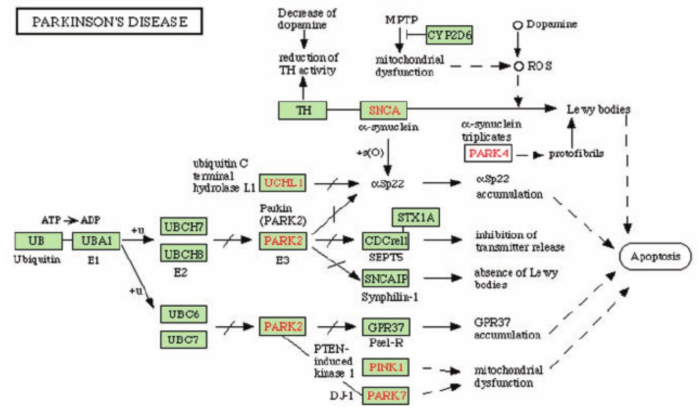
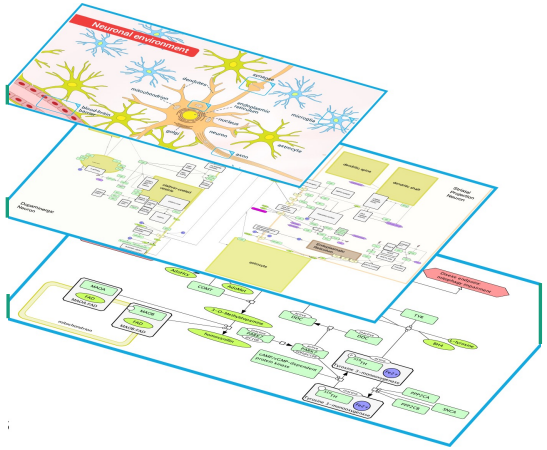
DIGIPD: Linking digital markers to molecular markers in Parkinson‘s disease

DIGIPD

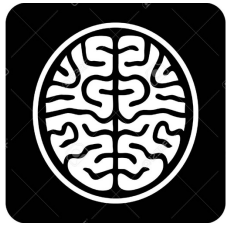
Validating DIGItal biomarkers for better personalized treatment of Parkinson’s Disease

- **Mapping of molecular markers to disease mechanisms**





- **Dependencies between molecular mechanisms, DMs, and other data types**



Fraunhofer SCAI · Paris Brain Institute ICM · uni.lu Université du Luxembourg · Universitätsklinikum Erlangen · Télécom SudParis · Institut Polytechnique de Paris · Université de Namur · Portabiles HealthCare Technologies · Asociación Parkinson Madrid

Linking DMs to molecular markers & other data types

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- **Overview**:
  - **Statistical analyses**:
    - Pre-processing and filtering of clinical, miRNA/mRNA, epigenetics, metabolomics & digital gait data
    - Differential expression/abundance analysis for miRNA (+ target prediction), mRNA and epigenetics
    - Pathway enrichment analyses for miRNA, mRNA, epigenetics and metabolomics data
    - Correlation analysis: gait data vs. pathway omics activities (metabolomics & miRNA)
  - **ML analyses**:
    - Patient unsupervised clustering analyses at baseline (clinical data)
    - Supervised cross-validation analysis for ML-based PD vs. Control prediction (all data types)
    - Supervised cross-validation analysis of UPDRS 3 total motor scores (all data types)
    - Prediction of other non-motor symptoms / co-morbidities (baseline & future visits, all data types)

Common themes in pathway enrichment analyses

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Overlap across different analyses:

Shared pathways	**Top-ranked for which omics data / pathway databases**
Endoplasmic reticulum stress	PD-map (blood RNAseq, iPSC RNAseq, brain scRNAseq), GO (blood RNAseq)
Mitochondrion / oxidative stress	PD-map (blood RNAseq, iPSC RNAseq, brain scRNAseq, brain proteomics)
Cell death / apoptosis	PD-map (iPSC RNAseq), KEGG (iPSC RNAseq), KEGG (blood RNAseq)
Glycosylation-related pathways	GO (iPSC RNAseq), KEGG (iPSC RNAseq), Reactome (iPSC RNAseq), KEGG (miRNA)
MAPK signaling	KEGG (blood RNAseq), BioCarta (blood RNAseq), NeuroMMSig-PD (blood RNAseq)
Ribosomal pathways	GO (blood RNAseq), KEGG (iPSC RNAseq)
Nonsense mediated mRNA decay	GO (blood RNAseq), Reactome (blood RNAseq), PD-map (brain proteomics)
MTOR/FOXO signaling	PD-map (blood RNAseq), PD-map (iPSC RNAseq), KEGG (miRNA)
Synaptic vesicle pathways	PD-map (brain scRNAseq, brain proteomics), GO (iPSC RNAseq)

Linking markers to pathways: Pathway-based prediction

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Prediction of PD vs. control using **median** GO set expression (min. 10 mappable genes, 3-fold CV, random equal-sized partitions, iPSC RNAseq RNA-seq)

AUC statistics:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GBM	0.545	0.549	0.552	0.608	0.640	0.727
SVM	0.375	0.528	0.682	0.580	0.682	0.682
ROTF	0.604	0.615	0.625	0.667	0.699	0.773
XGB	0.545	0.616	0.688	0.691	0.764	0.841
DEEP	0.705	0.750	0.795	0.795	0.840	0.885
RF	0.385	0.471	0.557	0.579	0.676	0.795
GAUS	0.188	0.412	0.636	0.517	0.682	0.727
LBOOST	0.344	0.501	0.659	0.611	0.744	0.830

