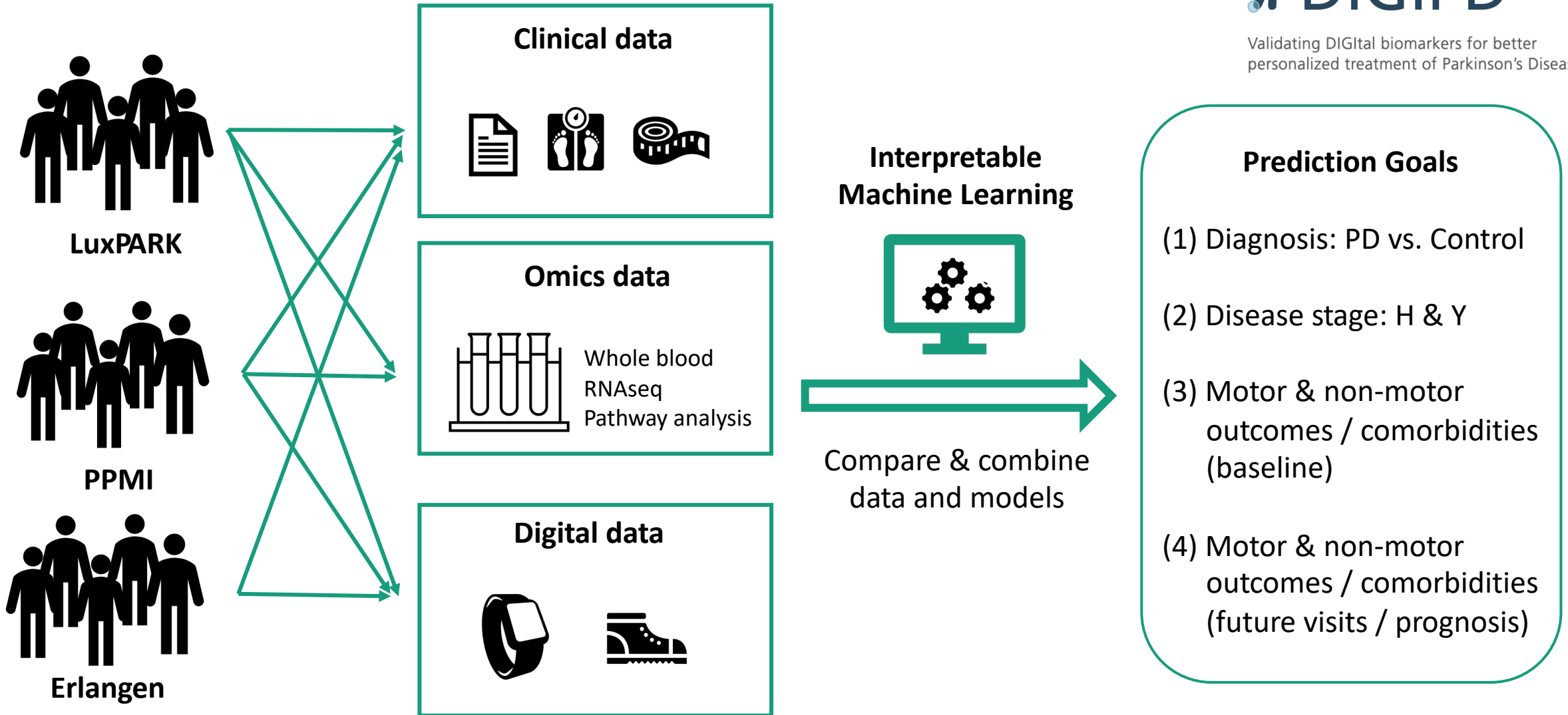


Overview: Machine Learning Analyses



LuxPARK digital biomarker data



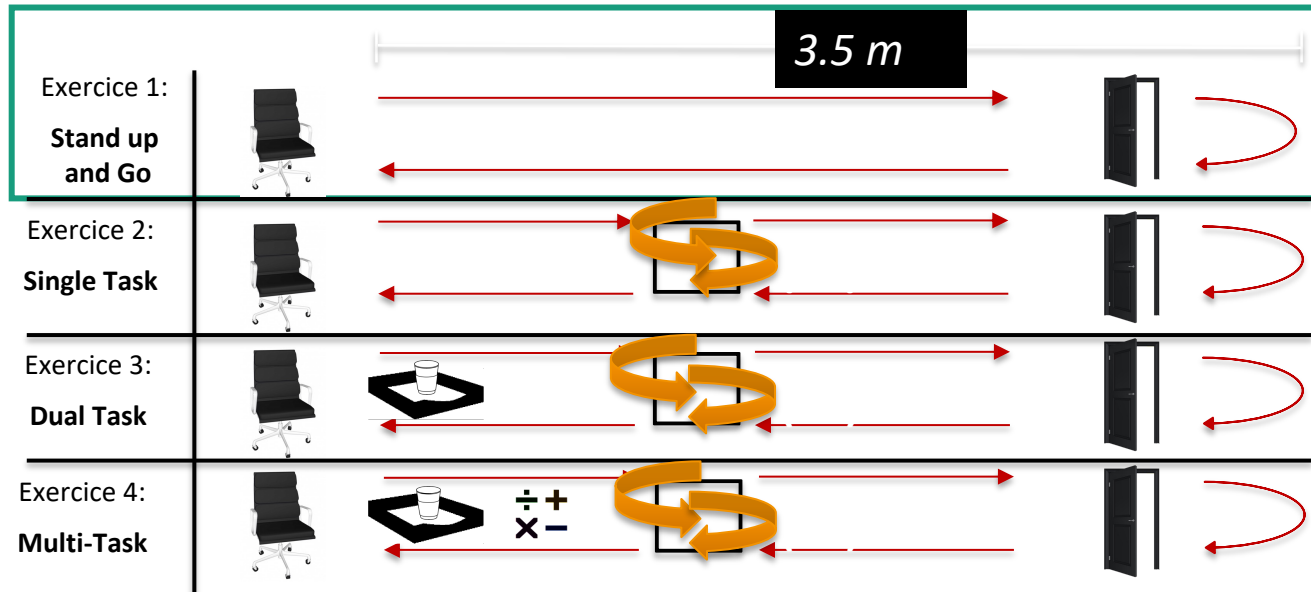
Validating DIGital biomarkers for better personalized treatment of Parkinson's Disease

- **Type of data:** eGaIT (embedded Gait analysis using Intelligent Technologies, Portables)
- **Sample sizes:** 301 subjects (205 PD patients + 95 controls)

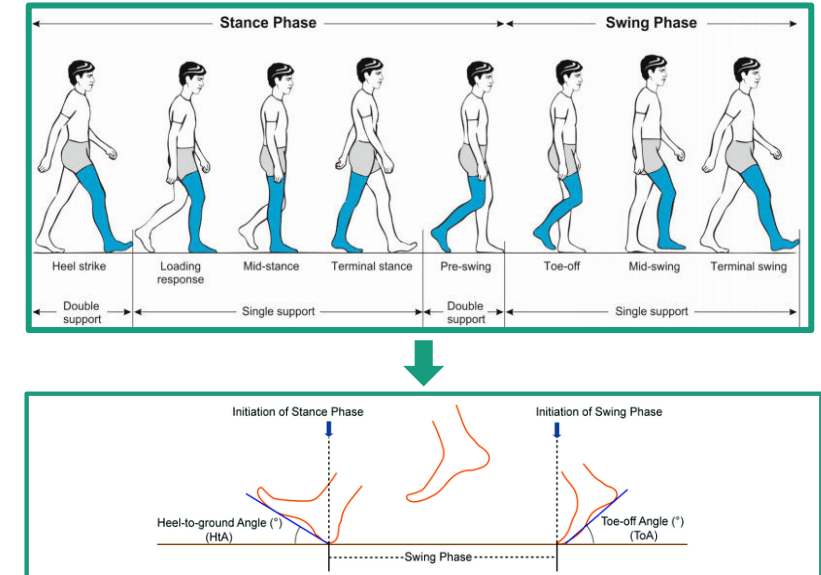


Raw sensor data + standard feature extraction (Portables)

