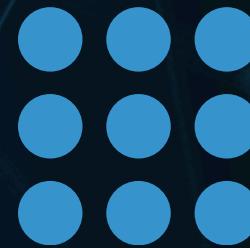
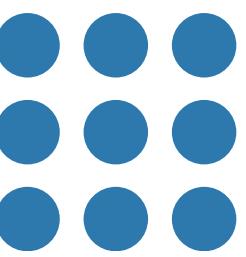


Toward Early Detection of Parkinson's Disease

Integrating Novel Proteomic
Biomarkers for Predictive Analysis

by Joan Jaylani
Priyanka Gowd Mitta M.D.



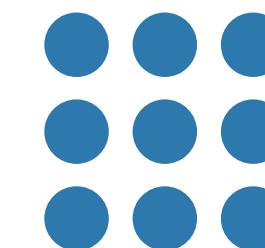


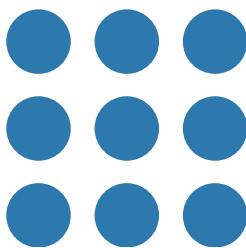
Background

Parkinson's Disease (PD) affects 1-2% of the population. *

Traditionally diagnosed after motor symptoms appear. *

Early neuronal loss often occurs before symptoms. *

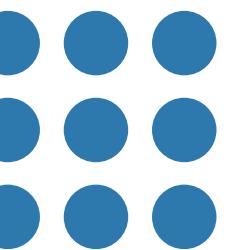


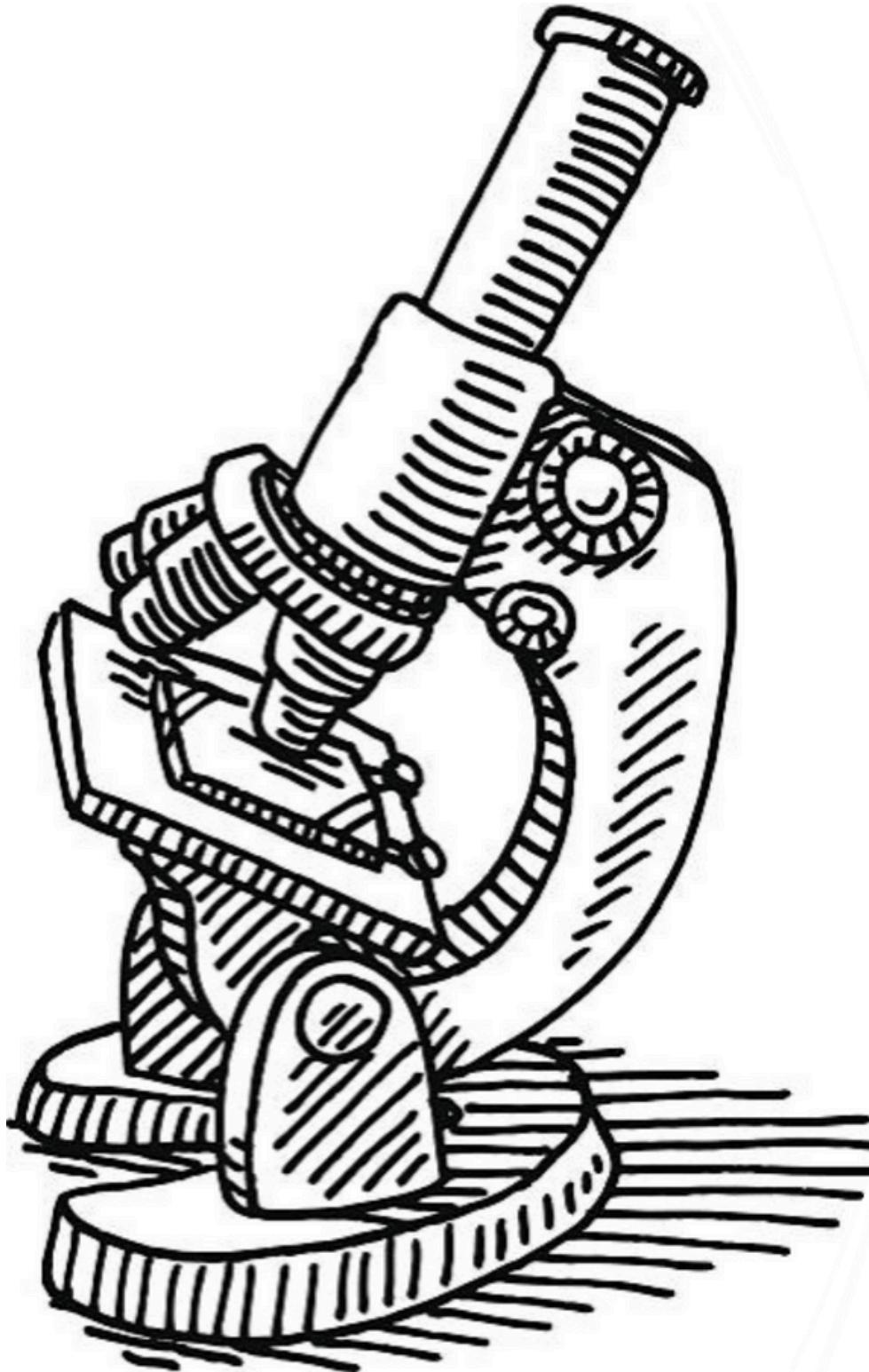
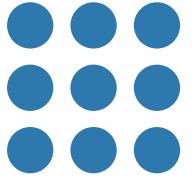


