AI FOR SUPPORT OF DECISION MAKING IN PERSONALIZED HEALTHCARE

Prof. Dr. Holger Fröhlich
Head of AI & Data Science Group, Deputy Head of Department of Bioinformatics
Fraunhofer Institute for Algorithms and Scientific Computing (SCAI)

AI & Data Science Group @Fraunhofer SCAI
Mission: Bringing Better Treatments to the Right Patients

- Omics
- Longitudinal clinical
- Real World Evidence
- Digital biomarkers
- Unstructured text

>130 publications

Methodologies, e.g.

Hybrid AI: combining human knowledge with AI

(Generative) time series models

Multi-modal, multi-scale data fusion

AI & Data Science

Application focus

Precision medicine

Early Drug Discovery

- Pharma
- Biotech
- Insurances
- Hospitals

Our value chain:

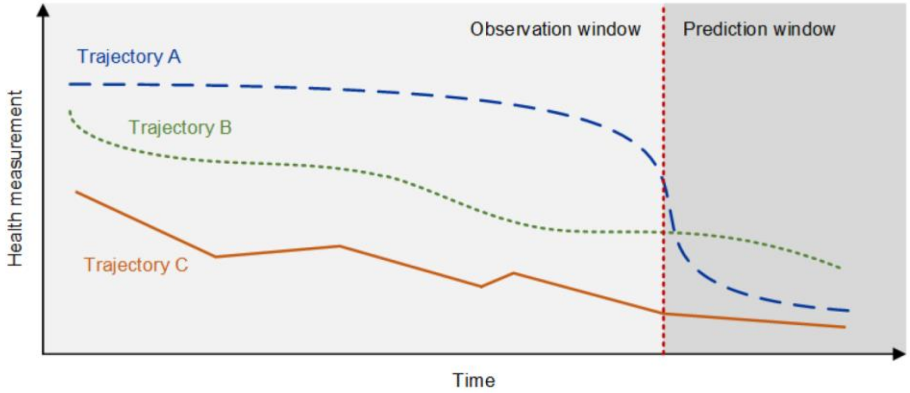
Biological system reconstruction & modeling → Drug target identification → The right drug for the right patients → Clinical trial design: simulation of synthetic cohorts

https://www.scai.fraunhofer.de/de/geschaeftsfelder/bioinformatik/forschungsfelder/AI-DAS.html

Fraunhofer

Motivation: AI for Analyzing Patient Trajectories

- Highly heterogenous disease trajectories in many disease areas
  - Example: neurogenerative diseases
- Patient trajectories provide rich information about
  - Disease dynamics
  - Disease state
  - Future disease outcomes
- Possible use of patient trajectories
  - Clustering
  - Patient pathways
  - Risk models



Allam et al., Journal of Medical Internet Research, 2021

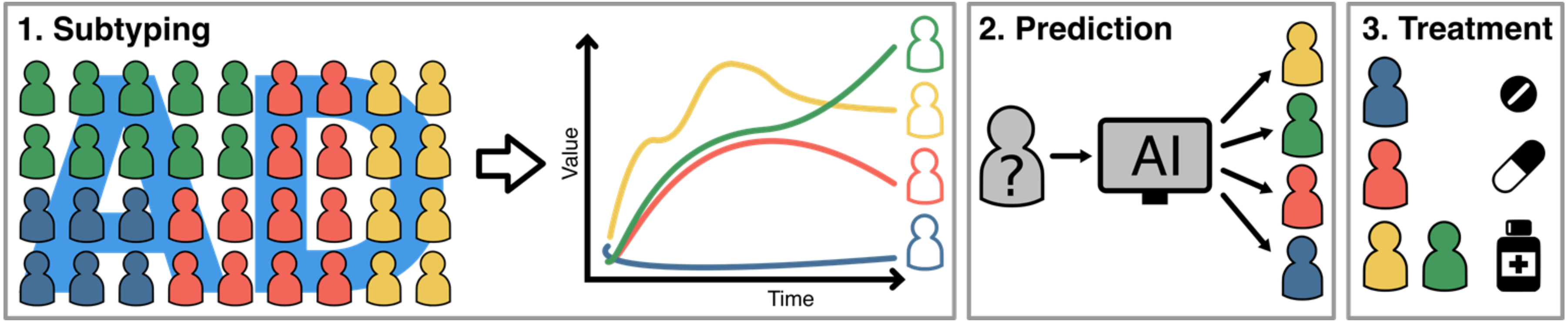
De Jong, …, Fröhlich, Giga Science, 2019

Birkenbihl, Salimi, Fröhlich, Alzheimer's & Dementia, 2022

e.g.: Linden, …, Fröhlich, Frontiers in Artificial Intelligence, 2021

Motivation: Clustering of Patient Trajectories

- Identify patients with similar patient journeys
- Personalized prediction of type of progression
- Treatment tailored to type of progression