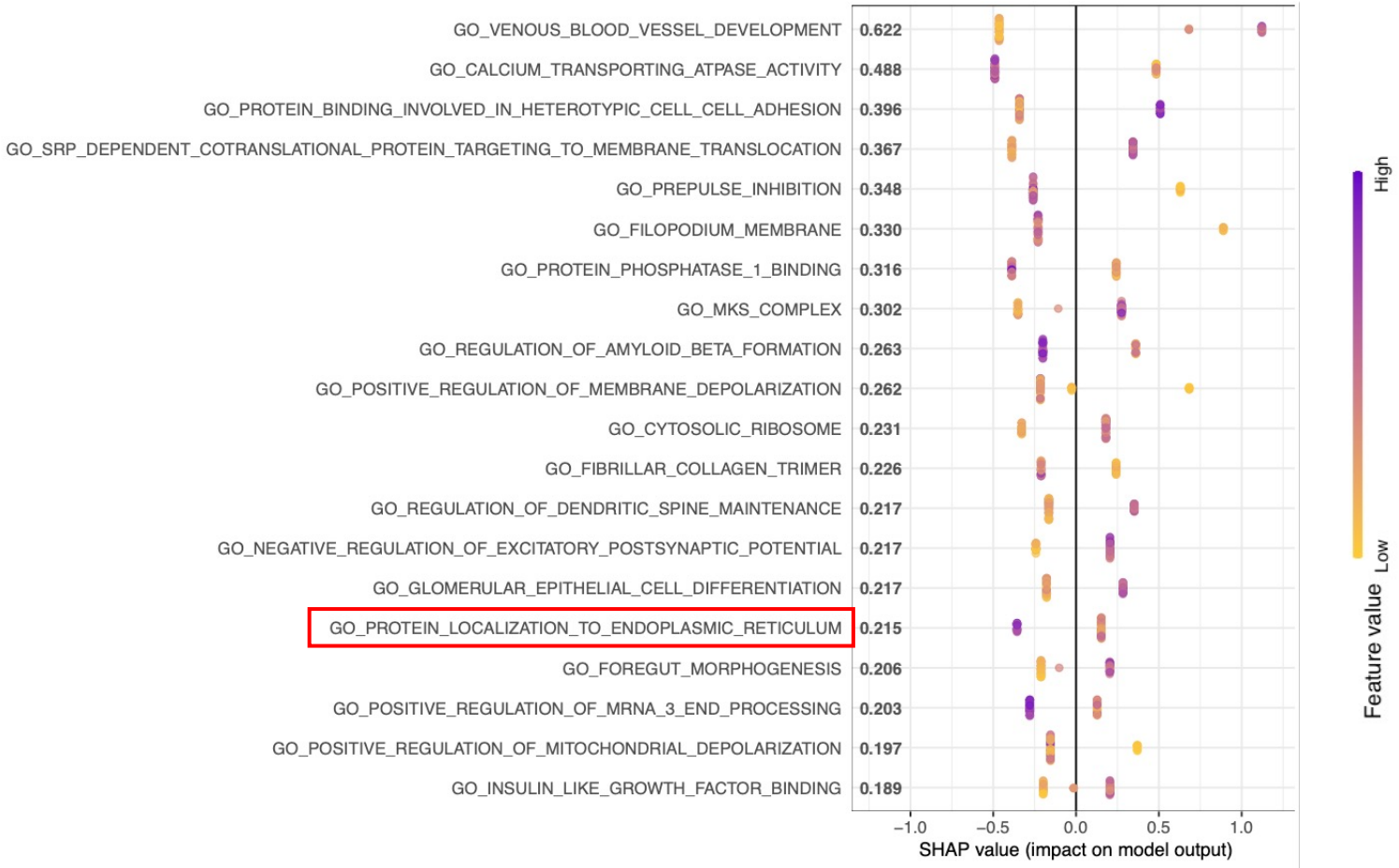
Median AUCs of 0.6 to 0.8 across different ML algorithms

PD vs. control prediction – feature impacts on outcome

DIGIPD

Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Shapley additive explanations (SHAP) value analysis of feature impact on outcome



Prediction of UPDRS 3 total score using pathways

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Prediction of UPDRS 3 sum > median using GO set expression (min. 10 mappable genes, 3-fold CV, random equal-sized partitions, iPSC RNAseq RNA-seq)

AUC statistics:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GBM	0.50	0.56	0.61	0.59	0.64	0.67
SVM	0.50	0.53	0.57	0.59	0.63	0.69
ROTF	0.40	0.55	0.70	0.62	0.73	0.75
XGB	0.42	0.56	0.70	0.65	0.77	0.83
DEEP	0.42	0.44	0.47	0.45	0.47	0.47
RF	0.50	0.50	0.50	0.51	0.52	0.53
GAUS	0.31	0.39	0.47	0.42	0.48	0.50
LBOOST	0.57	0.62	0.67	0.64	0.67	0.68

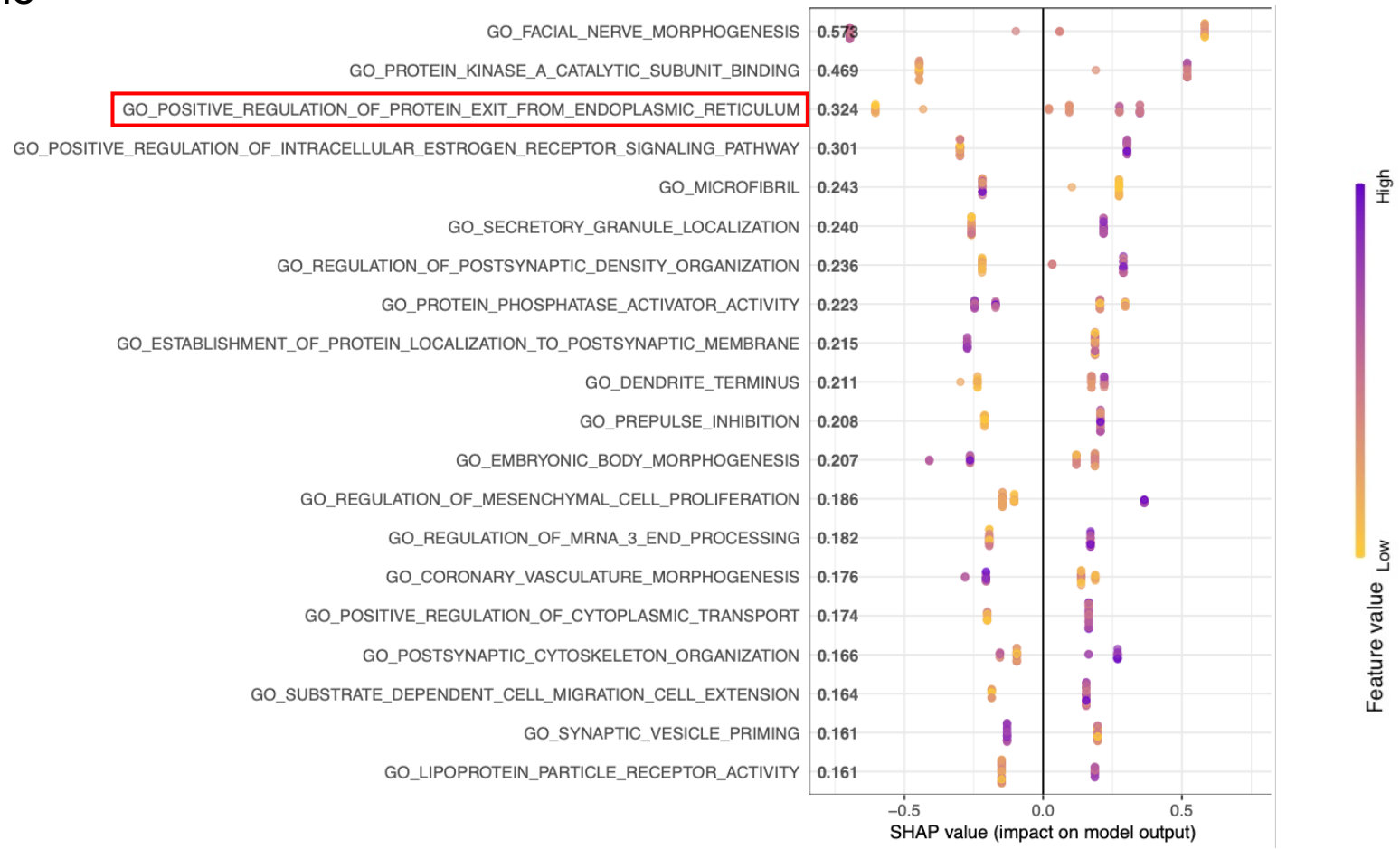
Median AUCs of 0.5 to 0.7 across different ML algorithms

