

Pro-T-Omics: A Tool for Biomarkers Discovery

Alex Montoya Franco, Faridun Mamadbekov, Ikram UI Haq, Mudita Shakya

Biomarkers for early detection?

Did you know that most people diagnosed with Parkinson's are 60 years or older?

There is no single definitive test for diagnosing Parkinson's Disease (PD). Symptoms vary from person to person, and they are similar to other neurological disorders.

By the time a diagnosis is made, PD has typically already progressed to a point of no return.

Biomarkers

Detectable and measurable changes in the body that can be used to predict, diagnose, and monitor disease activity and progression.

So, aren't there ways to diagnose it early?

Currently, there are none. But there is hope. Research is focused on the discovery of Biomarkers.

Want to know more? Keep reading

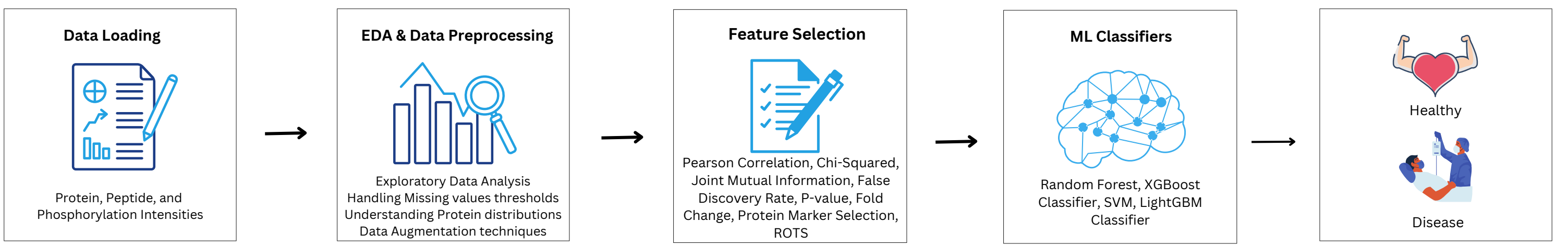
Introduction

This Research Project uses protein intensities extracted from urine and blood samples. We aim at identifying significant protein features that could potentially be biomarkers. These features are then used to classify between two groups, for instance: healthy and Parkinson's patients, enabling early detection.

Objectives

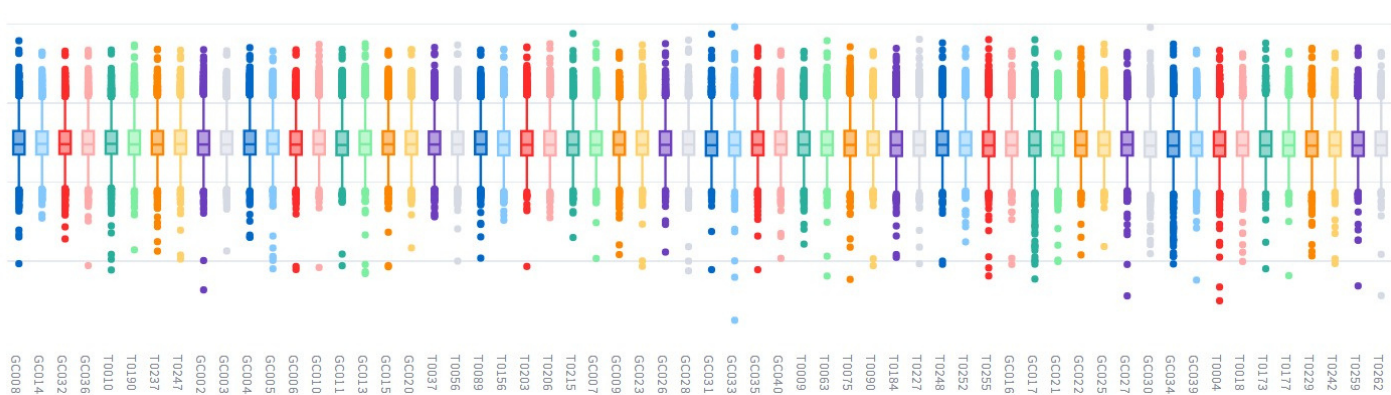
- Discovery of significant Protein Biomarkers.
- Early Diagnosis of diseases.
- Building an ML-based pipeline for analysis and predictions.
- Delivering a user-friendly interface.