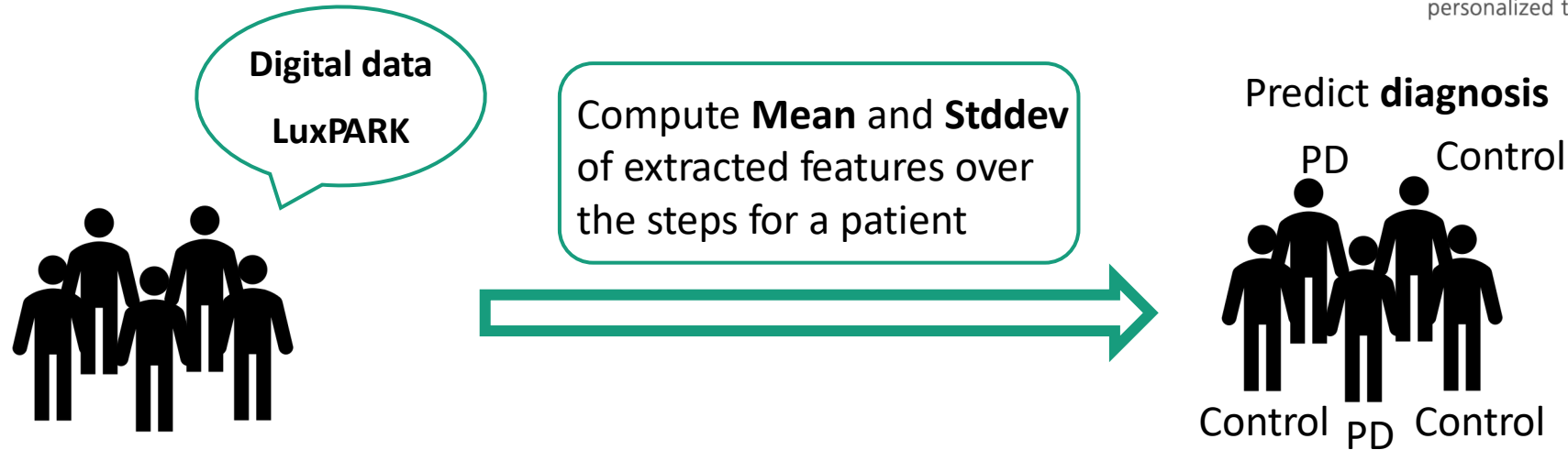
Tasks:



Extracted features:



PD vs. Control classification (Standard feature extraction)



Cross-validated predictive performance:

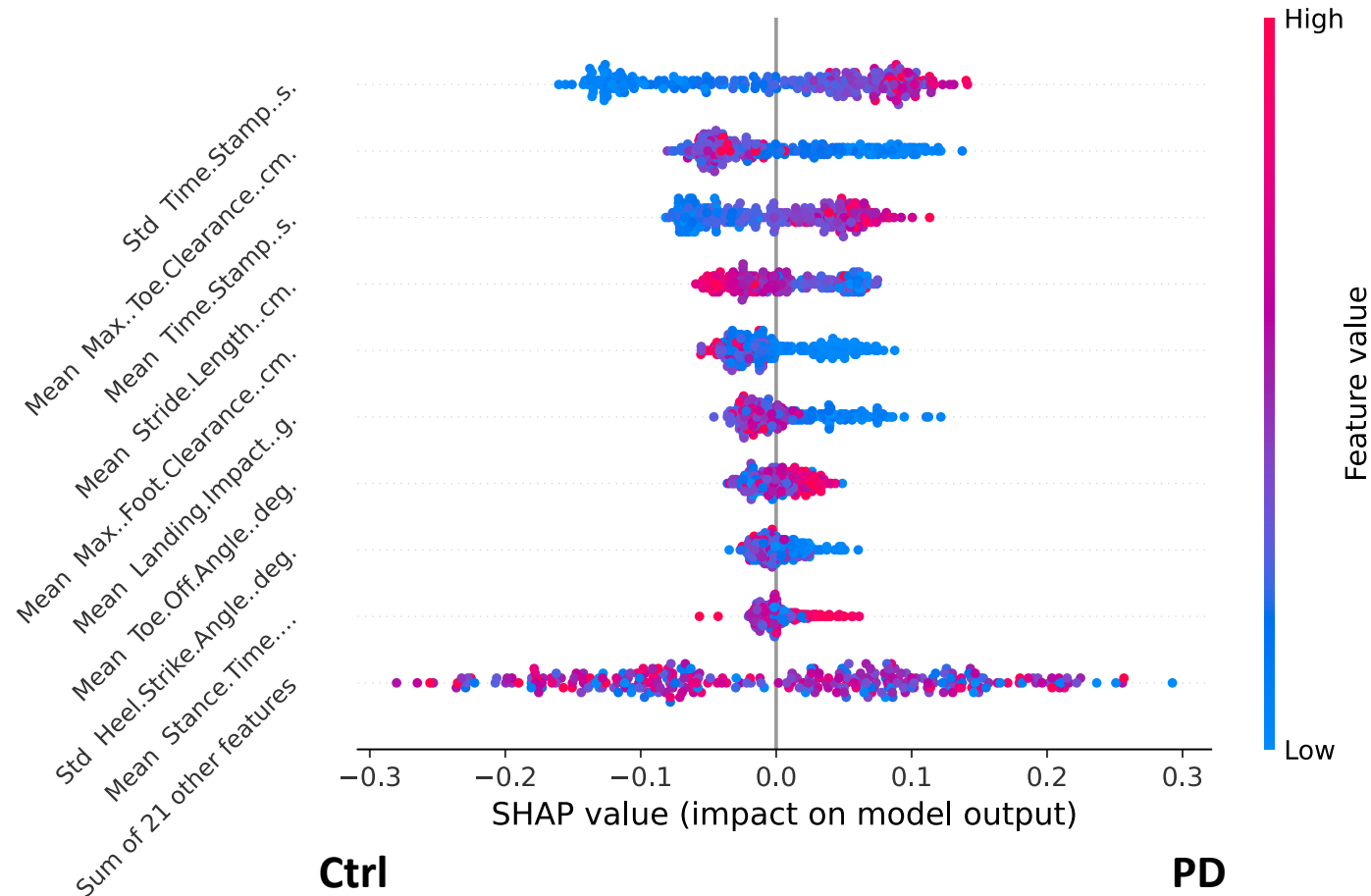
Models	Linear SVM	RBF SVM	Random Forest	Gradient Boosting	Adaboost	Logistic
10-fold CV AUC mean(std)	0.701 (0.15)	0.784 (0.06)	0.794 (0.04)	0.798 (0.06)	0.783 (0.05)	0.761 (0.08)

UK Parkinson's Disease Society Brain Bank diagnostic criteria: 76% specificity (Berg et al., 2013)

