National Institutes of Health
*Turning Discovery Into Health*

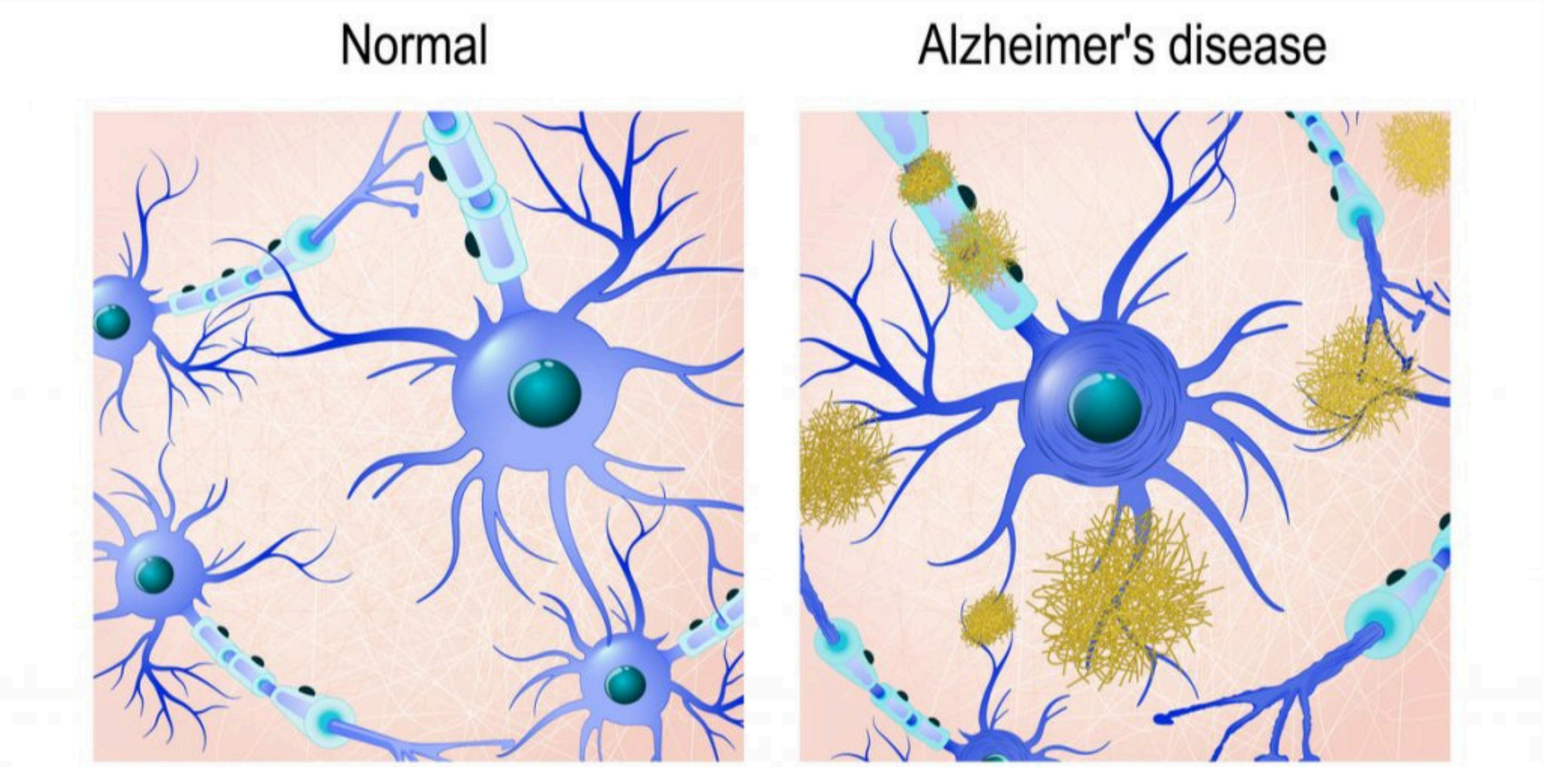
# Urinalysis Unlocking the Next Phase of Alzheimer's Diagnosis


Presentation by: Andranique Green

# Table
# Of Content

# Subject Matter
# Background

Alzheimer’s Disease is caused by the build up of abnormal proteins




collected from Aviv Clinics

# Problem Statement

## Alzheimer’s Disease

It is estimated that Alzheimer’s related dementia accounts for 70% of total dementia cases and number of people living with the disease is steadily rising. There are various tests that assist in the diagnosing process; unfortunately the majority of these tests tend to be inaccessible, invasive, and expensive. This project aims to change this fact.