Motivation for Biomarker Research

Enhancing Diagnostic Accuracy Content

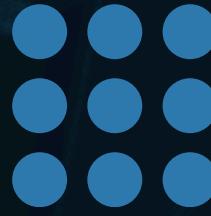
- * Current diagnostic methods rely on clinical observation and DaTscan imaging.
- * These approaches are reactive, not predictive.
- * Biomarkers—especially proteins, peptides, and metabolites—may indicate PD before motor symptoms emerge.





Study Objective

To evaluate the predictive potential of proteomic biomarkers—including peptides, proteins, and metabolites using machine learning for early detection of Parkinson's disease.



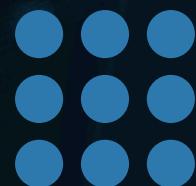
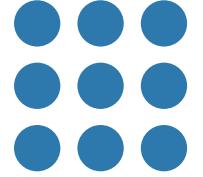


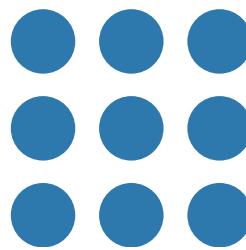
Parkinson's
Progression
Markers
Initiative



Data Sources

- * Parkinson's Progression Markers Initiative (PPMI) datasets
 - SAA_Biospecimen_Analysis.csv (alpha-synuclein, n = 1,586)
 - Biospecimen_Analysis_Results.csv (proteomics, > 487k rows, n = 262, unique biomarkers 2,958)
- * Data collected from 33 global research sites (2014-2024).