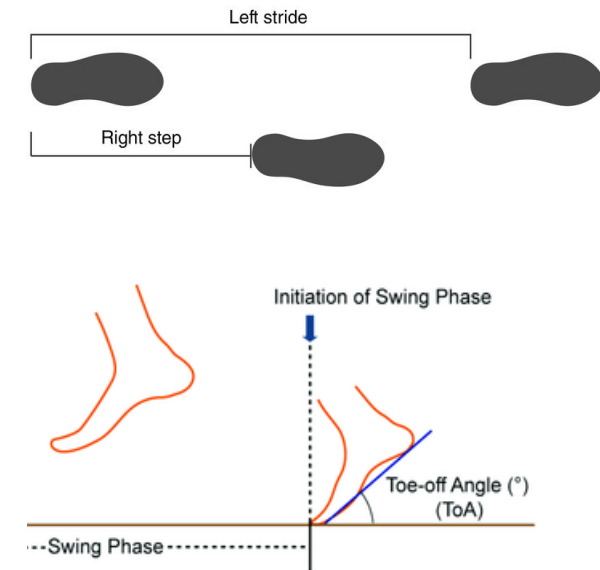
PD vs. Control – Most predictive features (SHAP analysis)



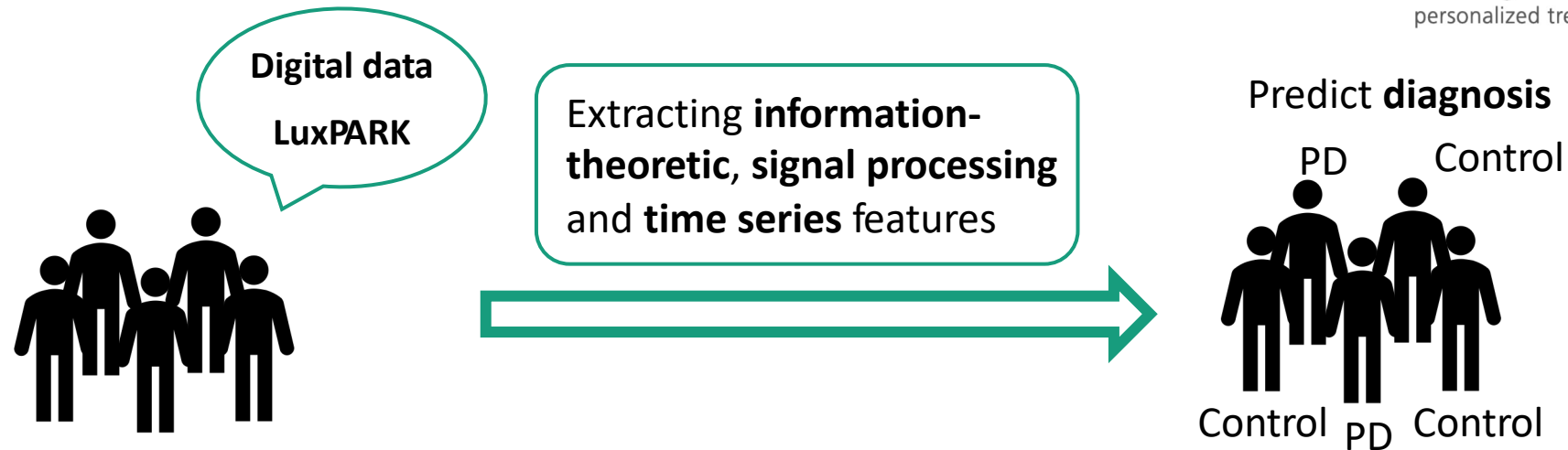
Validating DIGital biomarkers for better personalized treatment of Parkinson's Disease



Timestamp, foot and toe-clearance, and stride length related features most predictive



PD vs. Control classification (raw signal data)



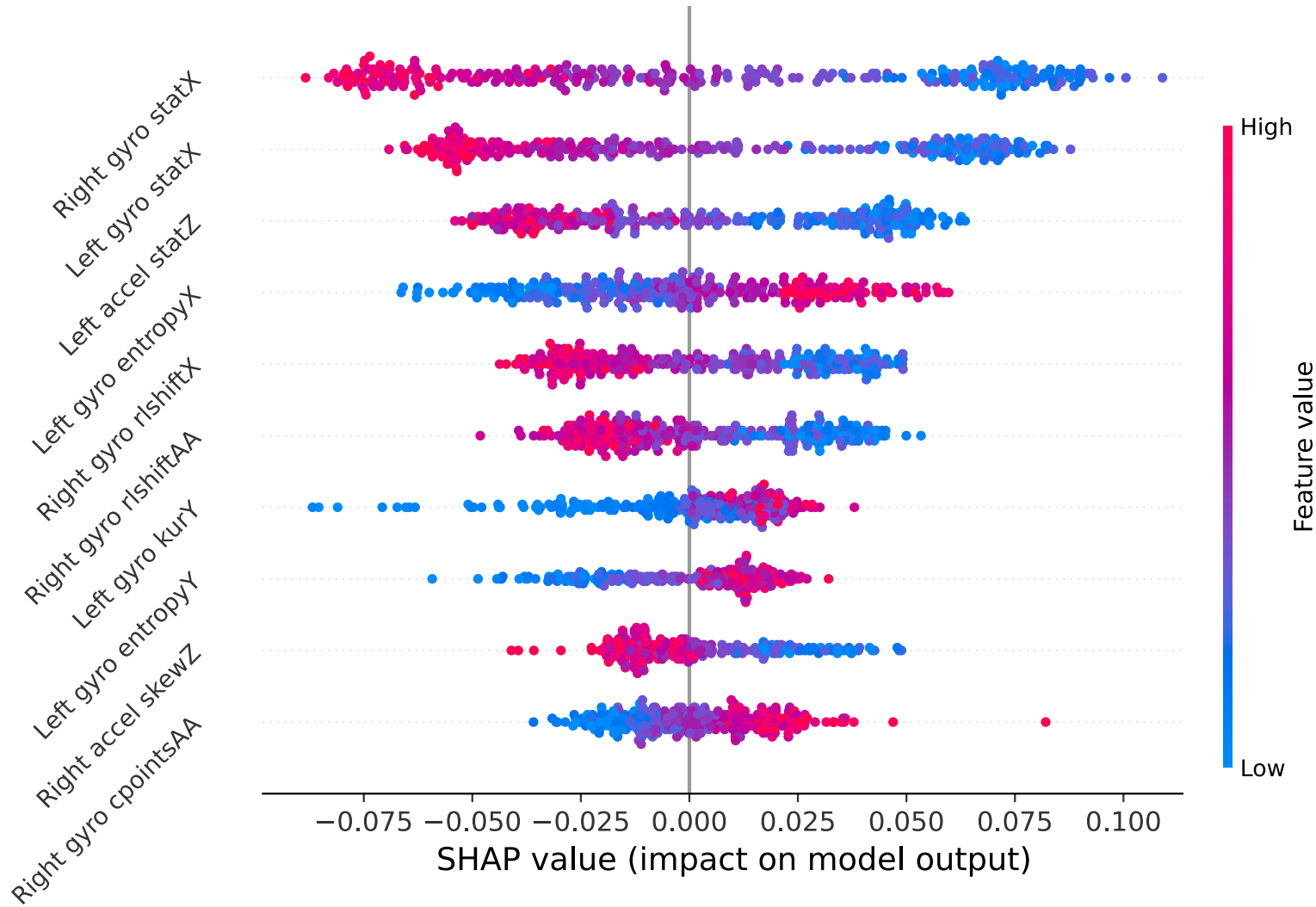
Cross-validated predictive performance:

Models	Linear SVM	RBF SVM	Random Forest	Gradient Boosting	Adaboost	Logistic
10-fold CV AUC mean(std)	0.881 (0.04)	0.870 (0.03)	0.878 (0.05)	0.852 (0.06)	0.859 (0.05)	0.882 (0.04)