## How

The goal of this project is to identify biomarkers for detecting Alzheimer’s dementia from urine and create a model that can correctly classify between individuals with Alzheimer’s and those without. Which reduce pain and cost associated with current testing, and lower accessibilty barriers

# Research and Development Process

# Experimental Data:

- Niigata University Graduate School of Medical and Dental Sciences
- Urine samples were collected from 36 individuals sex and aged matched
- The samples were analysis by mass spectrometry
- Results: There were changes in 109 proteins

**Healthy Control**

Noted as 0 in the cleaned dataset. Cognitively normal individuals

**Alzheimer's Sample Group**

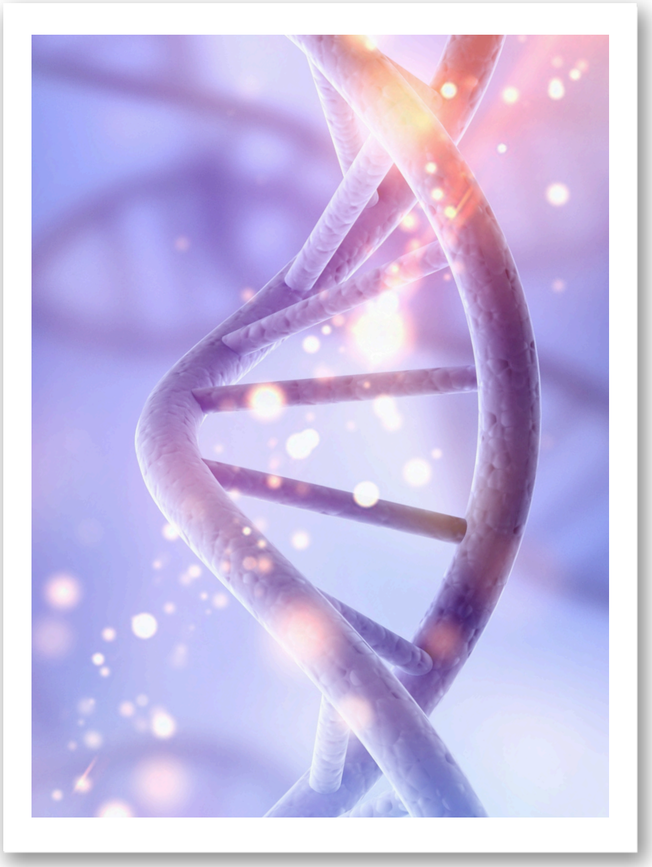
Noted as 1 in the cleaned dataset. Individuals Experiencing Alzheimer's related demetia

# Data
# Limitations

The experimental data I have was only partially completed.

While I had very comprehensive data on a singular patient there are still 36 total samples.

There isn’t much we can do about the dataset but being that it is early in test development we can use the results as a guiding path for research advancement



# Data Handling

The individual sample data was very extensive and the reports formatting made it very hard to extract.

For this reason I relied specifically on emPAI values or estimated protein ratios based on peptide presence.

## Data Included :

| |
|---|
| protein_desc- the names of proteins present |
| emPAI value |
| Alzheimer’s- Target column |

# Binary
# Classification Models

Keeping the data limitation in mind I used a simple machine learning model to prevent overfitting. I ended up comparing models based on their accuracy score solely to identify how well they could differentiate based on protein values without preference for TP and TN. With later models metrics like sensitivity would be better suited to meet the problem's needs.