Machine Learning Methodology Overview

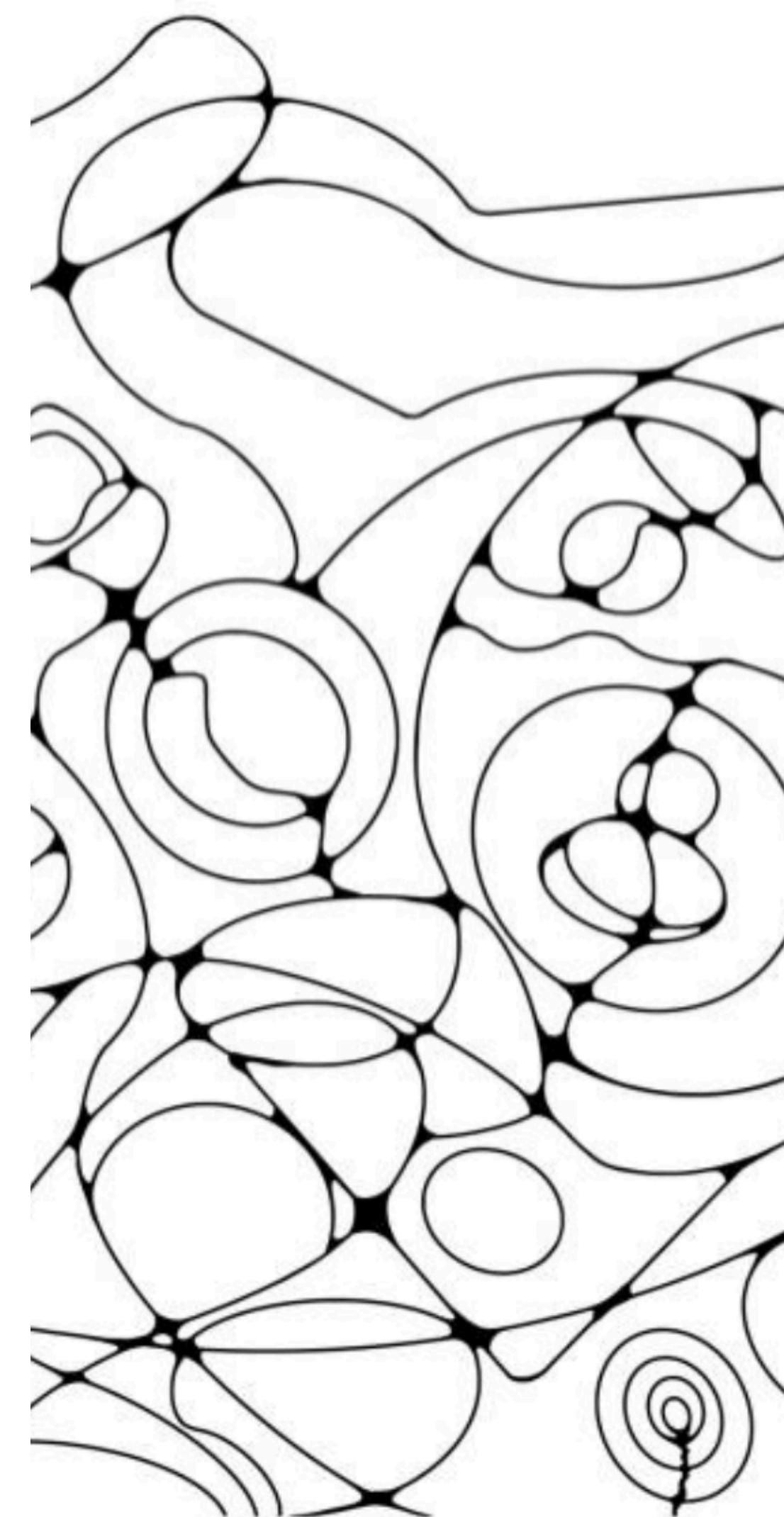
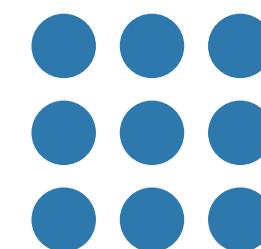
* Preprocessing

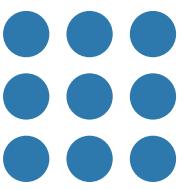
- Structured by patient ID and biomarker test type.
- Reviewed handling outliers, missing data