PD vs. Control – Most predictive features (raw signal data)



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



Models	Linear SVM	RBF SVM	Random Forest	Gradient Boosting	Ada-boost	Logistic
10-fold CV AUC mean(std)	0.881 (0.04)	0.870 (0.03)	0.878 (0.05)	0.852 (0.06)	0.859 (0.05)	0.882 (0.04)



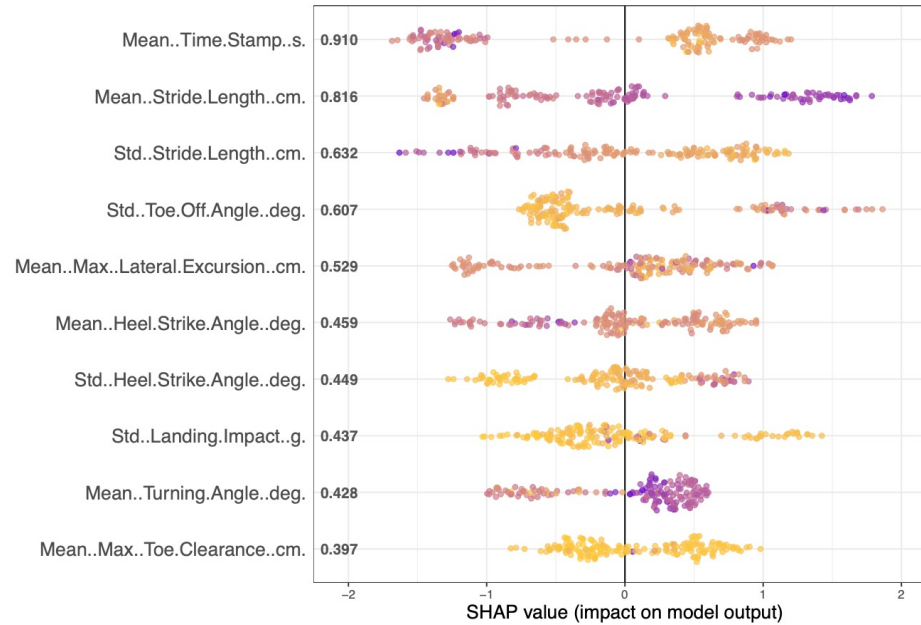
Higher predictive performance than for pre-processed gait features, but lower interpretability

Prediction of disease stage (Hoehn & Yahr scale)



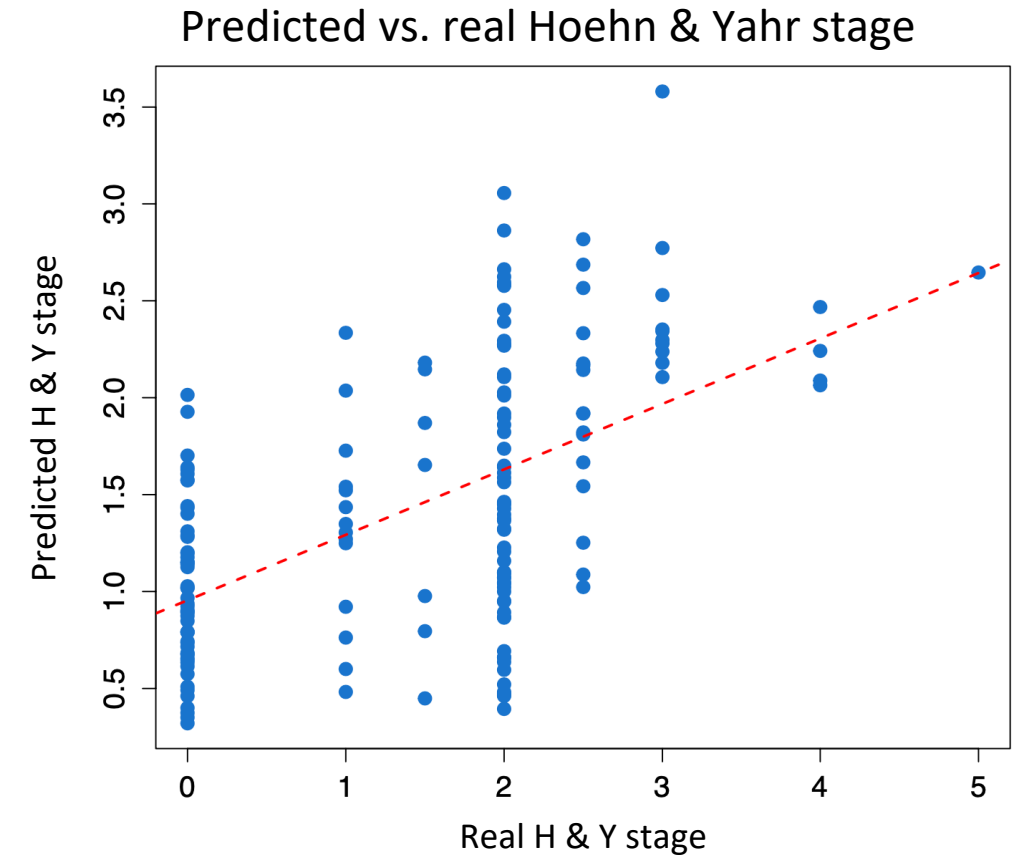
Validating DIGItal biomarkers for better personalized treatment of Parkinson's Disease

Most predictive features:



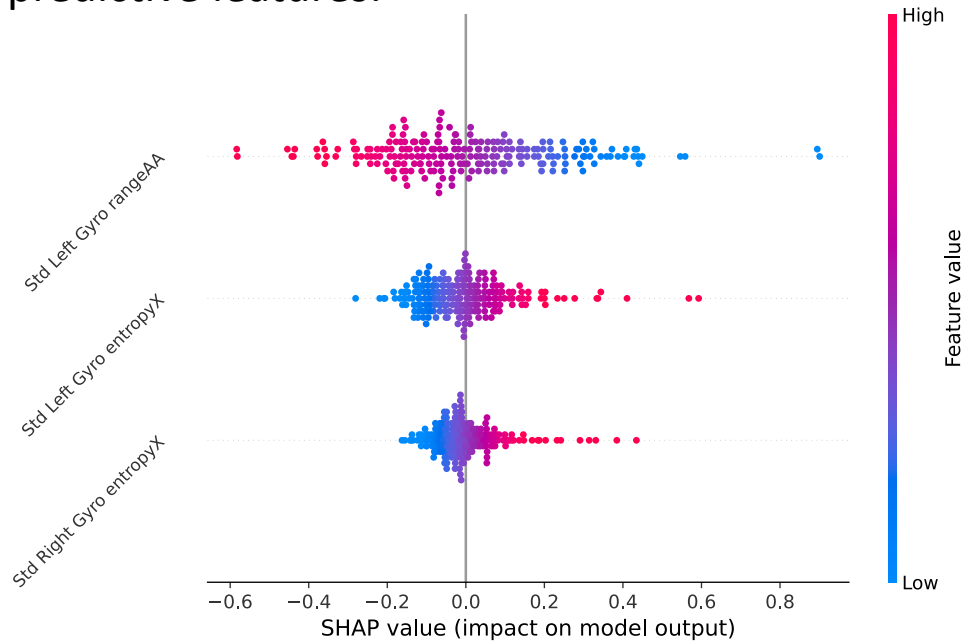
Cross-validation results (raw data):