I chose logistic regression for model and I built two versions. One without added penalties and another using ElasticNet. The models performed the same regardless of the penalties

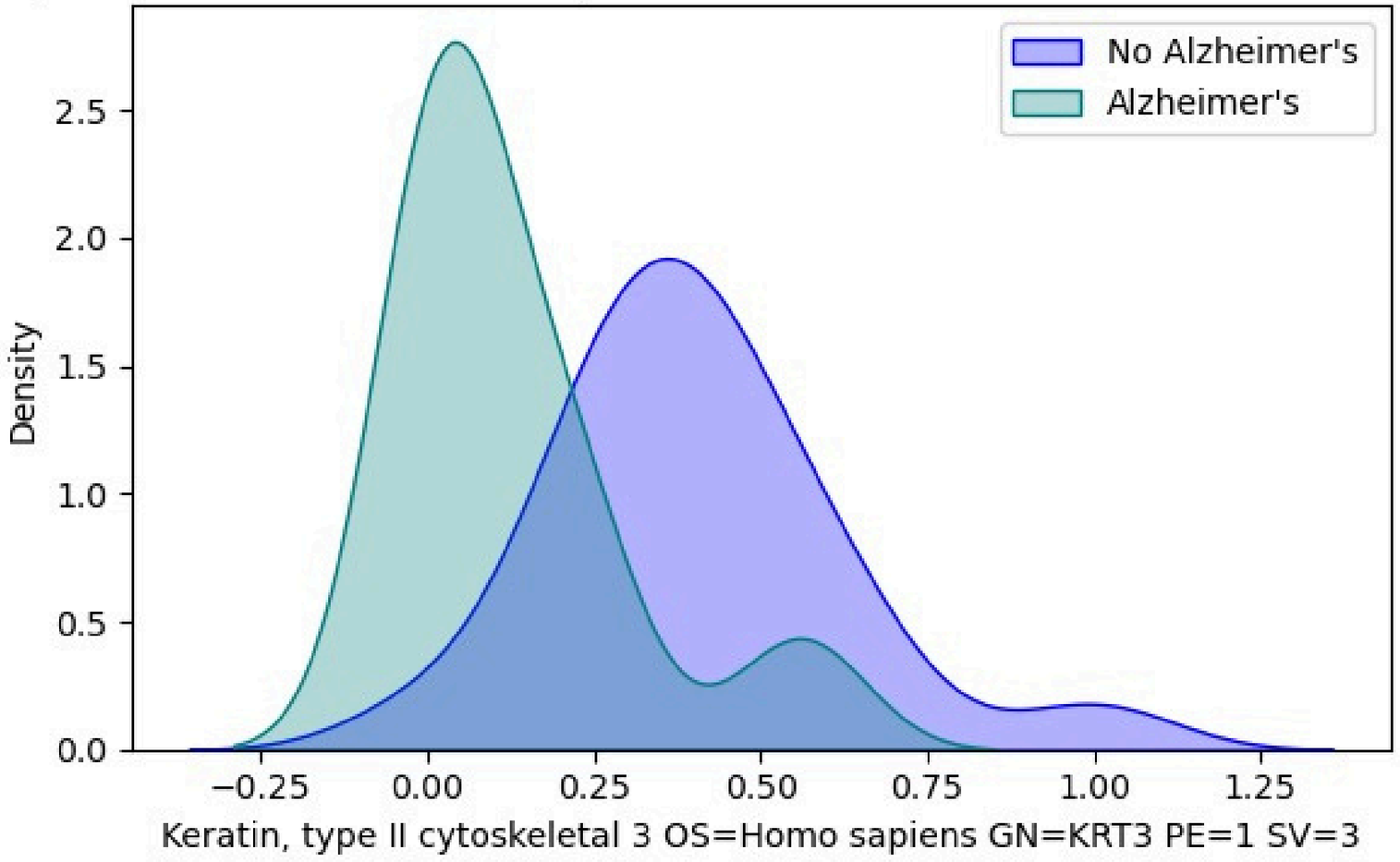
## Target Protein Group

Extracted through calculating mean variance between groups