Pipeline

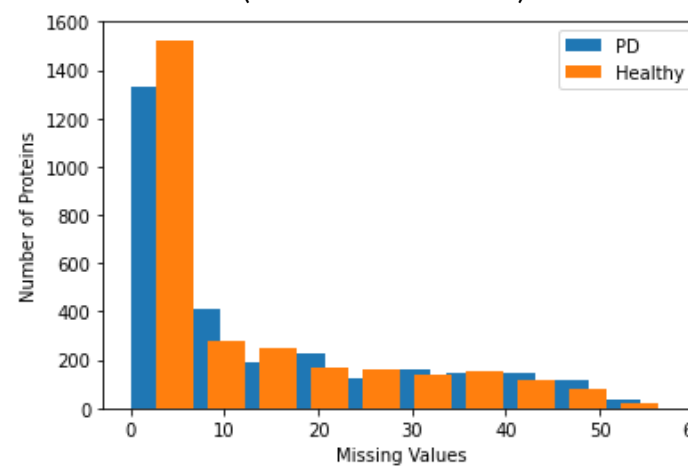


Exploratory Data Analysis and Data Preprocessing

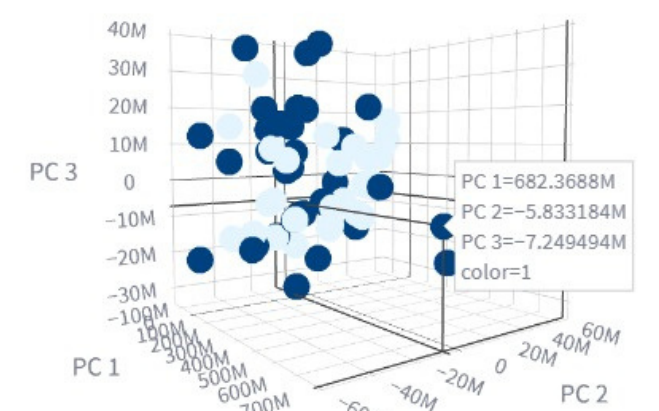
Box plot of Log₂ intensity distribution of all samples



Distribution of Missing Values in PD vs Healthy (out of 4000 Proteins)



Principal Component Analysis (PCA) plot of samples



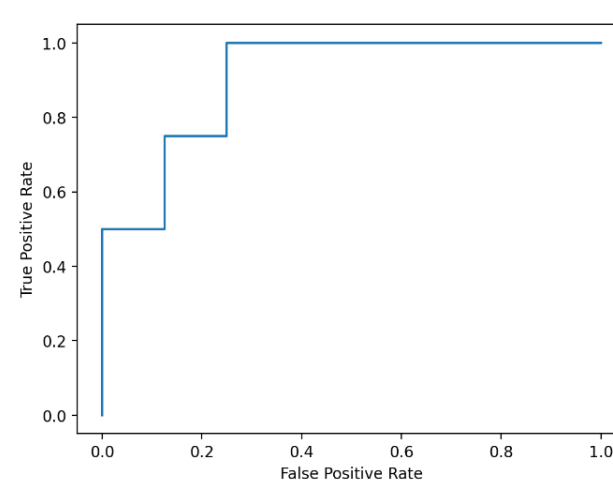
Feature Engineering

Selection of Protein Biomarkers based on multiple statistical and Machine Learning based feature selection techniques.

Sample ID	Pearson Correlation	PG.ProteinGroups	PG.Genes	PG.ProteinDescriptions	Chi Square Selection	Joint Mutual Information	Fold Change	P-value	False Discovery Rate	total
0	5,966	Q9UNX4	WDR3	WD repeat-containing protein 3						4
1	5,684	Q9NWJ2	ARL8B	ADP-ribosylation factor-like protein 8B						4
2	5,217	Q9BS26	ERP44	Endoplasmic reticulum resident protein 44						4
3	5,086	Q96PK6	RBM14	RNA-binding protein 14						4
4	4,847	Q92522	H1-10	Histone H1.10						4
5	4,671	Q8N8X0	SCCPDH	Saccharopine dehydrogenase-like oxidoreductase						4
6	4,354	Q6VP21;Q6VP21-3	KYAT3	Kynurenine-oxoglutarate transaminase 3/isoform 2						4
7	3,893	Q14684	RRP1B	Ribosomal RNA processing protein 1 homolog B						4
8	3,841	Q14116;Q14116-2	IL18	Interleukin-18/isoform 2 of Interleukin-18						4
9	3,693	Q12778	FOXO1	Forkhead box protein O1						4
10	3,123	P42574	CASP3	Caspase-3						4
11	3,017	P33241	LSP1	Lymphocyte-specific protein 1						4
12	2,937	P28838;P28838-2	LAP3	Cytosol aminopeptidase/isoform 2 of Cytosol amin						4

Results

Random Forest Classifier Blood Protein Data Accuracy: 83.33%



Random Forest Classifier Urine Protein Data Accuracy: 75.0%