UPDRS 3 total score prediction – feature impacts on outcome

DIGIPD

Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Shapley additive explanations (SHAP) value analysis of feature impact on outcome



Integrated ML analysis: eGaIT + Metabolomics + Clinical

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Combine scaled features from eGaIT data, filtered metabolomics & clinical data:
  → Prediction PD vs. control (10-fold CV, clinical data: age, gender, BMI, Sniffin Sticks)

AUC statistics:

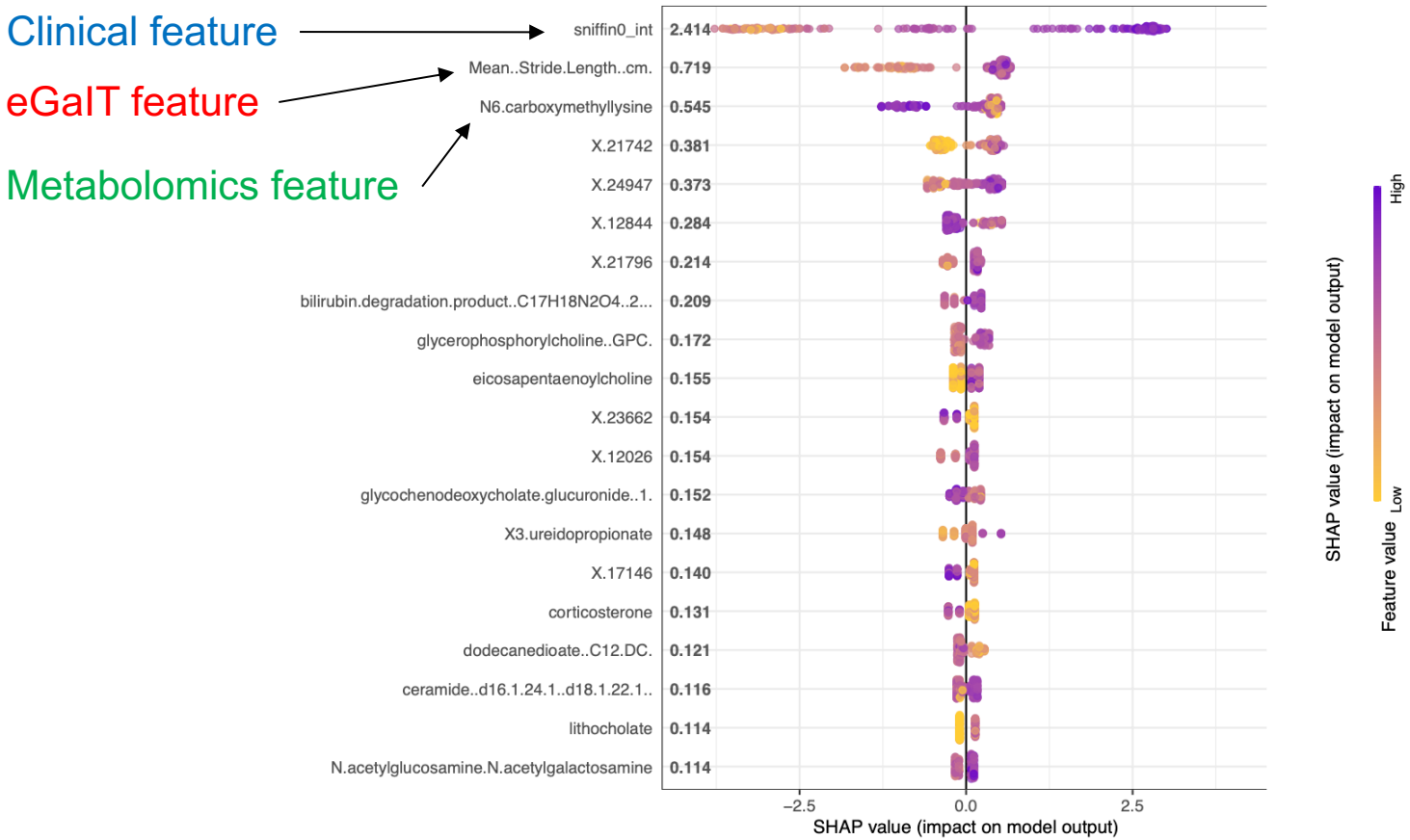
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GBM	0.852	0.895	0.932	0.930	0.966	1.000
SVM	0.716	0.772	0.846	0.837	0.898	0.938
ROTF	0.735	0.870	0.944	0.919	0.981	1.000
XGB	0.877	0.941	0.951	0.947	0.963	1.000
DEEP	0.877	0.917	0.938	0.940	0.972	1.000
RF	0.747	0.864	0.963	0.921	0.988	1.000
GAUS	0.593	0.802	0.883	0.857	0.926	1.000

Median AUCs of 0.85 to 0.96 across different ML algorithms

Integrated ML analysis: SHAP value analysis

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Combine scaled features from eGaIT data, filtered metabolomics & clinical data



Carboxymethylysine (CML) is a type of advanced glycation end product (AGE) that has been linked to age-related diseases, including PD.