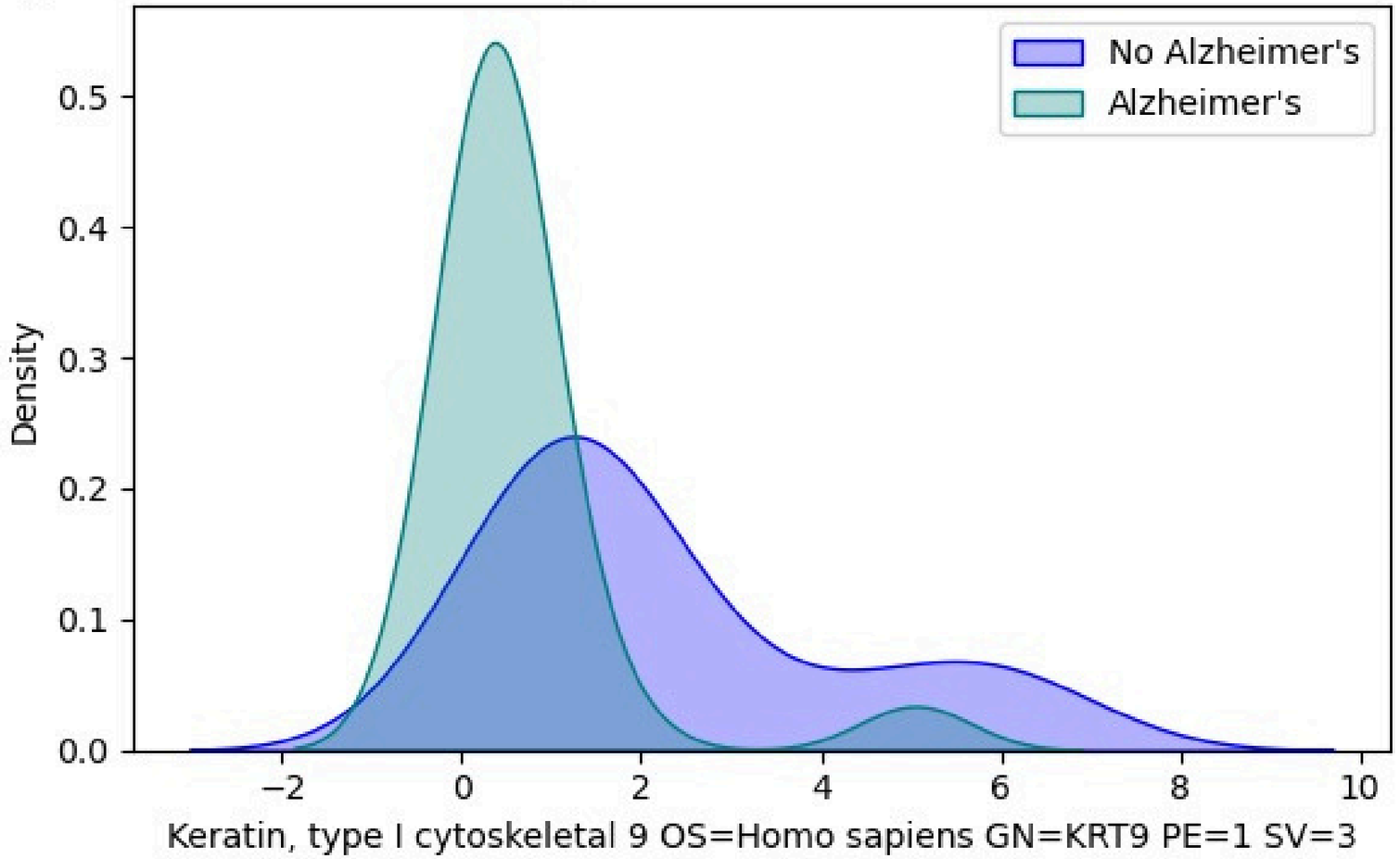
## Important Features Group

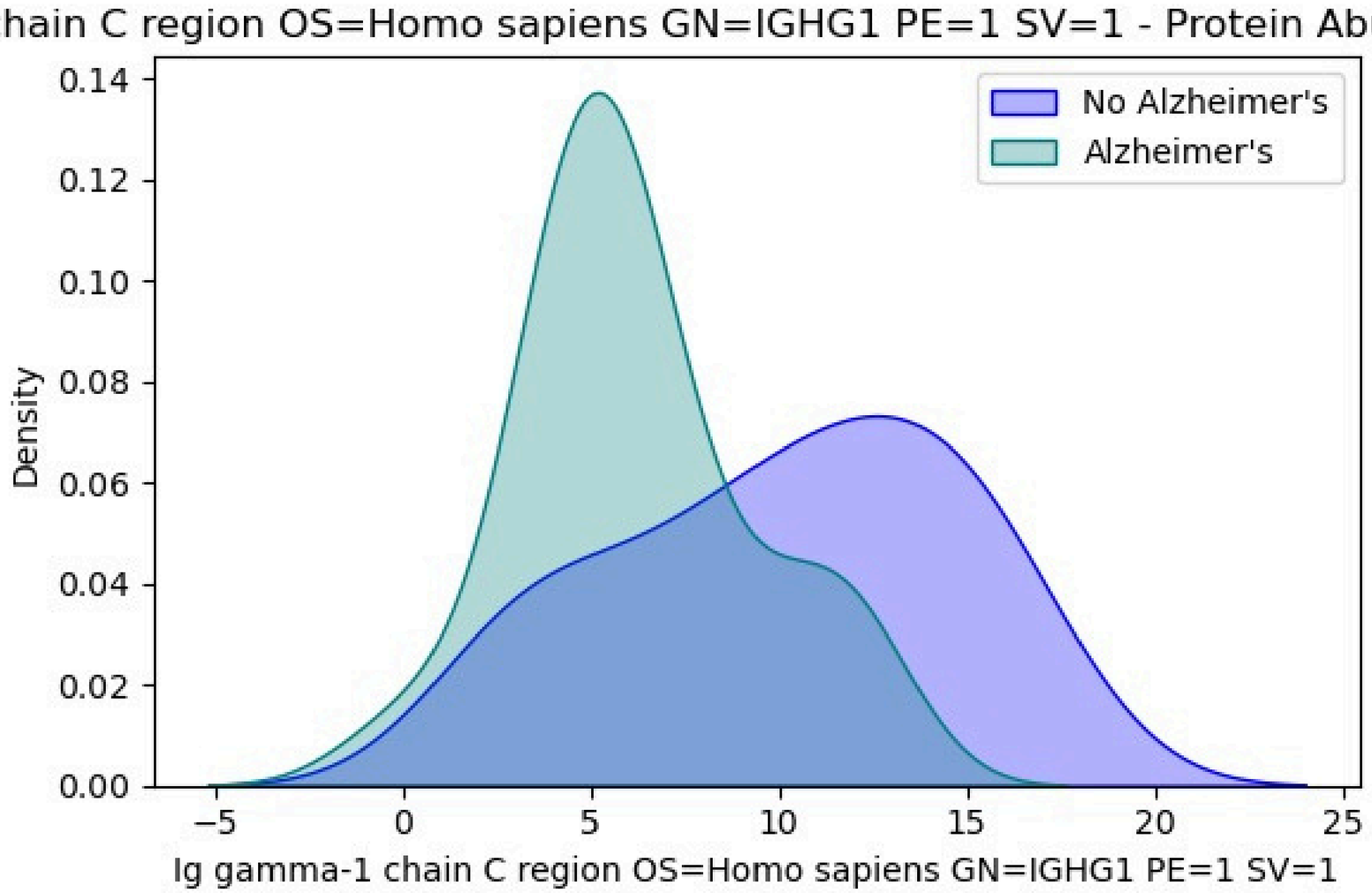
Decision Tree importance assigned proteins