* Models Tested

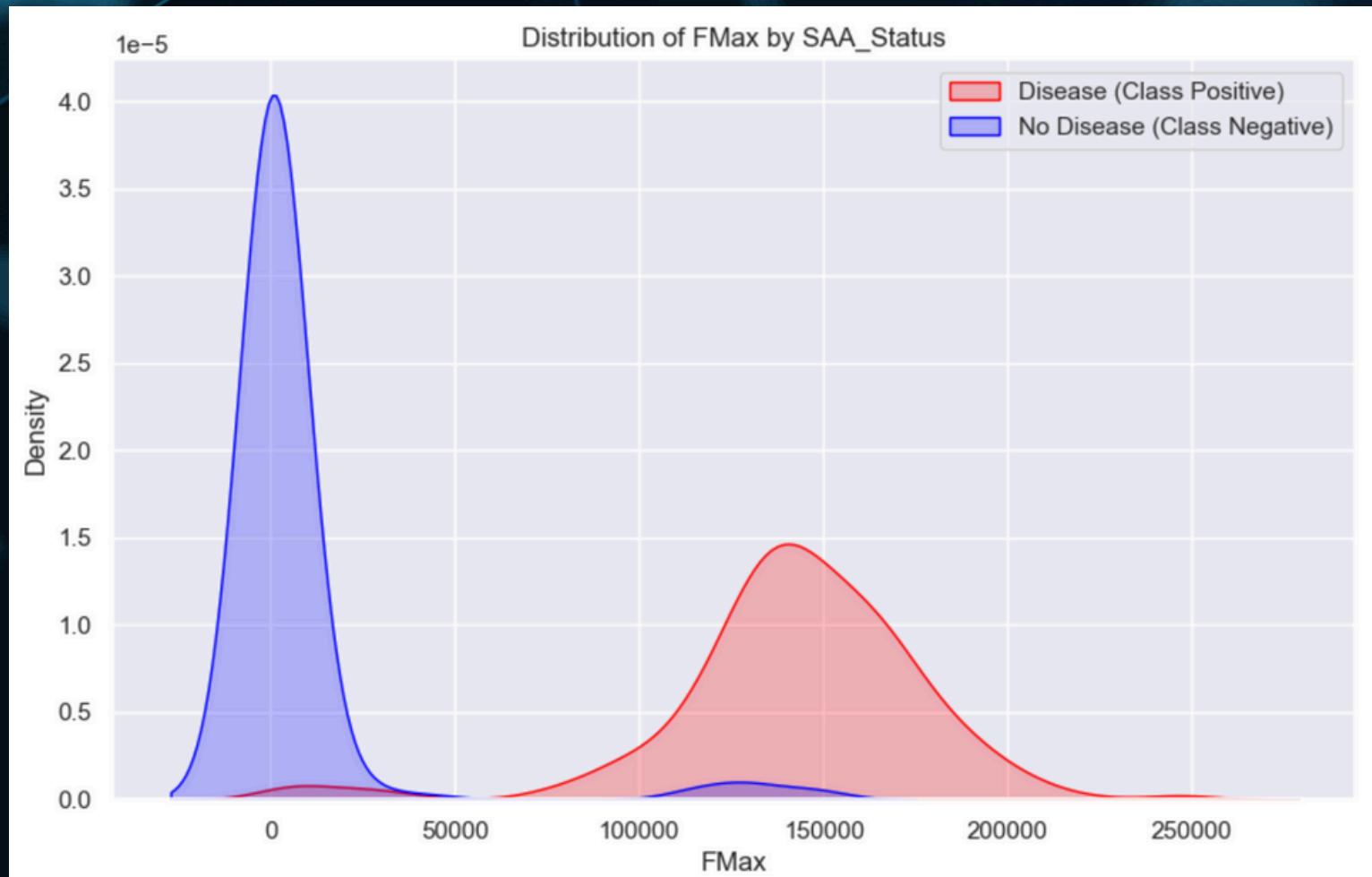
- Logistic Regression
- Random Forest
- Support Vector Classifier
- XGBoost

* 10-fold Stratified Cross-Validation for robustness



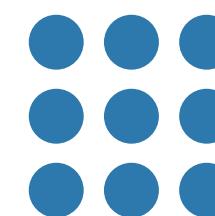


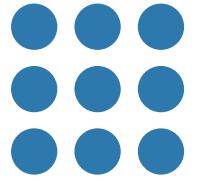
Maximum Fluorescence Threshold Value for Alpha Synuclein



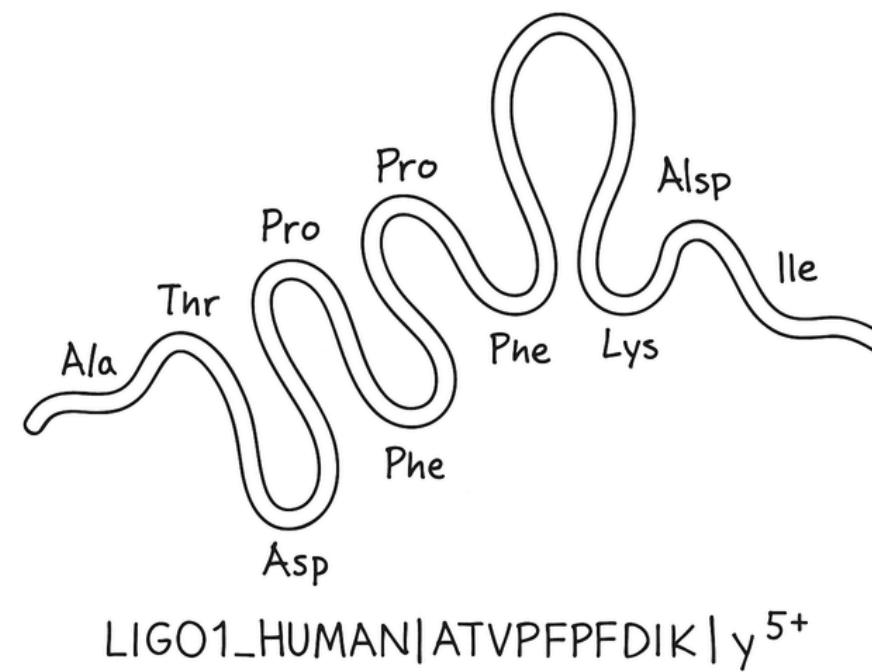
- Alpha-synuclein misfolds into Lewy bodies in PD. *
- Seed Amplification Assay (SAA) predicts PD with ~97% accuracy (Russo et al., 2021). *
- Logistic regression model replicated this result. *