Prediction of comorbidities using eGaIT data (follow-up visit)

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

Analysis: Prediction of further disease-associated outcomes in PD
(Algorithm: XGBoost, 10-fold CV)

AUC statistics:

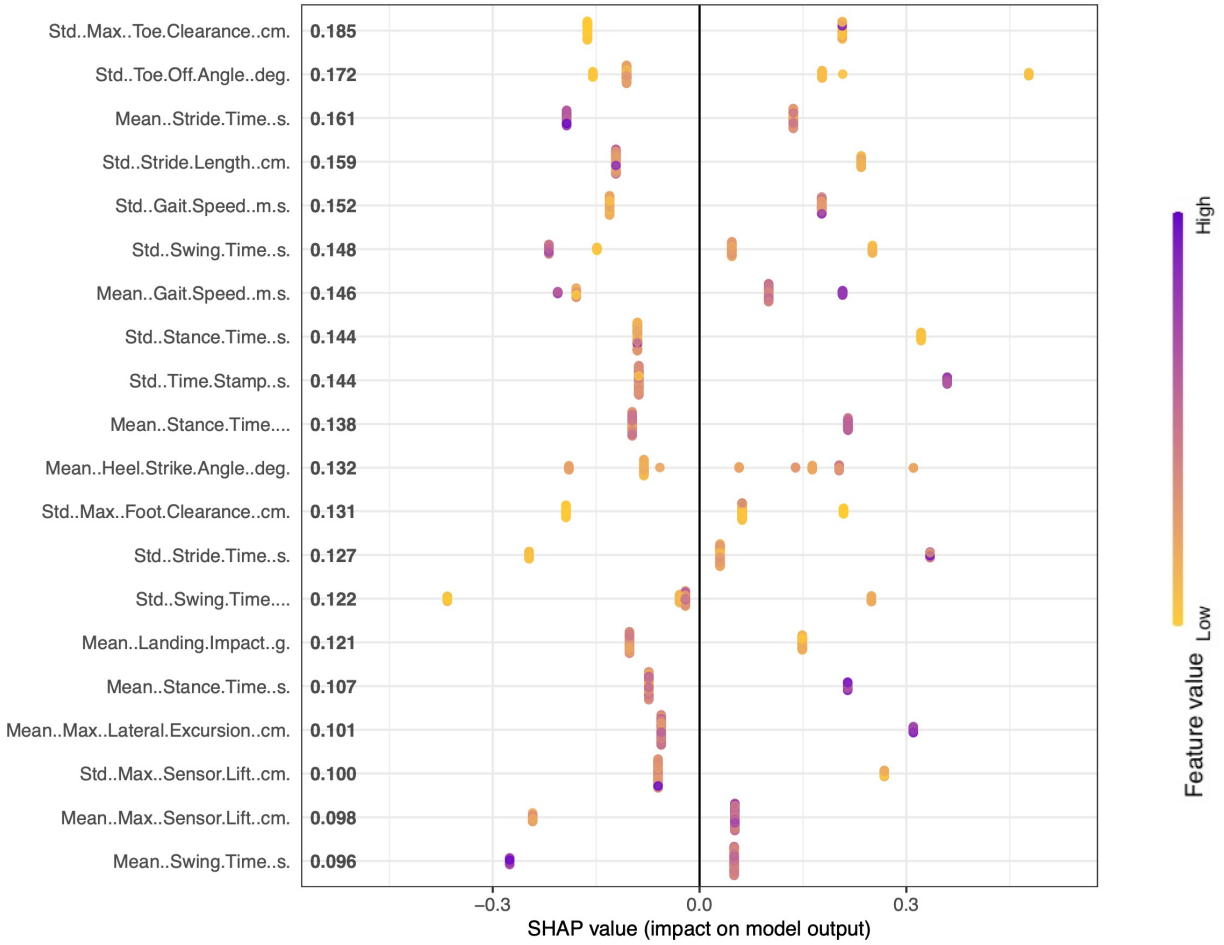
Outcome	Min.	1st Qu.	Mean	3rd Qu.	Max.
Cognitive decline (MoCA)	0.400	0.491	0.629	0.759	0.867
Impulse control disorders (QUIP)	0.192	0.519	0.642	0.817	1.000
Depression (BDI)	0.357	0.500	0.575	0.651	0.750
Hallucinations	0.333	0.444	0.536	0.573	0.833
Dyskinesias	0.282	0.427	0.557	0.646	0.962
Apathy (Starkstein scale)	0.375	0.481	0.574	0.634	0.850
Quality of life (PDQ-39)	0.375	0.542	0.593	0.673	0.729

Median AUCs between 53% to 64%, depending on the outcome

Prediction of comorbidities: MoCA example – feature impacts

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Shapley additive explanations (SHAP) value analysis of feature impact on outcome