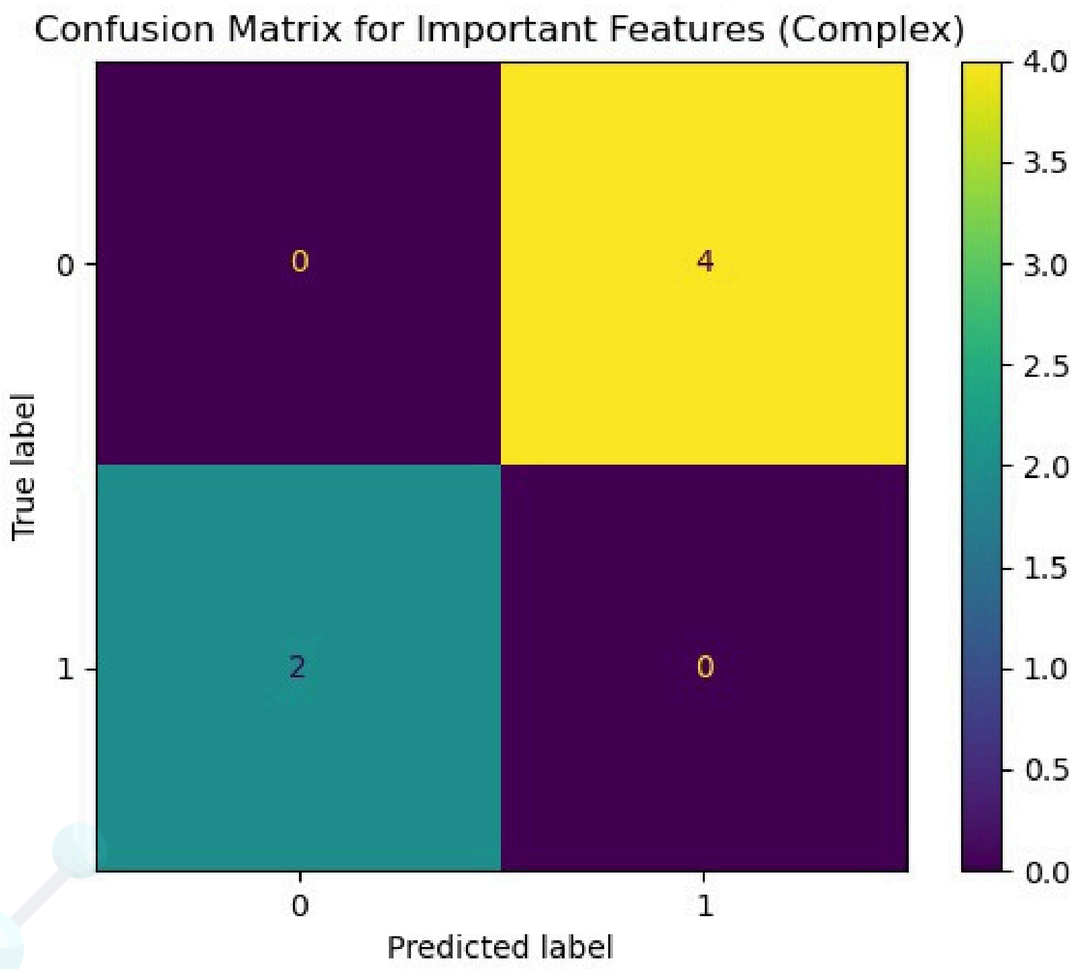
Keratin, type II cytoskeletal 3 OS=Homo sapiens GN=KRT3 PE=1 SV=3 - Protein Abundance Distribution