- Accuracy: 95.8%
- Recall (TP): 96.6%
- Recall (FN): 3.4%
- Result Variability: $\pm 2.6\%$





Proteomic Feature Engineering

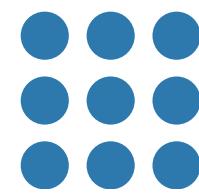


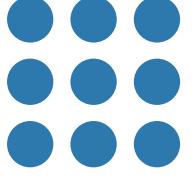
Reshaped dataset to patient × biomarker (2,958 cols). *

Focused on 735 peptide features. *

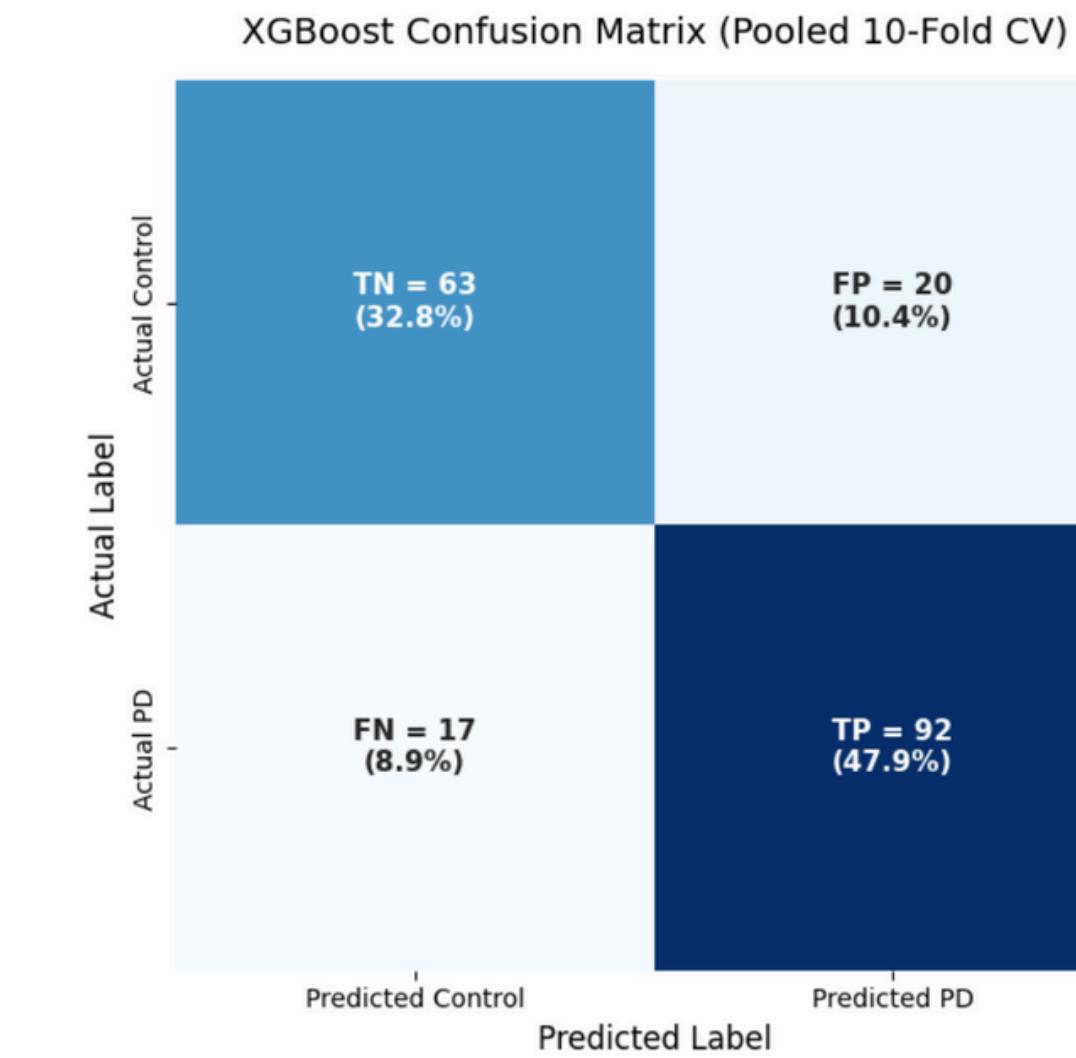
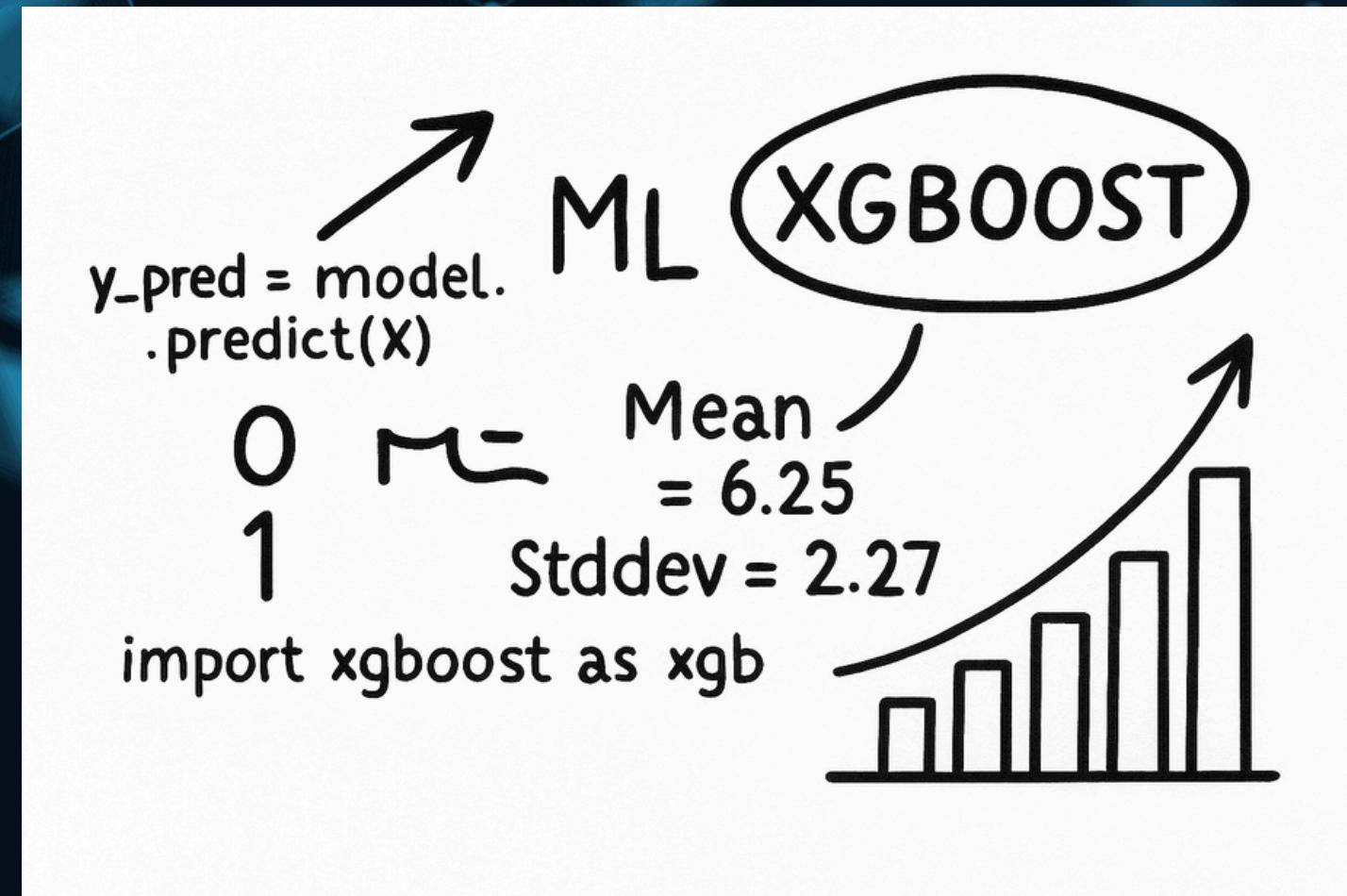
Feature selection → top 30 biomarkers. *