Models	GBM	SVM	ROTF	XGB	DEEP	RF	GAUSS
10-folds AUC mean(std)	0.747 (0.1)	0.773 (0.1)	0.736 (0.11)	0.722 (0.19)	0.76 (0.05)	0.743 (0.09)	0.772 (0.13)



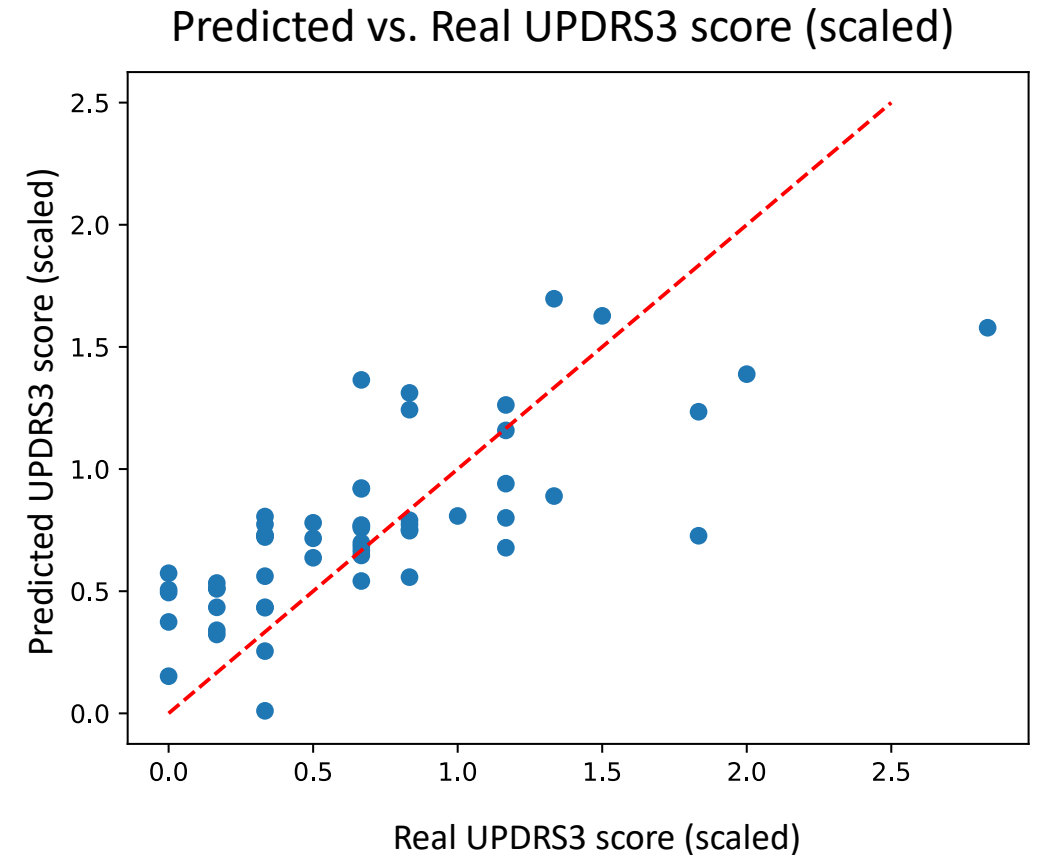
Prediction of UPDRS 3 score (digital surrogate biomarker)

Most predictive features:

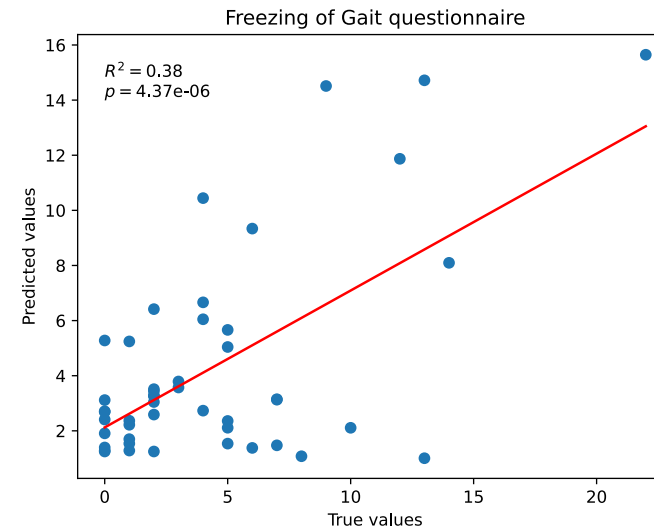
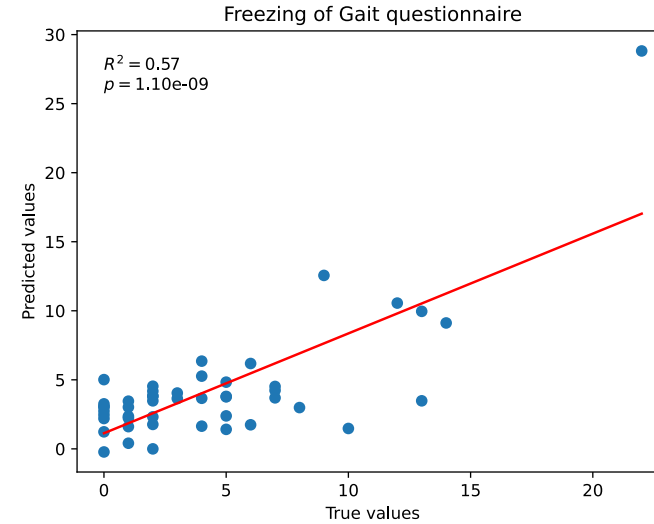
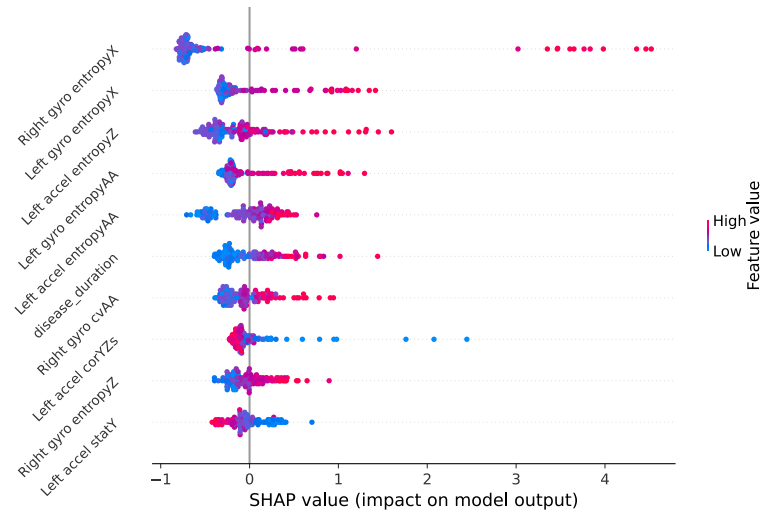
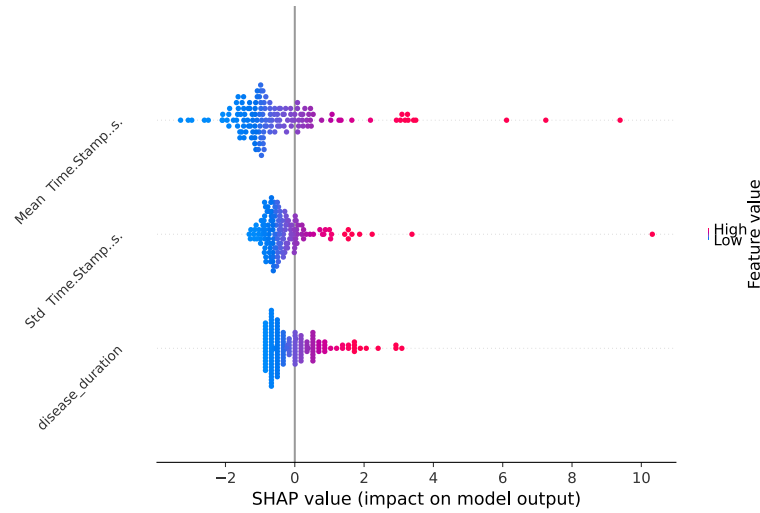