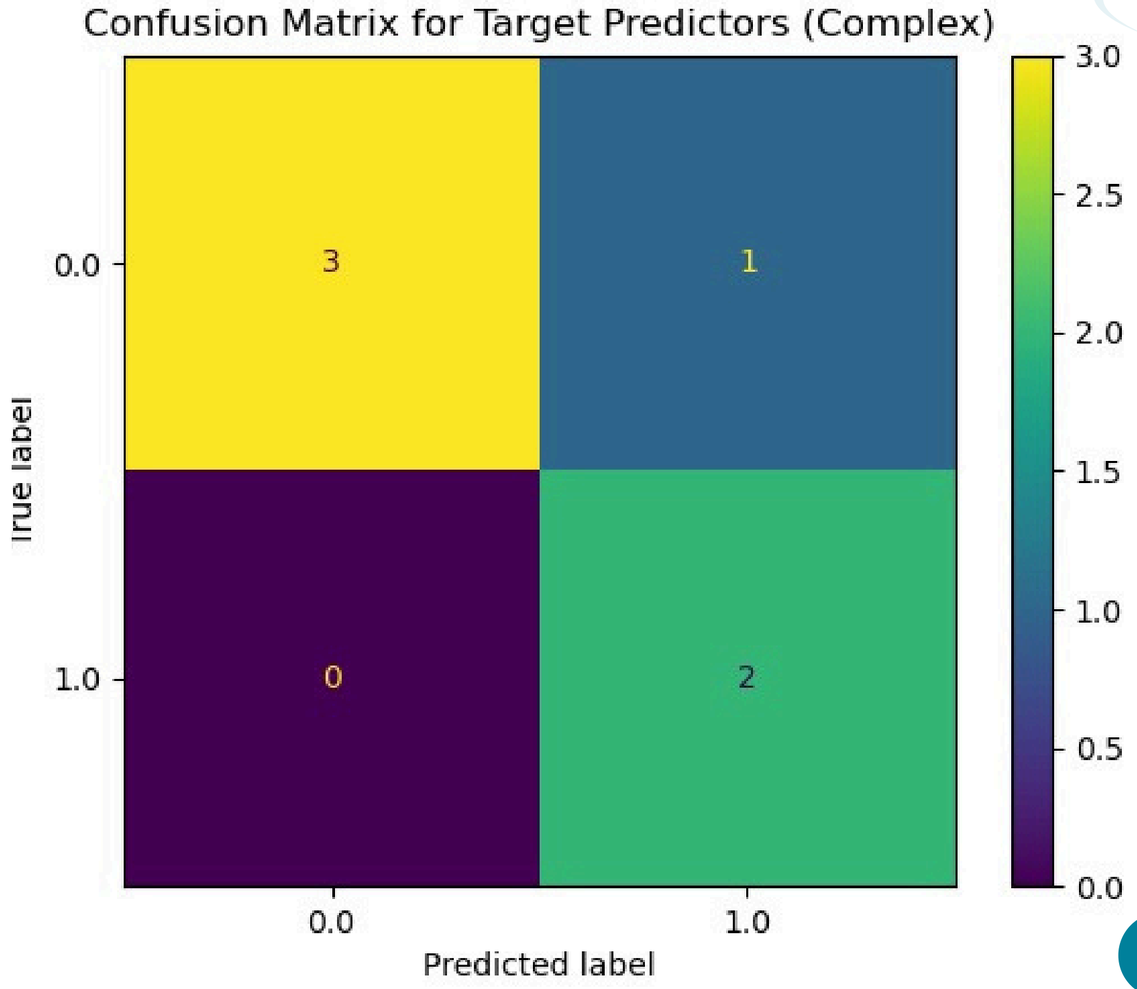


Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=3 - Protein Abundance Distribution




Ig gamma-1 chain C region OS=Homo sapiens GN=IGHG1 PE=1 SV=1 - Protein Abundance Distribution
No Alzheimer's
Alzheimer's
Density
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
−5
0
5
10
15
20
25
Ig gamma-1 chain C region OS=Homo sapiens GN=IGHG1 PE=1 SV=1


Confusion Matrix for Important Features (Complex)
True label
Predicted label
0
1
0
4
2
0
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0


Confusion Matrix for Target Predictors (Complex)
True label
Predicted label
0.0
1.0
3
1
0
2
3.0
2.5
2.0
1.5
1.0
0.5
0.0

# Conclusion

Move forward with additional research on the 3 proteins identified by decision trees

Metrics:
The Logarithmic model trained on the Decision Tree features had a 100% accuracy rate with testing and 87.5% with the training data