Challenges

Clinical time series data is short: typically < 10 time points

- Existing time series clustering methods designed for more time points

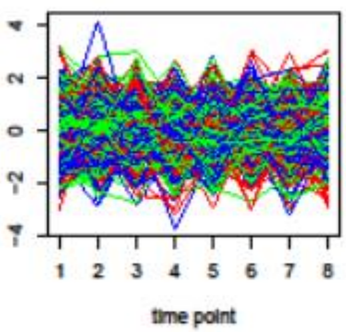
Data are limited: few hundred patients

Missing values, e.g. due to patient drop-out ➔ missing data <u>not</u> at random (MNAR)

- Correlation between missingness and observed data possible
- Multiple imputation methods exist, but errors will propagate further into clustering

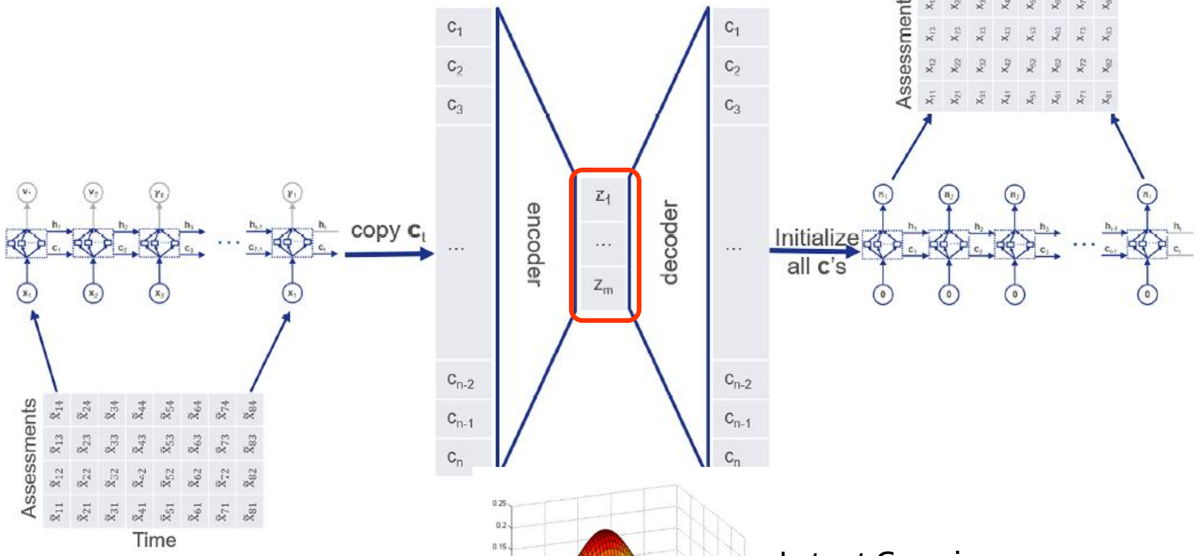
Variational Deep Embedding with Recurrence (VADER)

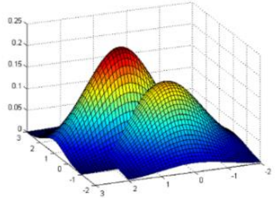
$$\mathcal{L}_{\mathrm{ELBO}}(\mathbf{x}) = E_{q(\mathbf{z},c|\mathbf{x})}[\log p(\mathbf{x}|\mathbf{z})] - D_{KL}(q(\mathbf{z},c|\mathbf{x})||p(\mathbf{z},c))$$



Multivariate short time series data

Embedding (peephole LSTM)





Latent Gaussian Mixture Model

Handling of missing data as <u>part of model training</u>

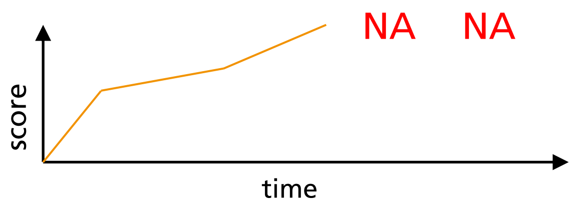
- Data augmentation + modification of VAE loss function

De Jong, …, Fröhlich, Giga Science, 2019

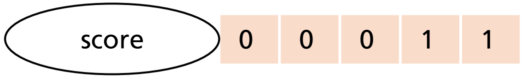
Implicit Imputation via „Imputation“ Layer

Model should only learn to reconstruct *observed* data

- Indicator variables encode missingness pattern
- Consider only observed data in reconstruction loss term in ELBO criterion



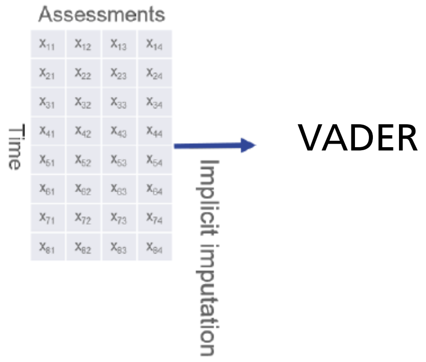
$$\frac{NM}{|A|}\sum_{l=1}^{L}\sum_{i=1}^{N}\sum_{j=1}^{M}(1-\mathbf{1}_A(x_{ij}^l))(x_{ij}^l-\widehat{x_{ij}^l})^2$$



„Implicit“ imputation („Imputation“ layer)