Cross-validation results (raw data):

Models	Linear SVR	RBF SVR	Random Forest	GBoost	Neural Network	Ada-boost	Ridge	OLS
CV-10: R^2 mean(std)	0.381 (0.16)	0.307 (0.20)	0.37 (0.10)	0.349 (0.14)	0.421 (0.16)	0.364 (0.12)	0.442 (0.15)	0.433 (0.15)



Freezing of Gait prediction



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

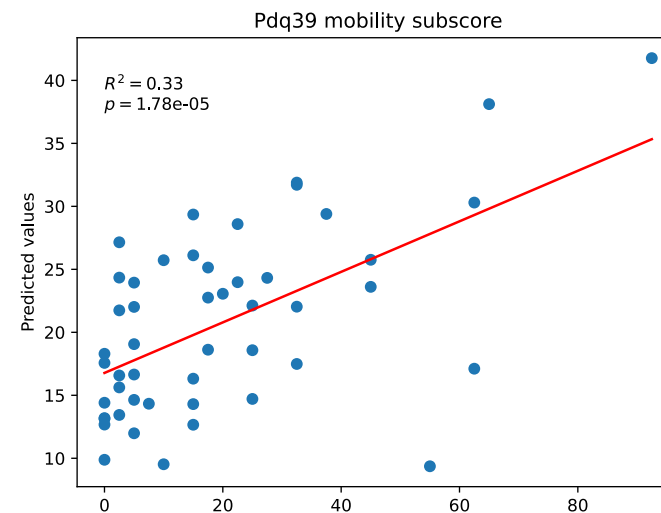
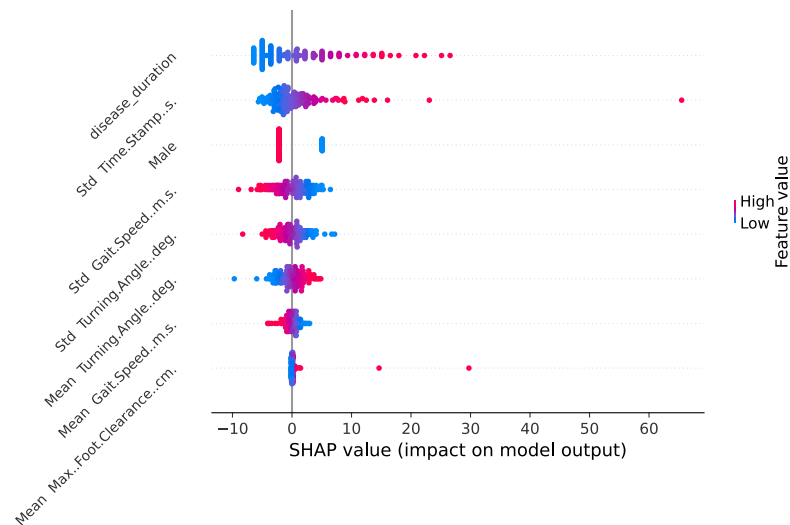
← Prediction using gait-specific features
→ benefits of gait-specific attributes

← Prediction using time series features

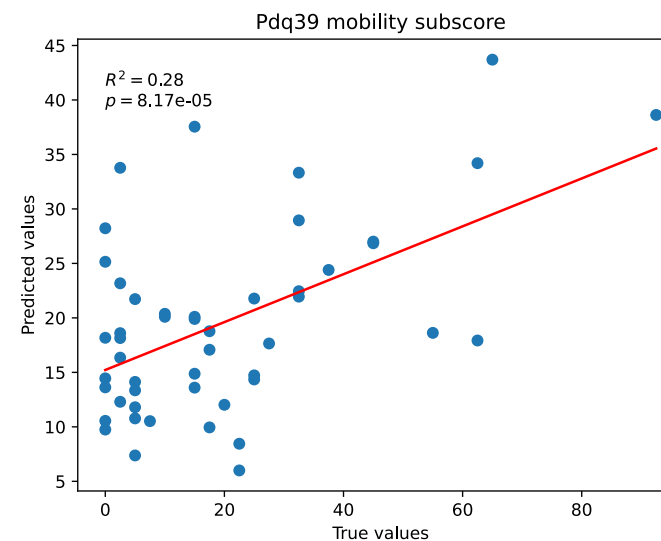
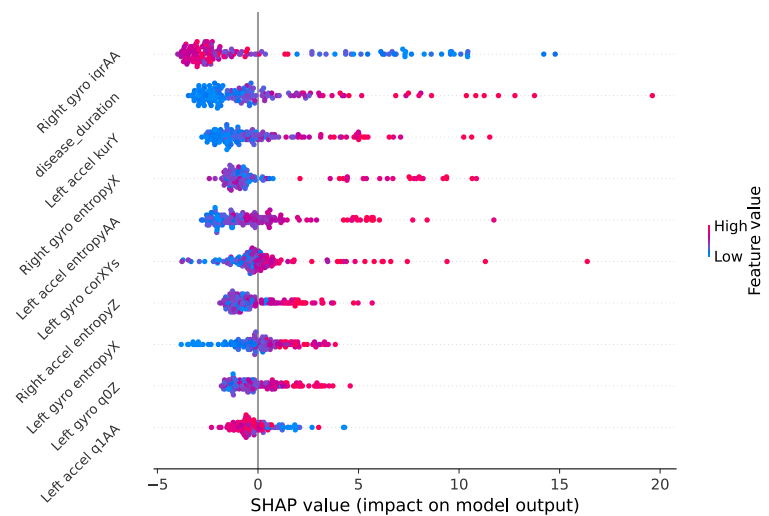
Quality of life (PDQ39) – Mobility sub-score prediction



