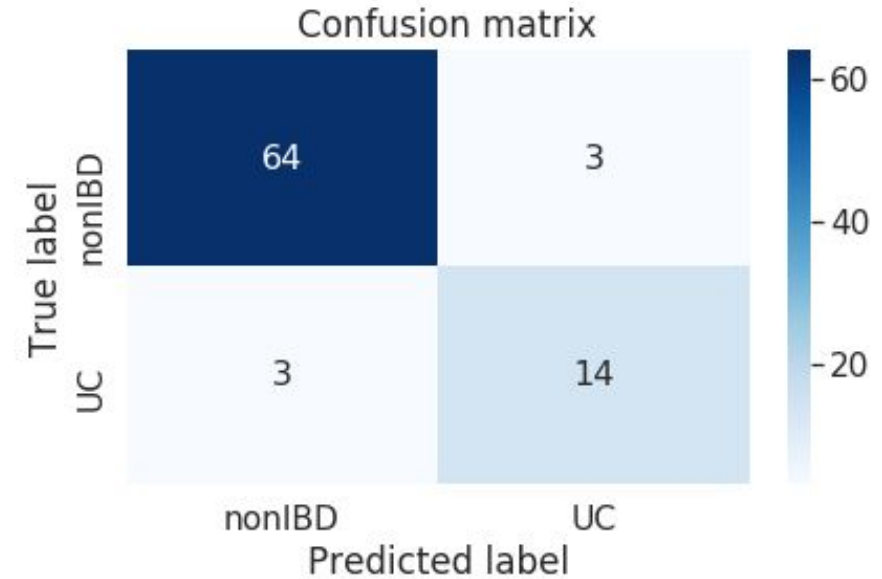
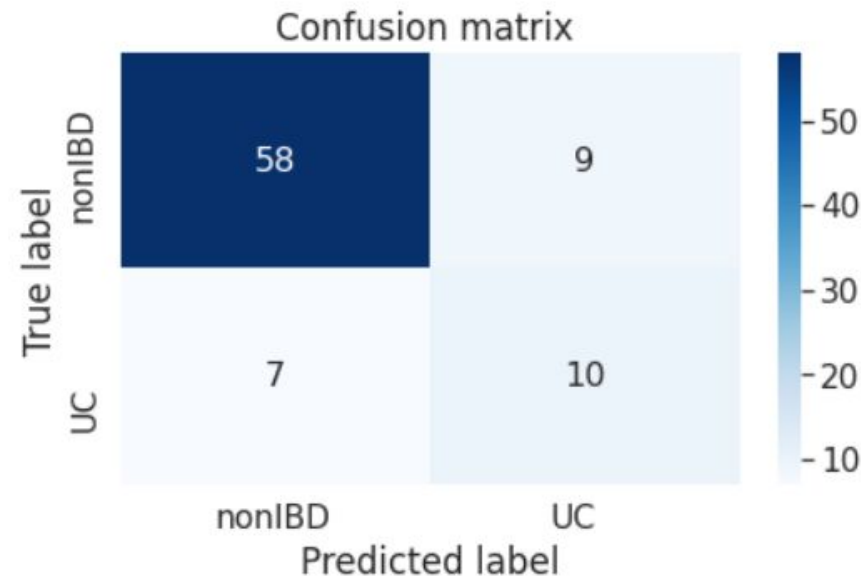
# AE vs ML (Taxonomy)



# Results – Classification (Pathway)



# AE vs ML (Pathway)



# Ideas for Future work

- Testing different parameters
- Classification with multiple layers
- Tanh
- More layers
- Regularization
- Sparse AE
- Semi supervised
- Ensemble
- Dropout





***Thank you  
for attention!***