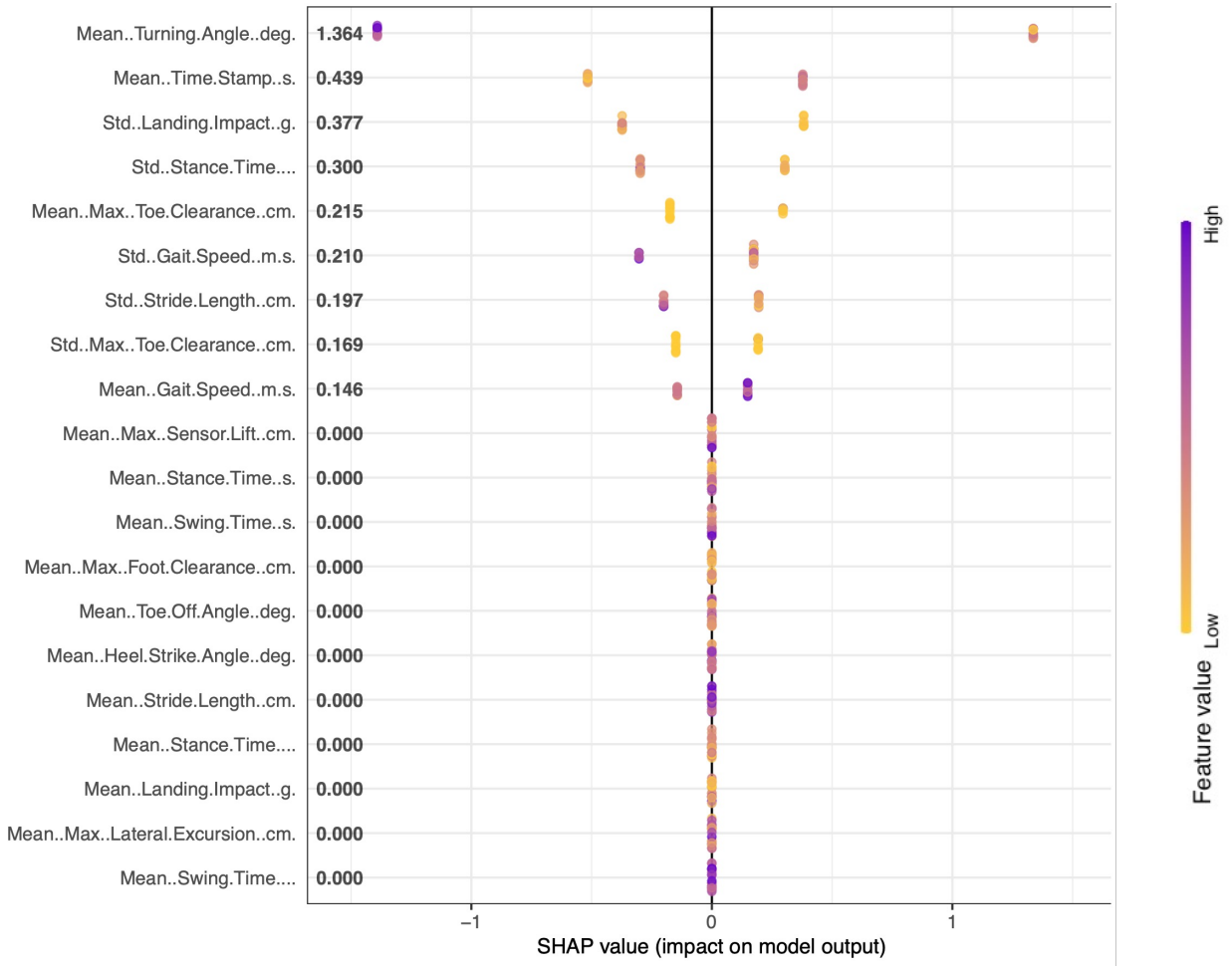


Toe & foot clearance, as well as speed-related features tend to be predictive

Prediction of comorbidities: QUIP example – feature impacts

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Shapley additive explanations (SHAP) value analysis of feature impact on outcome



Turning angle, landing impact, toe clearance, as well as speed-related features tend to be predictive

Prediction of comorbidities using clinical data (follow-up visit)

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

Analysis: Prediction of further disease-associated outcomes in PD
(Algorithm: XGBoost, 10-fold CV) using baseline numerical clinical data (no filtering)

AUC statistics:

Outcome	Min.	1st Qu.	Mean	3rd Qu.	Max.
Cognitive decline (MoCA)	0.496	0.535	0.608	0.680	0.793
Impulse control disorders (QUIP)	0.460	0.548	0.604	0.662	0.795
Depression (BDI)	0.426	0.559	0.596	0.644	0.736
Hallucinations	0.358	0.422	0.536	0.624	0.722
Dyskinesias	0.432	0.502	0.573	0.658	0.735
Apathy (Starkstein scale)	0.468	0.484	0.546	0.584	0.646
Quality of life (PDQ-39)	0.510	0.522	0.559	0.588	0.608

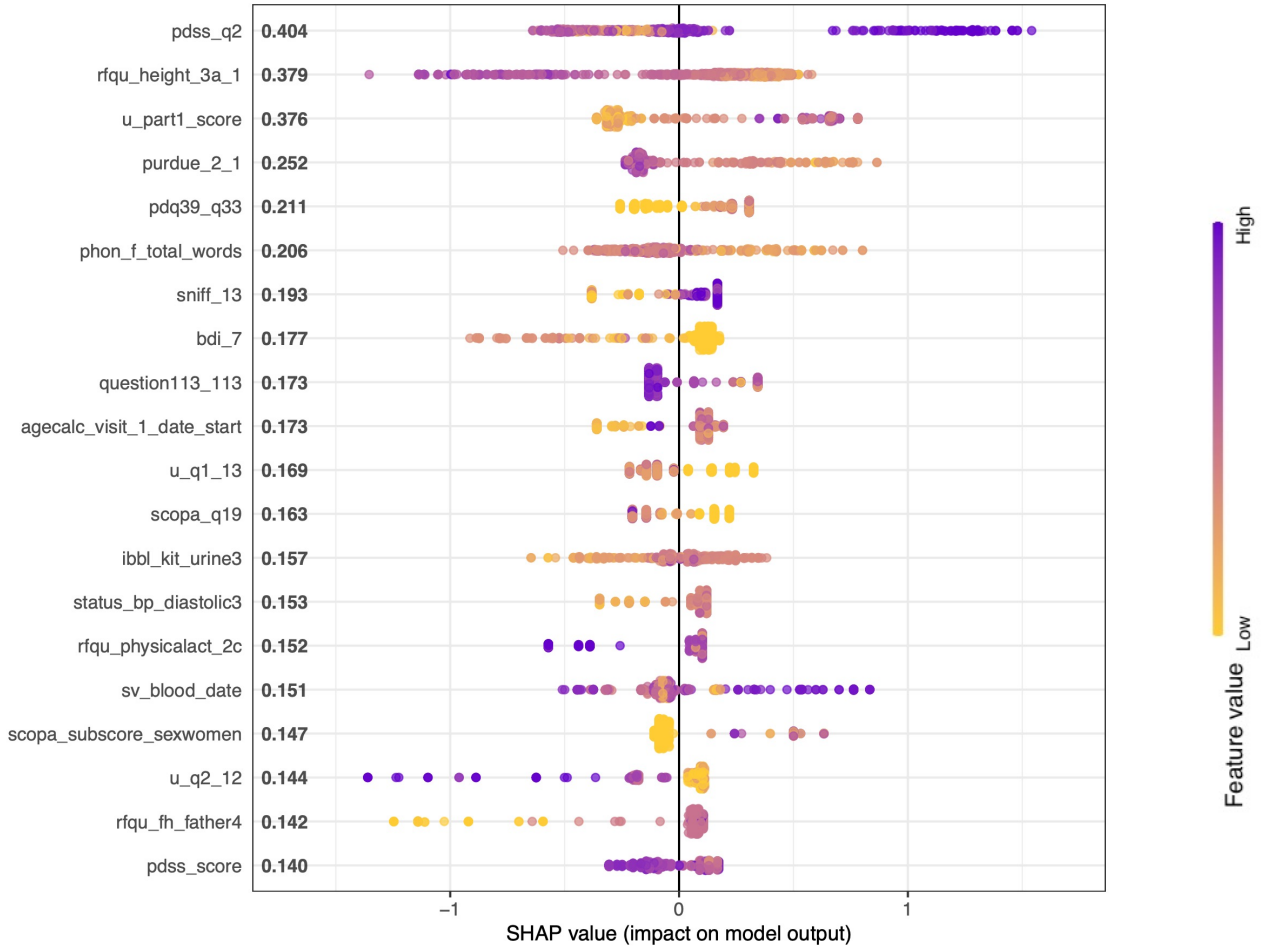
Median AUCs between 53% to 61%, depending on the outcome