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



← Prediction using gait-specific features
→ benefits of gait-specific attributes



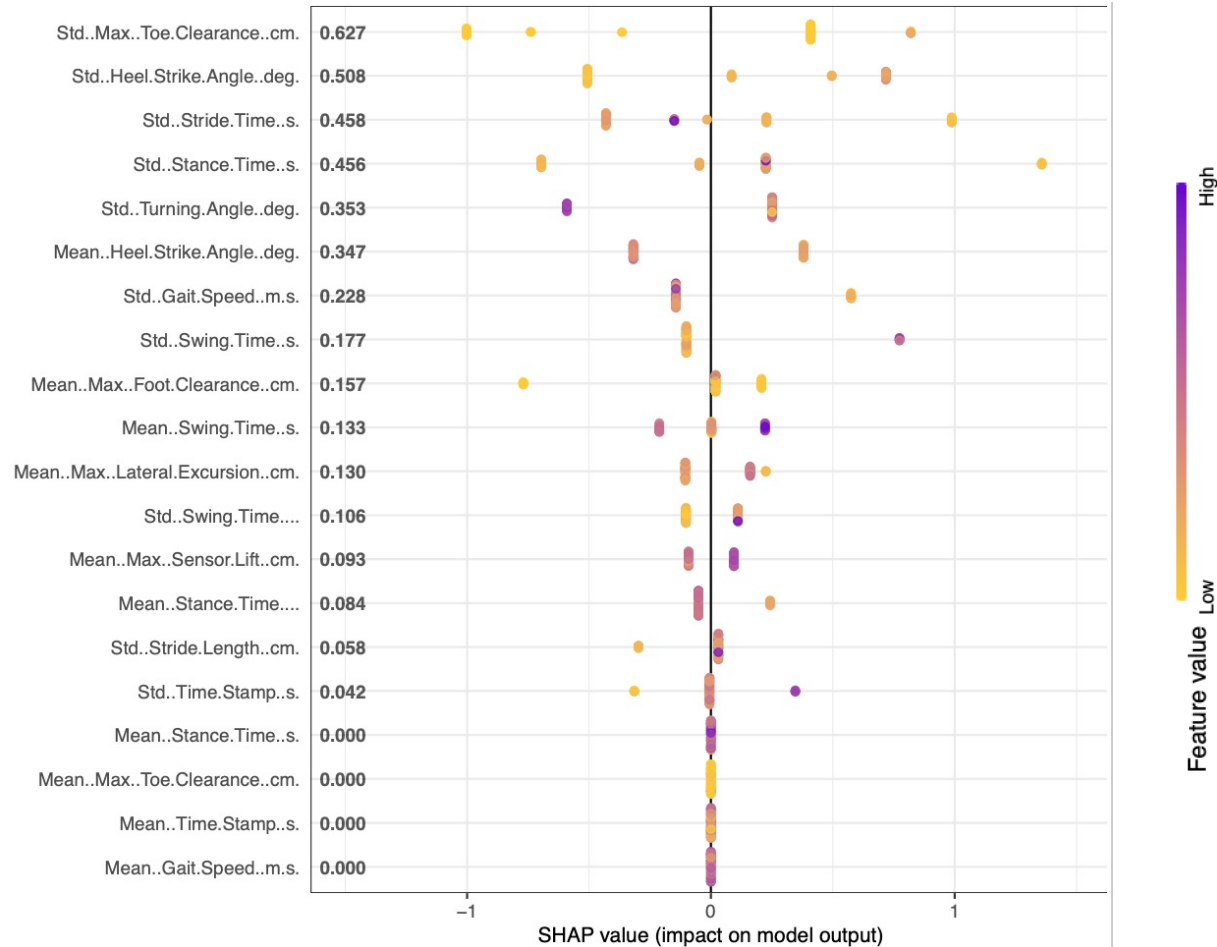
← Prediction using time series features

Prediction of akinetic-rigid vs. other phenotypes



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

Most predictive features (gait-specific):



Cross-validation results (gait-specific features):

Models	GBM	SVM	ROTF	XGB	DEEP	RF	GAUSS
10-folds AUC mean(std)	0.562 (0.24)	0.505 (0.18)	0.565 (0.1)	0.63 (0.14)	0.641 (0.21)	0.583 (0.2)	0.491 (0.16)

Cross-validation results (time series features):

Models	GBM	SVM	ROTF	XGB	DEEP	RF	GAUSS
10-folds AUC mean(std)	0.48 (0.17)	0.491 (0.21)	0.498 (0.2)	0.538 (0.22)	0.671 (0.22)	0.518 (0.2)	0.441 (0.19)

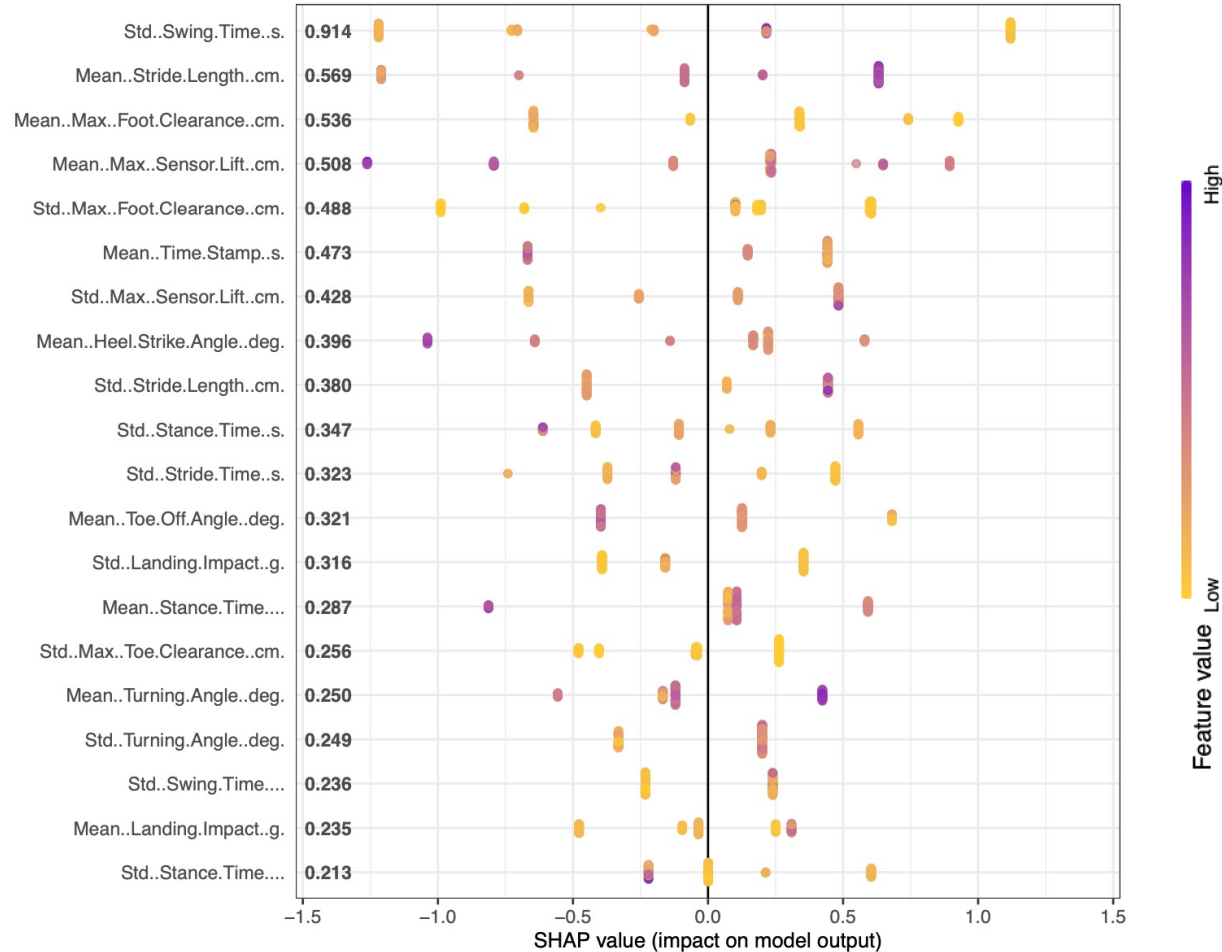
→ gait-specific attributes in most cases superior

Prediction of gait disorder occurrence in PD



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

Most predictive features (gait-specific):



Cross-validation results (gait-specific features):

Models	GBM	SVM	ROTF	XGB	DEEP	RF	GAUSS
10-folds AUC mean(std)	0.67 (0.1)	0.67 (0.19)	0.688 (0.16)	0.727 (0.1)	0.712 (0.12)	0.673 (0.11)	0.67 (0.15)