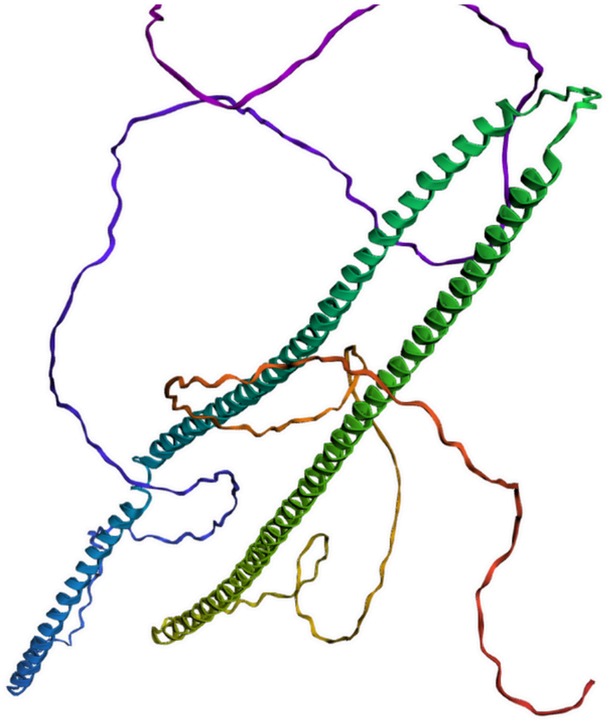
# R&D

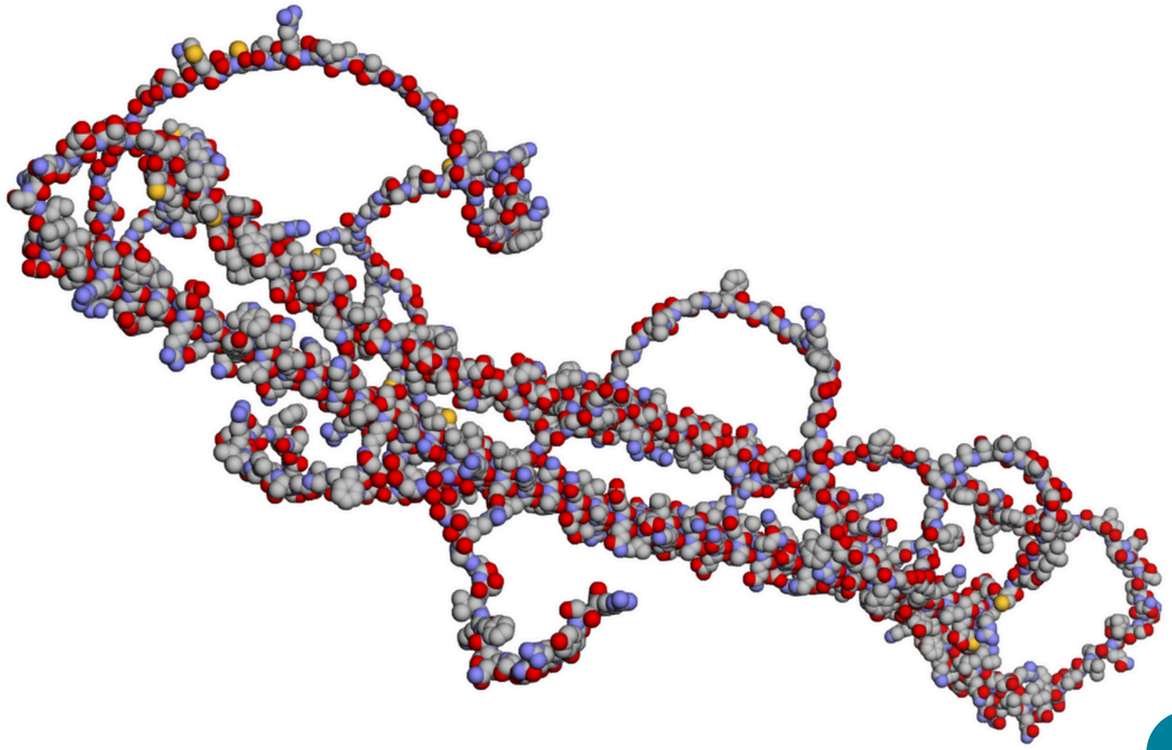
## Next Steps

- Expand Dataset:
- Collect Repeated Samples over the Disease Progression:
- Incorperate More Information Available from CSV Report:
- Employ Decision Curve Anaylsis and Patient Stratification:

# Protein

## Visualizations: Keratin, type II





# Thank You
# for Listening