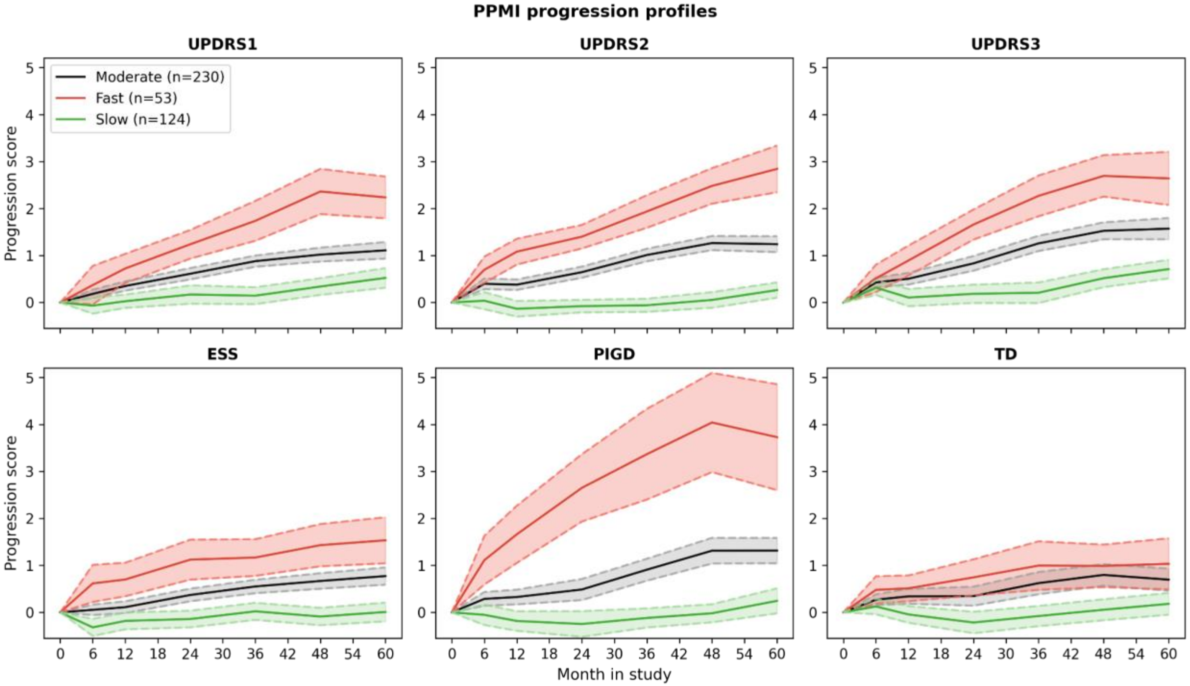
$$\tilde{x}_{ij}^l = x_{ij}^l \times (1-\mathbf{1}_A(x_{ij}^l)) + b_{ij} \times \mathbf{1}_A(x_{ij}^l)$$

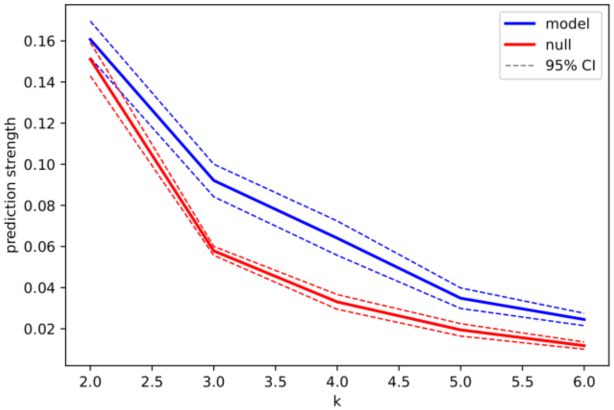
trainable



Fraunhofer

Application to Parkinson's Disease





Tibshirani & Walters, Journal of Computational and Graphical Statistics, 2005

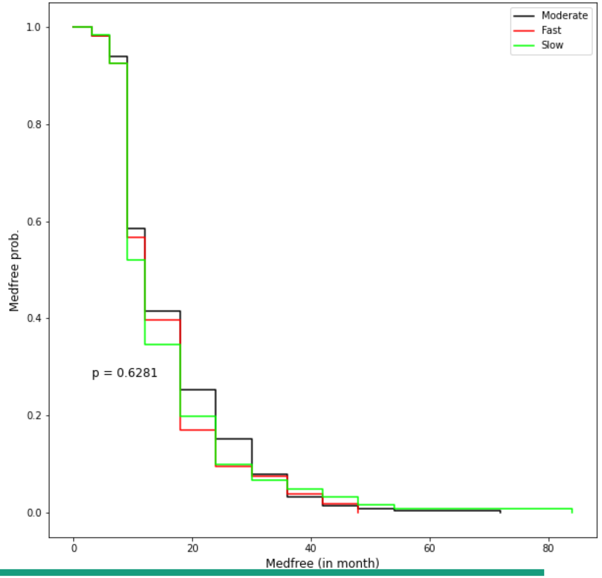
~1/3 of patients do not demonstrate significant disease progression

Fraunhofer

Are Clusters a Result of Confounding Effects?