Prediction of comorbidities: MoCA example – feature impacts

DIGIPD

Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Shapley additive explanations (SHAP) value analysis of feature impact on outcome

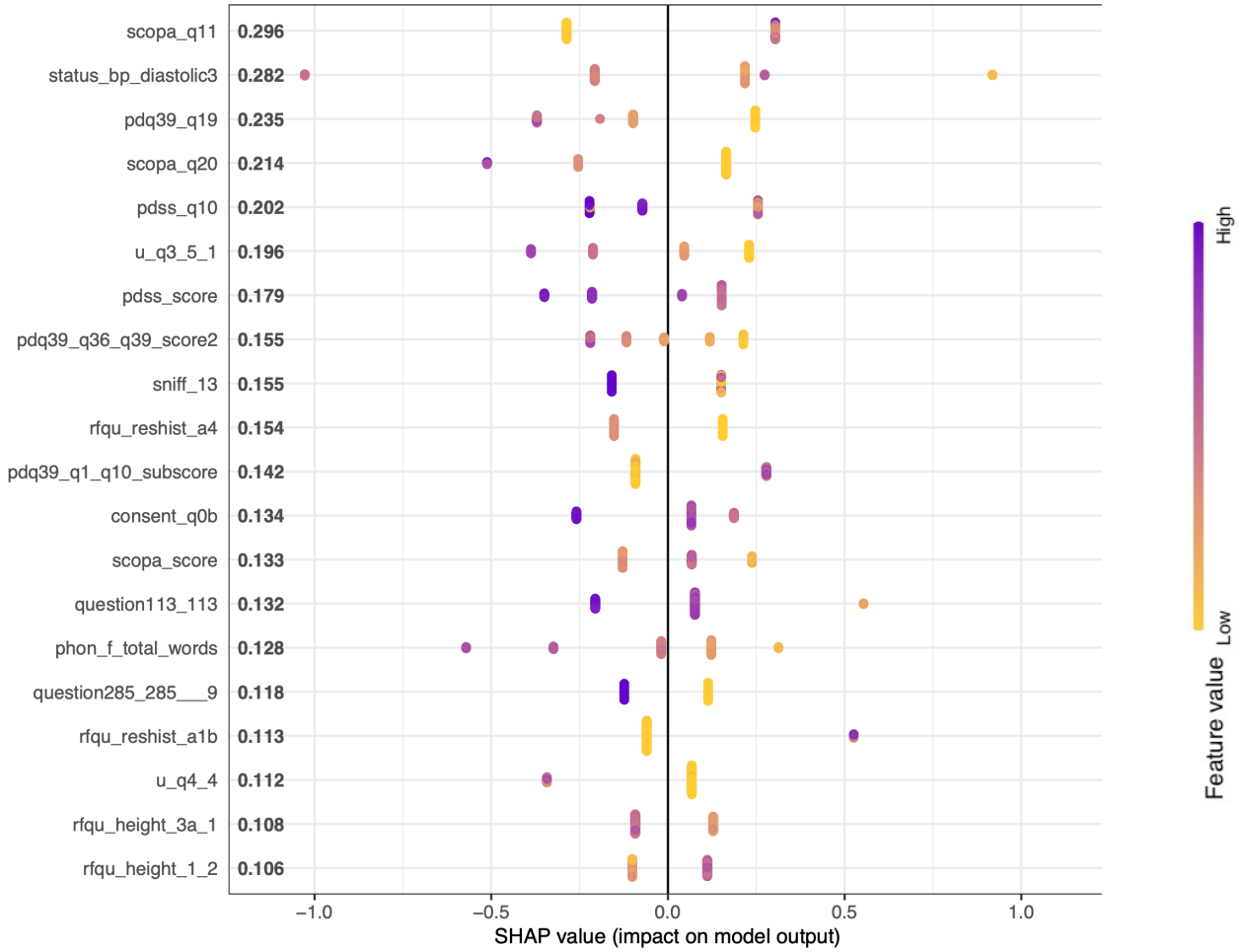


Sleep-related features, phonemic fluency, autonomic dysfunction, and overall non-motor phenotypes severity tend to be predictive

Prediction of comorbidities: QUIP example – feature impacts

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Shapley additive explanations (SHAP) value analysis of feature impact on outcome



Autonomic dysfunction, blood pressure, tearful feeling, sleep scores, sense of smell score tend to be predictive

Prediction of comorbidities using metabolomics (follow-up visit)

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

Analysis: Metabolomics-based prediction of disease-associated outcomes in PD (XGBoost, 10-fold CV)

AUC statistics:

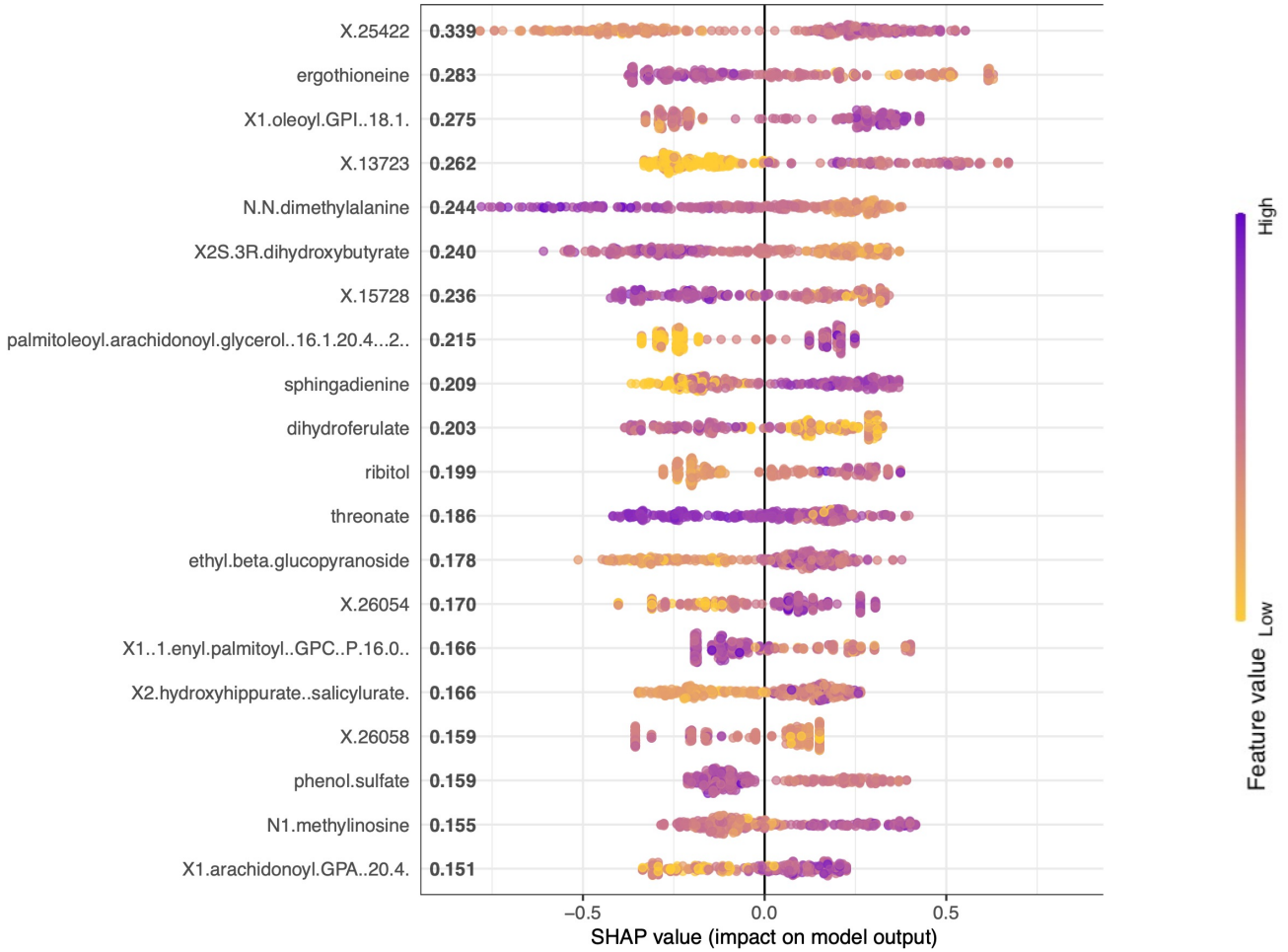
Outcome	Min.	1st Qu.	Mean	3rd Qu.	Max.
Cognitive decline (MoCA)	0.435	0.596	0.648	0.714	0.814
Impulse control disorders (QUIP)	0.161	0.457	0.563	0.721	0.899
Depression (BDI)	0.326	0.487	0.526	0.573	0.712
Hallucinations	0.399	0.481	0.578	0.642	0.767
Dyskinesias	0.326	0.548	0.669	0.820	0.907
Apathy (Starkstein)	0.495	0.552	0.593	0.621	0.718
Quality of life (PDQ-39)	0.401	0.514	0.574	0.628	0.734

Mean AUCs between 53% to 67%, depending on the outcome

Prediction of comorbidities: MoCA example – feature impacts

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Shapley additive explanations (SHAP) value analysis of feature impact on outcome



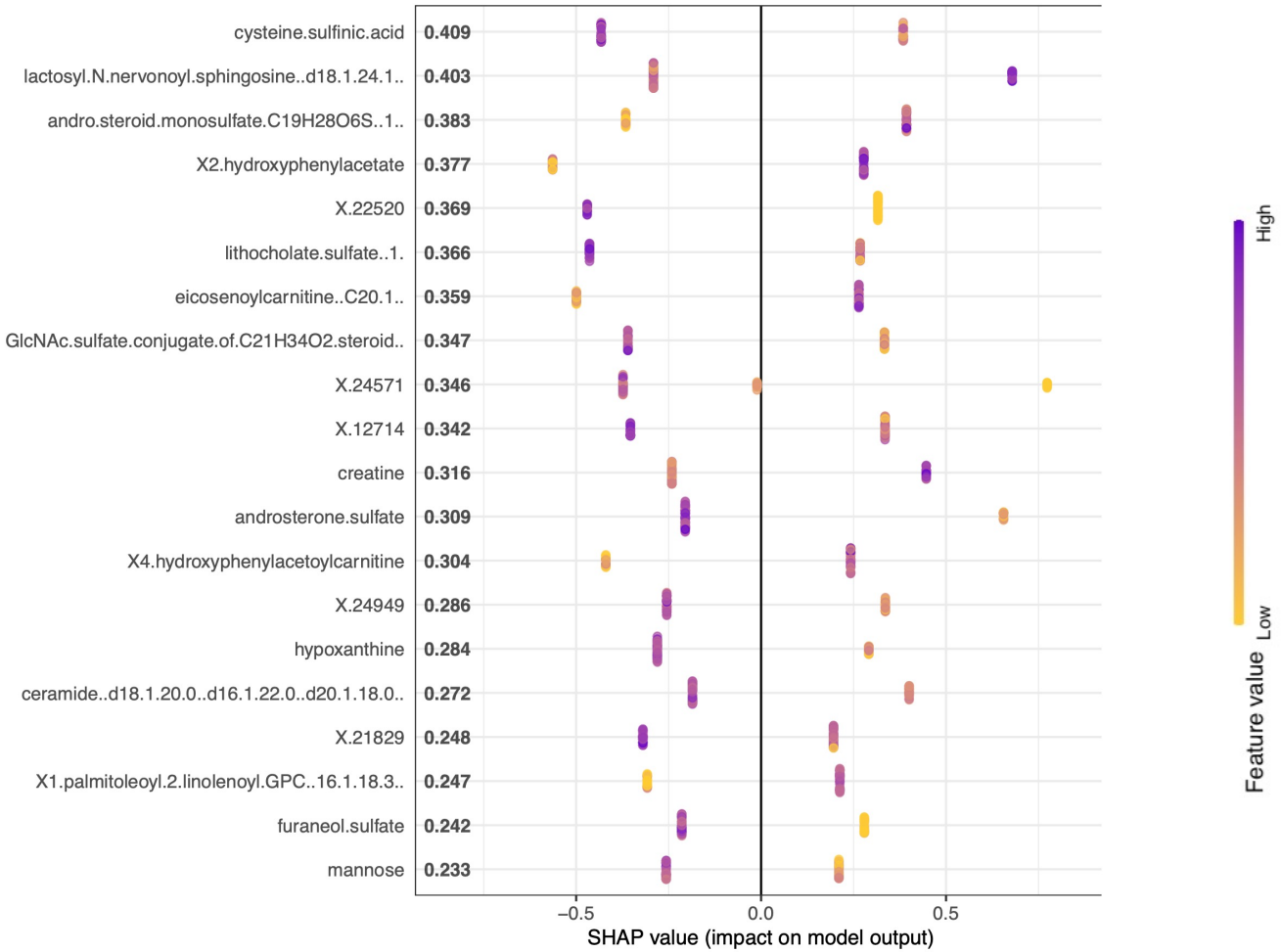
Top-ranked known metabolite ergothioneine has antioxidant properties and low levels have been associated with dementia (Wu et al., 2021), longitudinal consumption is protective in the 5xFAD model of Alzheimer's (Whitmore et al., 2022)