XGBoost used for optimized classification. *



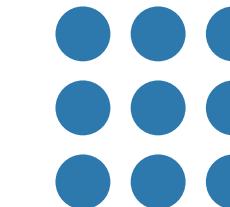


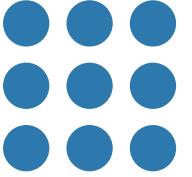
XGBoost Classification Results



Overall Accuracy = 80.7%

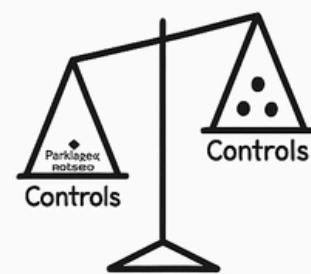
PD (Class 1): Precision = 82.1%, Recall = 84.4%, F1 = 83.3%
Control (Class 0): Precision = 78.8%, Recall = 75.9%, F1 = 77.3%



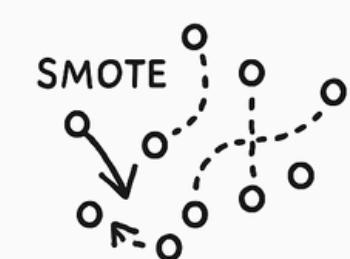


Study Limitations

STUDY LIMITATIONS



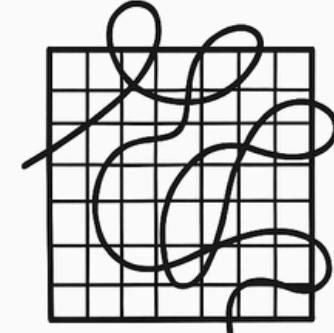
Data imbalance



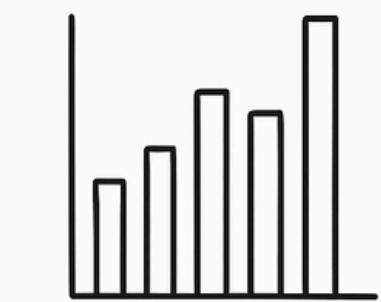
Synthetic data



Limited diversity



High dimensionality



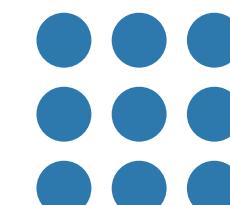
Variable sensitivity

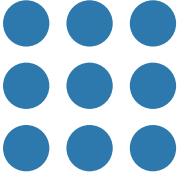
Imbalanced alpha-synuclein data (>80% PD+). *

Small proteomic sample size. *

Limited population diversity (mainly European ancestry). *

Some biomarkers lack disease specificity. *





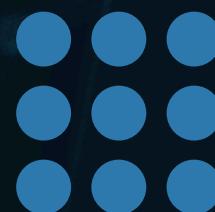
Clinical Implications

Early detection → timely intervention. *

Blood-based biomarkers could enable routine screening (≥ 65 years). *

Could identify PD up to 7 years before symptoms. *

Barriers: assay standardization, cost, regulation. *



Parkinson's Disease

Threshold Analysis Application



Instructions

Upload thresholds CSV

Drag and drop file here
Limit 200MB per file • CSV

Browse files

XGB_top_5_thresholds_extracted.csv 293.0B

X

Inputs

Homovanillic_acid

554.66045

LIGQ1_HUMAN|ATVPFPFDIK|y5+

0.22390

- +

CCKN_HUMAN|AHLGALLAR|y6+

0.69400

CBLN1_HUMAN|STFIAPR|y3+

1.03670

- +

Cysteine

4.27942

- +

Compute probability (threshold score)

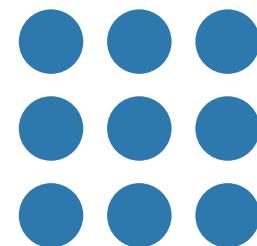
Threshold score (weighted fraction of hits)