- No significant differences in
  - Time since initial diagnosis
  - Time till start of therapy
  - Gender
- Slow progressors are younger
  - Later disease onset ➔ faster progression

Cluster	N	Age (Years)	Number of Females	Months since diagnosis	MDS-UPDRS I	MDS-UPDRS II	MDS-UPDRS III
Slow	124	60.2 ± 9.3	49 (39.5 %)	0.5 ± 0.5	5.4 ± 4.2	5.9 ± 4.2	21.0 ± 9.1
Moderate	230	62.7 ± 9.7	78 (33.9 %)	0.6 ± 0.5	5.1 ± 3.5	5.4 ± 3.9	20.6 ± 8.7
Fast	53	64.2 ± 10.8	13 (25.5 %)	0.7 ± 0.8	7.3 ± 4.7	7.2 ± 4.9	21.0 ± 8.8



Fraunhofer

Association with Distinct Symptoms, Genetic Loci and Biological Mechanisms

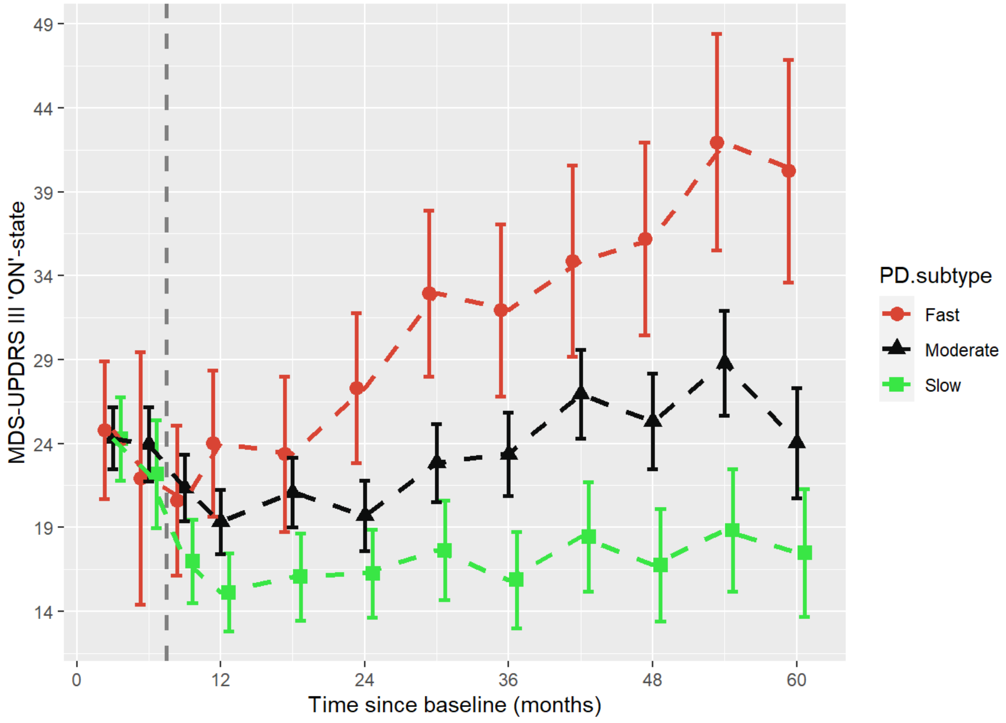
Slow	Moderate	Fast
• Hand movement, rigidity of ipsilateral extremities • Daytime sleepiness • Anxiety, depression, fatigue • Reduced semantic fluency • Polygenic risk score + several PD associated SNPs • vitamin and disaccharide metabolism (includes GBA) • Interleukin receptors	• eating • reduced agility in the ipsilateral leg • Several SNPs • Cholesterol metabolism • Vascular endothelial growth factor	• Hallucinations • Postural instability • Gait impairment • REM sleep behavior disorder (RBD) • Several PD associated SNPs • SNARE vesicle transport • Rap1 signaling / nigrostriatal dopaminergic pathway in medium spiny neurons • Low CSF amyloid beta

Method: bootstrapped sparse group lasso

Fraunhofer

Differential Response to Motor Symptom Therapy

- No differences in
  - treatment regime
  - LDOPA equivalent daily dose



Fraunhofer

Summary and Impact of this Project

Many diseases demonstrate highly heterogenous trajectories

Disease trajectories have to be understood as multivariate, short time series with missing data