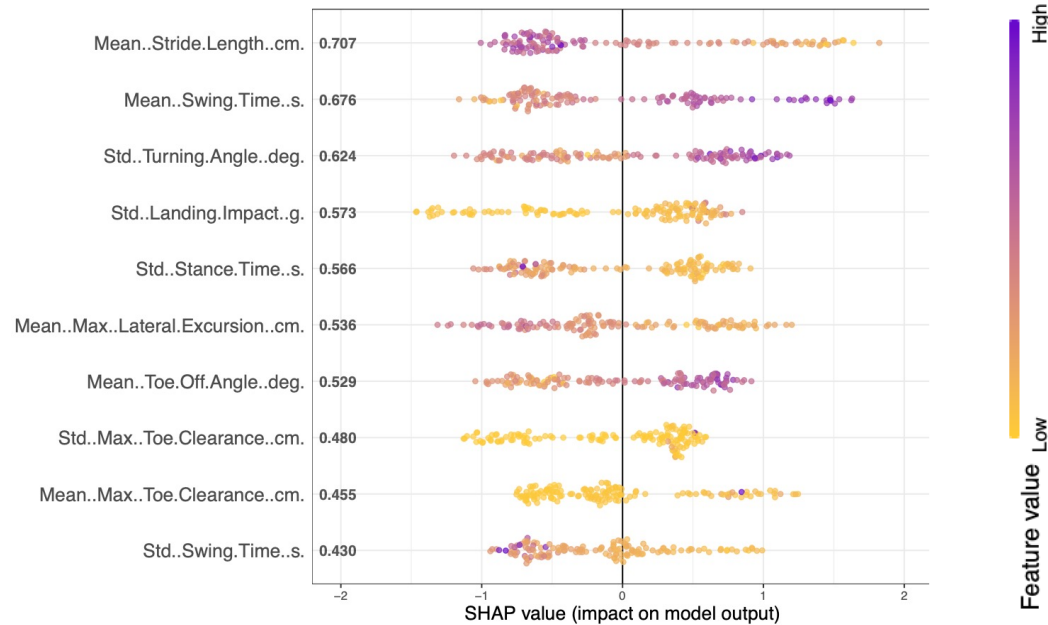
Cross-validation results (time series features):

Models	GBM	SVM	ROTF	XGB	DEEP	RF	GAUSS
10-folds AUC mean(std)	0.685 (0.09)	0.66 (0.17)	0.666 (0.19)	0.751 (0.17)	0.646 (0.22)	0.626 (0.17)	0.668 (0.19)

→ Similar performance for gait-specific attributes and time series attributes

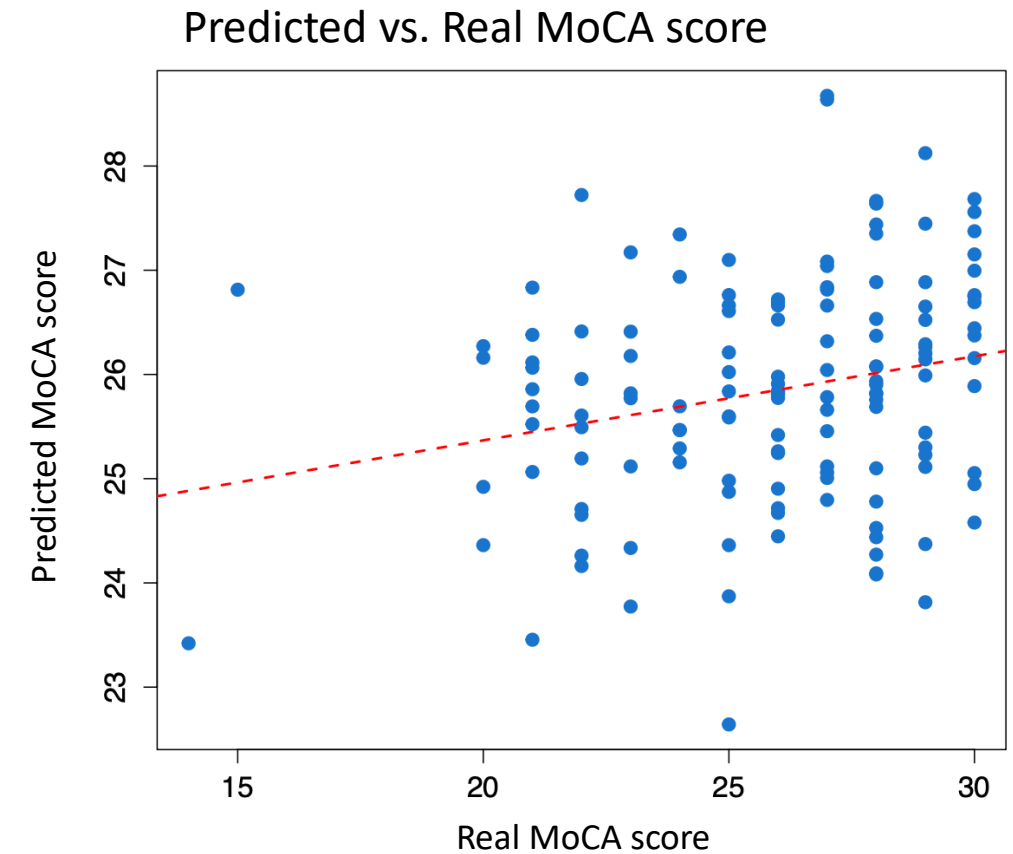
Prediction of Cognitive decline (MoCA, follow-up visit)

Most predictive features:



Cross-validation results (raw data):

Models	GBM	SVM	ROTF	XGB	DEEP	RF	GAUSS
10-folds AUC mean(std)	0.614 (0.15)	0.595 (0.14)	0.643 (0.13)	0.629 (0.17)	0.631 (0.15)	0.642 (0.08)	0.617 (0.19)



Ongoing - Future work



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

Digital data
LuxPARK



Clinical data



Omics data



Digital data



Predictive modeling

- Prediction of further motor- and non-motor outcomes
- Improving the definition of gait features (collaboration with J. Klucken group)
- Deep learning based prediction? (Cosine loss, Capsule networks)
- Rule-based ML for improved model interpretation

Integrative machine learning analysis (more tomorrow for WP4)

- Ensemble learning
- Multivariate prediction of related disease outcomes

References

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. Köglberger, 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
15. E. Glaab, A. Rauschenberger, R. Banzi, C. Gerardi, P. Garcia, J. Demotes-Mainard, and the PERMIT Group, Biomarker discovery studies for patient stratification using machine learning analysis of omics data: a scoping review, BMC Open (2021), 11, e053674
16. D. M. Hendrickx, P. Garcia, A. Ashrafi, A. Sciortino, K. J. Schmit, H. Kollmus, N. Nicot, T. Kaoma, L. Vallar, M. Buttini, E. Glaab, A new synuclein-transgenic mouse model for early Parkinson's reveals molecular features of preclinical disease, Molecular Neurobiology (2020), 58, 576-602
17. C. Brzenczek, Q. Klopfenstein, T. Hähnel, S. Sapienza, J. Klucken, H. Fröhlich, E. Glaab, *Integrating digital gait sensor data with metabolomics and clinical data to predict clinically relevant outcomes in Parkinson's disease*, npj Digital Medicine (2024), 7, 235
18. S. Le Bars, E. Glaab, *Single-Cell Cortical Transcriptomics Reveals Common and Distinct Changes in Cell-Cell Communication in Alzheimer's and Parkinson's Disease*, Molecular Neurobiology (2024), 10.1007/s12035-024-04419-7
19. E. Gómez de Lope, ... , R. Krüger, E. Glaab, *Comprehensive blood metabolomics profiling of Parkinson's disease reveals coordinated alterations in xanthine metabolism*, npj Parkinson's Disease (2024), 10, 68