0.843

Estimated PD probability

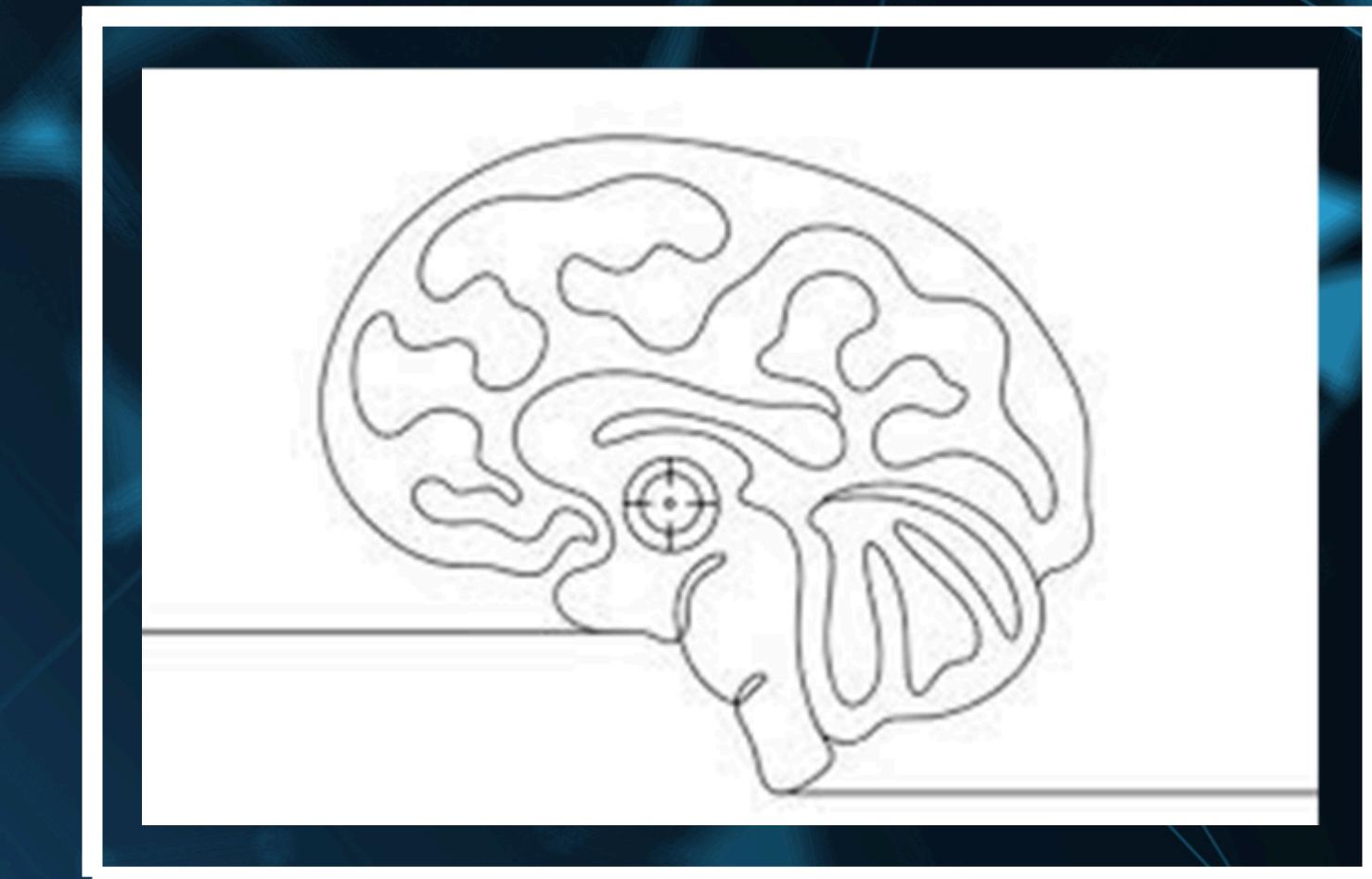
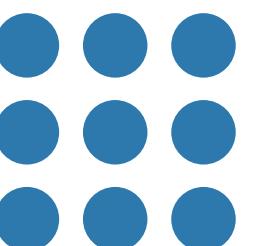
0.699

	biomarker	value	threshold	direction	match	weight
0	Homovanillic_acid	554.6604	64.8824	>	<input checked="" type="checkbox"/>	0.0946
1	LIGQ1_HUMAN ATVPFPFDIK y5+	0.2239	0.3365	<	<input checked="" type="checkbox"/>	0.0541
2	CCKN_HUMAN AHLGALLAR y6+	0.694	0.938	<	<input checked="" type="checkbox"/>	0.0563
3	CBLN1_HUMAN STFIAPR y3+	1.0367	0.9925	>	<input checked="" type="checkbox"/>	0.0144
4	Cysteine	4.2794	5.4819	>	<input type="checkbox"/>	0.0408



Conclusion & Future Work

- * Proteomic biomarkers show strong potential.
- * Alpha-synuclein + multi-marker panels enhance accuracy.
- * Future steps:
 - Larger, diverse validation cohorts
 - Automate mass spectrometry
 - Integrate thresholds into diagnostic tools.



Thank You