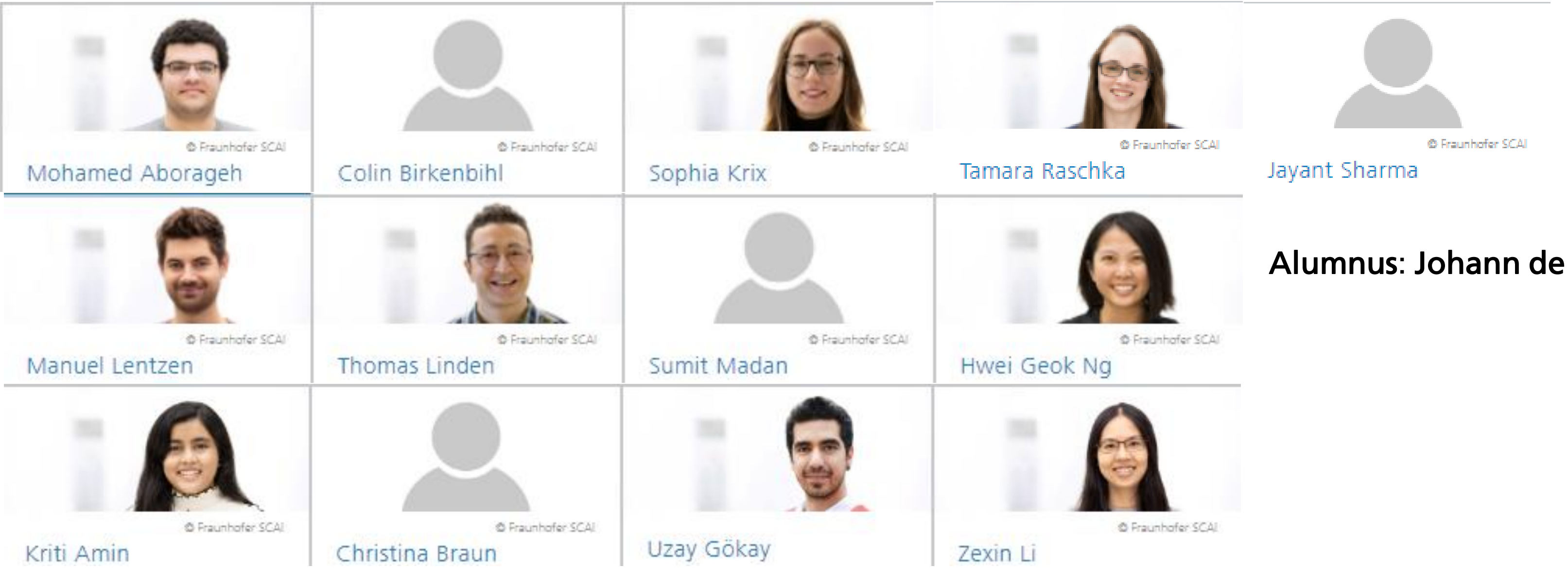
- VADER is a novel approach to cluster such data

Impact:

- Exclude patients that are likely slow progressors
- Use of stratification in statistical analysis at EOS

- Association with molecular mechanisms opens the door to develop subgroup specific treatments

Acknowledgements



Alumnus: Johann de Jong

AETIONOMY

The Virtual Brain Cloud

RADAR-AD

DIGIPD

nfdi4health

intern

Fraunhofer

Backup

ELBO-Criterion for VADER

$$
\begin{aligned}
p(c) &= \mathrm{Cat}(c|\boldsymbol{\pi}) \\
p(\mathbf{z}|c) &= \mathcal{N}\left(\mathbf{z}|\boldsymbol{\mu}_c, \boldsymbol{\sigma}_c^2\mathbf{I}\right) \\
p(\mathbf{x}|\mathbf{z}) &= \mathrm{Ber}(\mathbf{x}|\boldsymbol{\mu}_x) \quad or \quad \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_x, \boldsymbol{\sigma}_x^2\mathbf{I})
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L}_{\mathrm{ELBO}}(\mathbf{x}) &= E_{q(\mathbf{z},c|\mathbf{x})}\left[\log \frac{p(\mathbf{x},\mathbf{z},c)}{q(\mathbf{z},c|\mathbf{x})}\right] \\
&= E_{q(\mathbf{z},c|\mathbf{x})}\left[\log p(\mathbf{x},\mathbf{z},c) - \log q(\mathbf{z},c|\mathbf{x})\right] \\
&= E_{q(\mathbf{z},c|\mathbf{x})}[\log p(\mathbf{x}|\mathbf{z}) + \log p(\mathbf{z}|c) \\
&\quad + \log p(c) - \log q(\mathbf{z}|\mathbf{x}) - \log q(c|\mathbf{x})]
\end{aligned}
$$

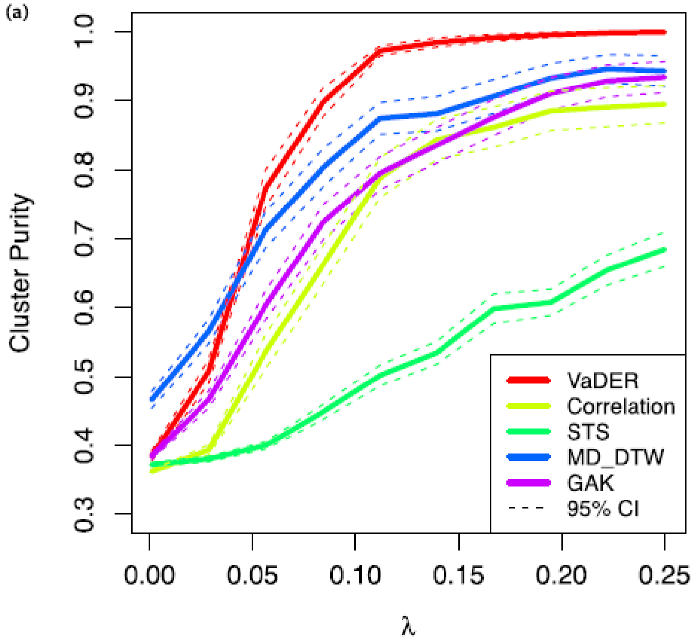
Probabilistic cluster assignment:

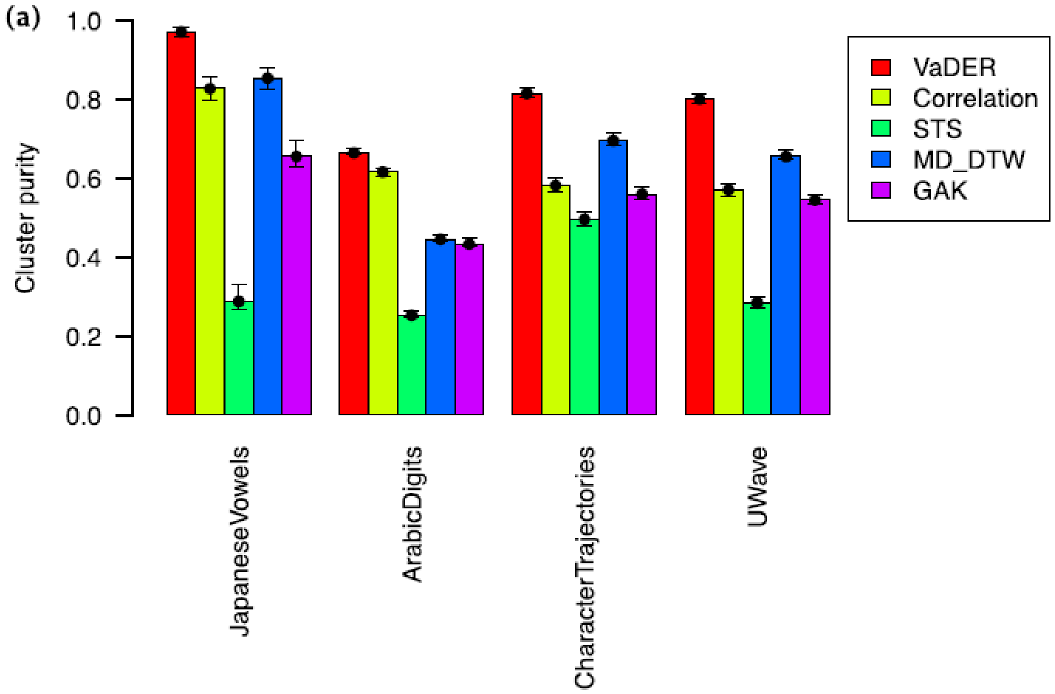
$$
q(c|\mathbf{x}) = p(c|\mathbf{z}) \equiv \frac{p(c)p(\mathbf{z}|c)}{\sum_{c'=1}^{K} p(c')p(\mathbf{z}|c')}
$$

$$
\mathcal{L}_{\mathrm{ELBO}}(\mathbf{x}) = \underbrace{E_{q(\mathbf{z},c|\mathbf{x})}[\log p(\mathbf{x}|\mathbf{z})]}_{\text{(Modified) reconstruction error}} - \underbrace{D_{KL}(q(\mathbf{z},c|\mathbf{x})||p(\mathbf{z},c))}_{\text{Deviance from Gaussian mixture prior = regularization}}
$$

Jiang Z, Zheng Y, Tan H, et al. Variational deep embedding: an unsupervised and generative approach to clustering. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17). 2017:1965–72.

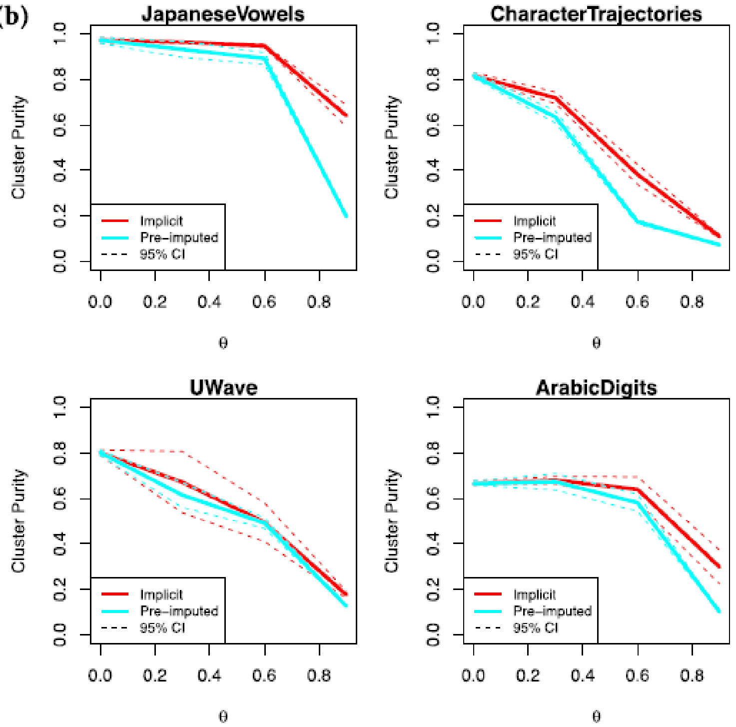
Validation of VADER



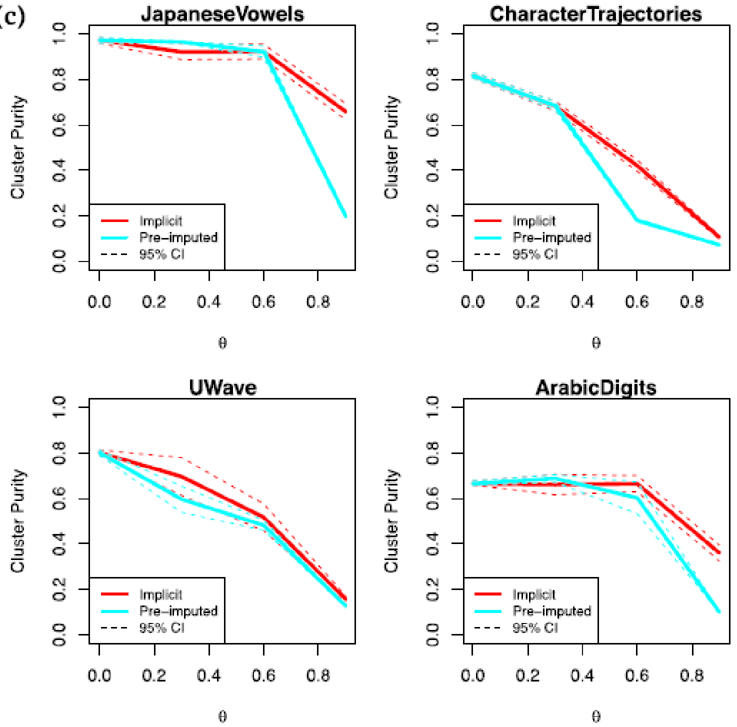


VADER Dependency on Missingness

(b)

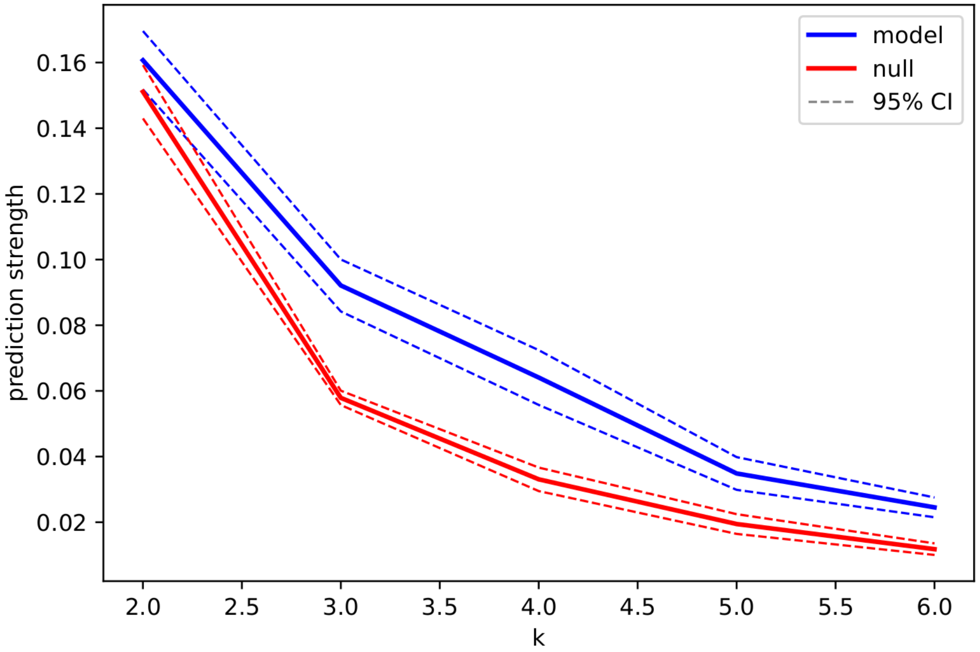


(c)



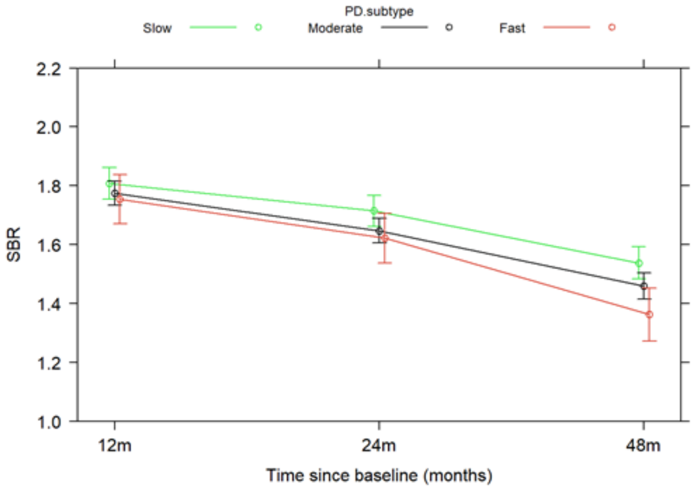
intern

Choice of number of clusters

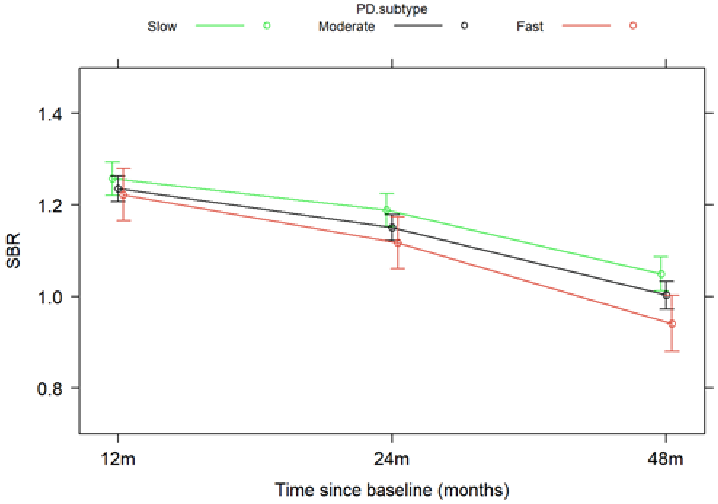


Differences in Dopaminergic Dificiency after 48 months

Caudate



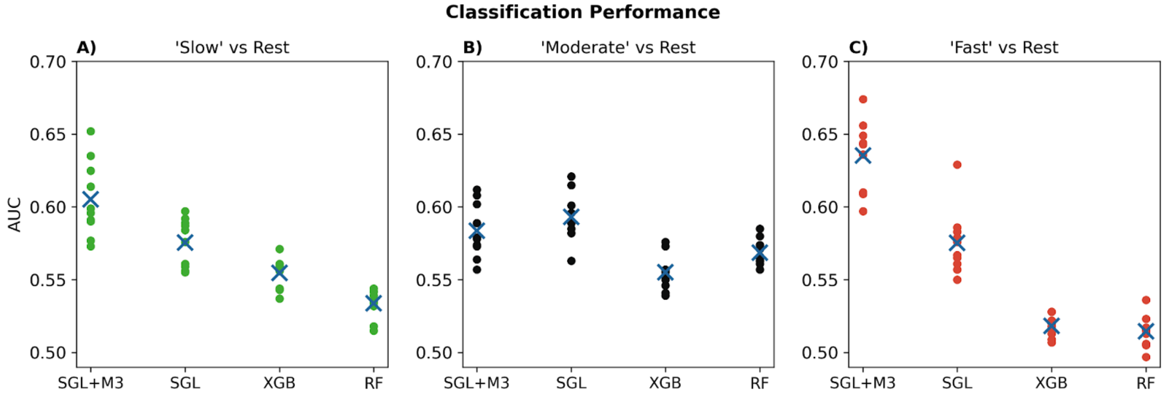
Striatum

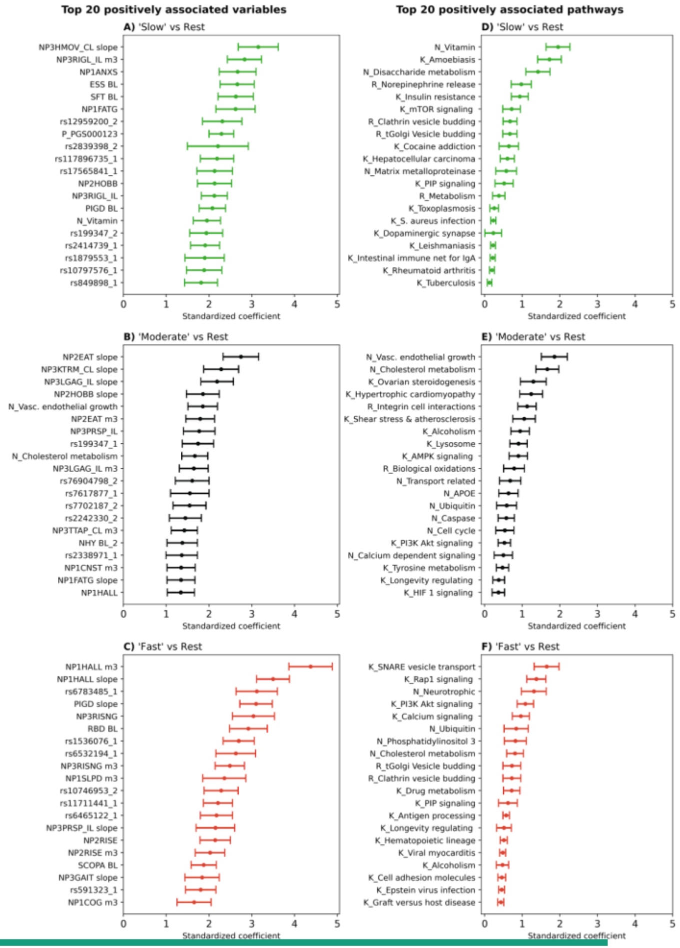


intern

Fraunhofer

Association with Distinct Symptoms and Genetic Loci





intern