Prediction of comorbidities: Dyskinesia example

DIGIPD

Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Shapley additive explanations (SHAP) value analysis of feature impact on outcome



Formation of stabilized cysteine sulfinic acid is critical for the mitochondrial function of the Parkinsonism protein DJ1 (Blackington et al., 2009); oral creatine attenuates L-Dopa induced dyskinesia in 6-OHDA rats (Valastro et al., 2009)

Integrated prediction of comorbidities (follow-up visit)

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

Analysis: Combine metabolomics with eGaIT data and selected clinical variables (age, gender, BMI, disease duration, Sniffin Sticks; XGBoost, 10-fold CV)

AUC statistics:

Outcome	Min.	1st Qu.	Mean	3rd Qu.	Max.
Cognitive decline (MoCA)	0.240	0.450	0.614	0.745	0.920
Impulse control disorders (QUIP)	0.250	0.500	0.725	1.000	1.000
Depression (BDI)	0.389	0.479	0.581	0.708	0.861
Hallucinations	0.444	0.771	0.869	1.000	1.000
Dyskinesias	0.500	0.847	0.897	1.000	1.000
Apathy (Starkstein)	0.167	0.447	0.527	0.600	0.880
Quality of life (PDQ-39)	0.320	0.562	0.628	0.710	0.833

Mean AUCs between 53% to 90%, depending on the outcome

Multivariate analysis of comorbidities in PD

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Identify outcomes with consistent correlations: QoL, Apathy, Hallucination, Depression, ICD
- Apply multivariate elastic net regression (R-package 'joinet', Rauschenberger & Glaab, 2021)

Mean absolute errors:

Method	QoL	Apathy	Hallucination	Depression	ICD
univariate	0.705	0.763	0.595	0.743	0.569
multivariate	0.719	0.758	0.585	0.725	0.557

Improvements: green
No improvement: red

Linking DMs to clinical outcomes, imaging & molecular markers

DIGIPD
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

- Outlook:
  - Next steps: 1. Further associative analyses (e.g. canonical correlation analysis)
    2. Interpretable rule-based machine learning
    3. Network-based machine learning
    4. Estimate causal relations (Double ML / causal reasoning analyses)
  - Correlate PPMI Digital Sensor data from Verily Study Watch (sensor-based smart watch) against other molecular data?

Publications

1. A. Rauschenberger, Z. Landoulsi, M. A. van de Wiel, E. Glaab. *Penalized regression with multiple sources of prior effects*, Bioinformatics (2022), 39(12), doi: 10.1007/s12035-022-02985-2.
2. M. Ali, O. Uriarte Huarte, T. Heurtaux, P. Garcia, B. Pardo Rodriguez, K. Grzyb, R. Halder, A. Skupin, M. Buttini, E. Glaab. *Single-Cell Transcriptional Profiling and Gene Regulatory Network Modeling in Tg2576 Mice Reveal Gender-Dependent Molecular Features Preceding Alzheimer-Like Pathologies*, Mol Neurobiol (2022), doi:10.1007/s12035-022-02985-2.
3. A. Rauschenberger, E. Glaab. *Predicting Dichotomised Outcomes from High-Dimensional Data in Biomedicine*, Journal of Applied Statistics, (2023), doi: 10.1080/02664763.2023.2233057.
4. L. C. Tranchevent, R. Halder, E. Glaab. *Systems level analysis of sex-dependent gene expression changes in Parkinson's disease*, NPJ Parkinson's Disease, (2022), 9, 8.
5. A. Rauschenberger, E. Glaab, *Predicting correlated outcomes from molecular data*, Bioinformatics (2021), 37(21), 3889–3895
6. R. Diaz-Uriarte, E. Gómez de Lope, R. Giugno, H. Fröhlich, P. V. Nazarov, I. A. Nepomuceno-Chamorro, A. Rauschenberger, E. Glaab, *Ten Quick Tips for Biomarker Discovery and Validation Analyses Using Machine Learning*, PLoS Computational Biology (2022), doi:10.1371/journal.pcbi.1010357
7. E. Glaab, J.P. Trezzi, A. Greuel, C. Jäger, Z. Hodak, A. Drzezga, L. Timmermann, M. Tittgemeyer, N. J. Diederich, C. Eggers, Integrative analysis of blood metabolomics and PET brain neuroimaging data for Parkinson's disease, Neurobiology of Disease (2019), Vol. 124, No. 1, pp. 555
8. S. Köglsberger, M. L. Cordero-Maldonado, P. Antony, J. I. Forster, P. Garcia, M. Buttini, A. Crawford, E. Glaab, *Gender-specific expression of ubiquitin-specific peptidase 9 modulates tau expression and phosphorylation: possible implications for tauopathies*, Molecular Neurobiology (2017), 54(10), pp. 7979
9. N. Vlassis, E. Glaab, *GenePEN: analysis of network activity alterations in complex diseases via the pairwise elastic net*, Statistical Applications in Genetics and Molecular Biology (2015), 14(2), 221
10. E. Glaab, *Using prior knowledge from cellular pathways and molecular networks for diagnostic specimen classification*, Briefings in Bioinformatics (2015), 17(3), pp. 440
11. E. Glaab, R. Schneider, *Comparative pathway and network analysis of brain transcriptome changes during adult aging and in Parkinson's disease*, Neurobiology of Disease (2015), 74, 1-13
12. E. Glaab, R. Schneider, *RepExplore: Addressing technical replicate variance in proteomics and metabolomics data analysis*, Bioinformatics (2015), 31(13), pp. 2235
13. E. Glaab, A. Baudot, N. Krasnogor, A. Valencia. Extending pathways and processes using molecular interaction networks to analyse cancer genome data, BMC Bioinformatics, 11(1):597, 2010
14. E. Glaab, A. Baudot, N. Krasnogor, R. Schneider, A. Valencia. EnrichNet: network-based gene set enrichment analysis, Bioinformatics, 28(18):i451-i457, 2012

DIGIPD

Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease