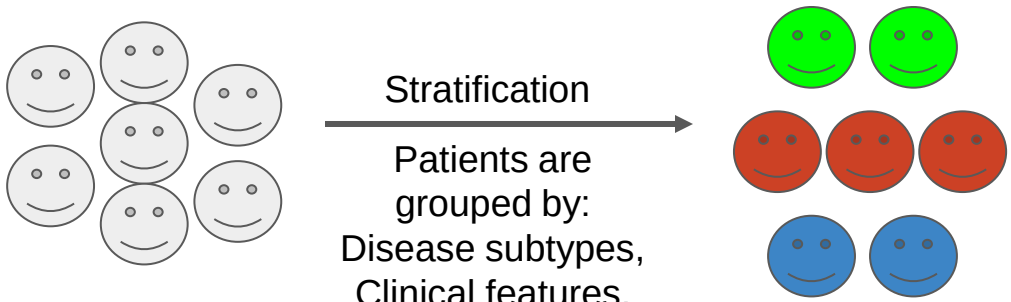
Patient Stratification via building Genomic Profile using Deep Learning in Neurodegenerative Diseases

Asif Emon

Fraunhofer SCAI

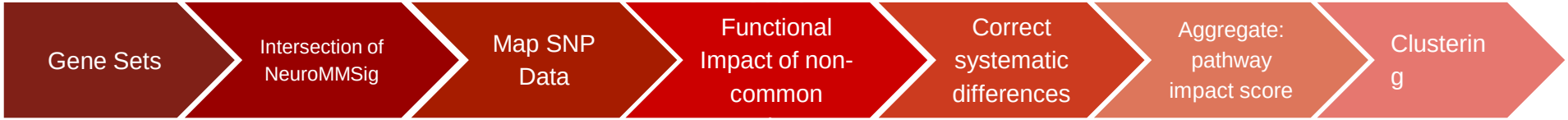
Introduction:

- **Neurodegenerative disease** is an umbrella term for a range of disorders that are characterized by the progressive degeneration of the structure and function of the central or peripheral nervous system.
- **Single nucleotide polymorphisms (SNPs)** are the most common type of genetic variation among people. Each SNP represents a difference in a single DNA building block, called a nucleotide.
- Building **genomic profile** that could corresponds to specific patient subgroups would allow to find/develop better cure to treat these diseases.



Fraunhofer SCAI

Workflow:



• **Pathway Databases**	• **Overlaps with NeuroMMSig common AD and PD genes** • **Common edges between AD/PD**	• **Alzheimer's and Parkinson's Genomic Data** • **Gene mapping of 71K common SNPs:** - dbSNP (proximity) - Cis-eQTL mapping (GTex)	• **Combined Annotation Dependent Depletion (CADD)** • **SIFT** • **PolyPhen2**	• **Batch correction** via FSVA surrogate variable approach	• **Fraction of putatively damaging SNPs per pathway mappable SNPs** **- ssGSEA** **- PCA** **- Autoencoder**	• **Non negative matrix factorization (NMF)**

Fraunhofer SCAI

Autoencoder (AE) neural network:

- Autoencoders are a specific type of feedforward neural networks where the input is the same as the output.
- They compress the input into a lower-dimensional code and then reconstruct the output from this representation in an unsupervised manner.
- AEs are mainly a dimensionality reduction (or compression) algorithm with a couple of important properties: Data-specific, Lossy and Unsupervised.
- Mathematically minimizing the reconstruction error in AE is the same as PCA. However, AE can introduces nonlinearities in the encoding, whereas PCA can only represent linear transformations.

Autoencoder architecture:

An autoencoder consists of 3 components: encoder, code and decoder. The encoder compresses the input and produces the code, the decoder then reconstructs the input only using this code.

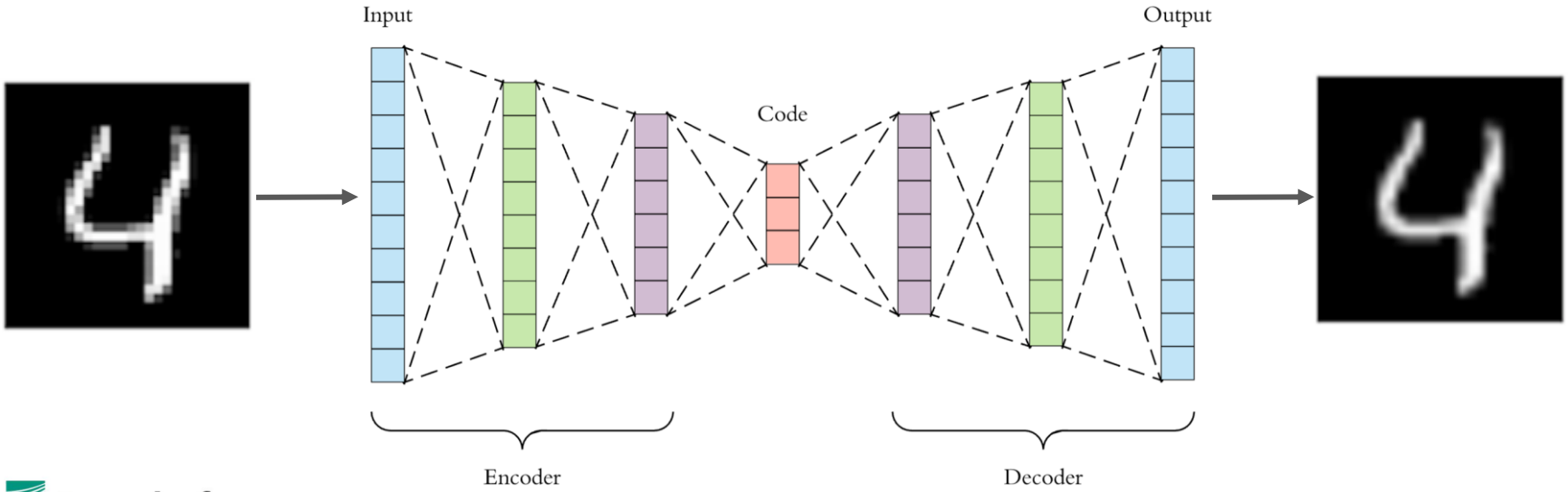


Image courtesy: https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798

Results:

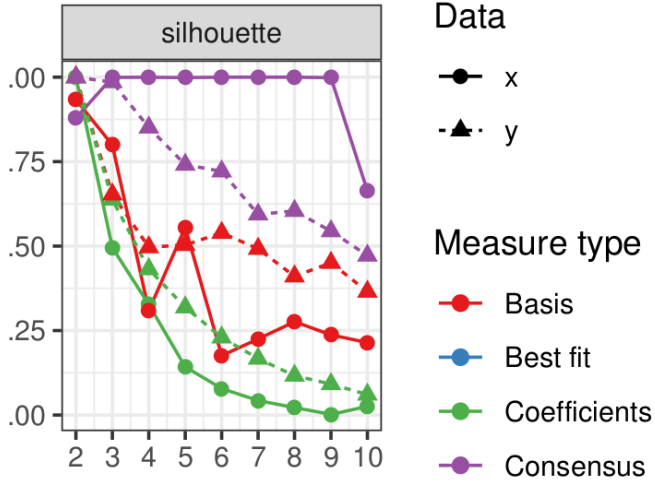
autoencoder + sparse NMF:

Clusters	AD	PD	Total	SI
Cluster 1	95	27	122	1
Cluster 2	119	167	286	1
Clusrer 3	106	32	138	1
Cluster 4	166	132	298	1

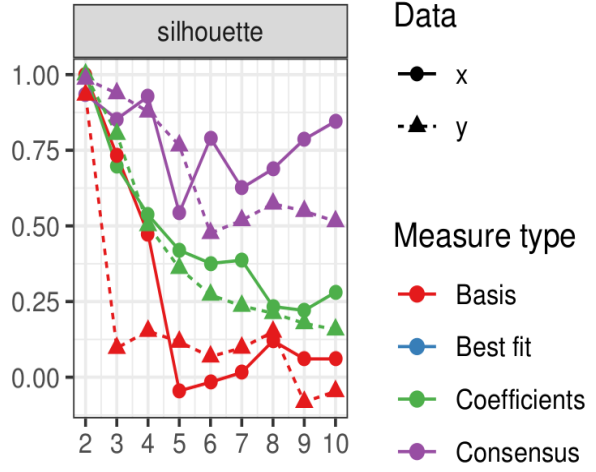
PCA + sparse NMF:

Clusters	AD	PD	Total	SI
Cluster 1	102		102	1
Cluster 2	384	358	742	0.95

Clustering with 10 times shuffled dataset

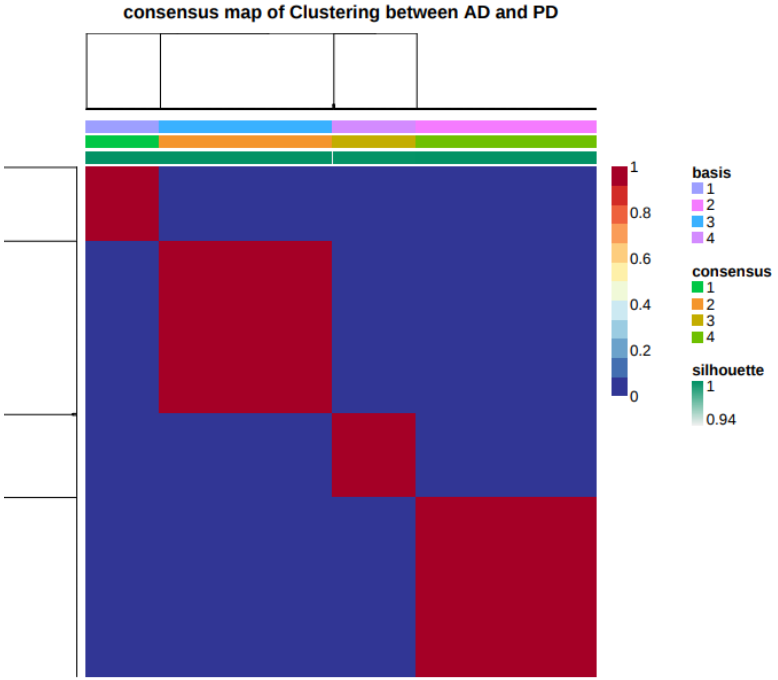


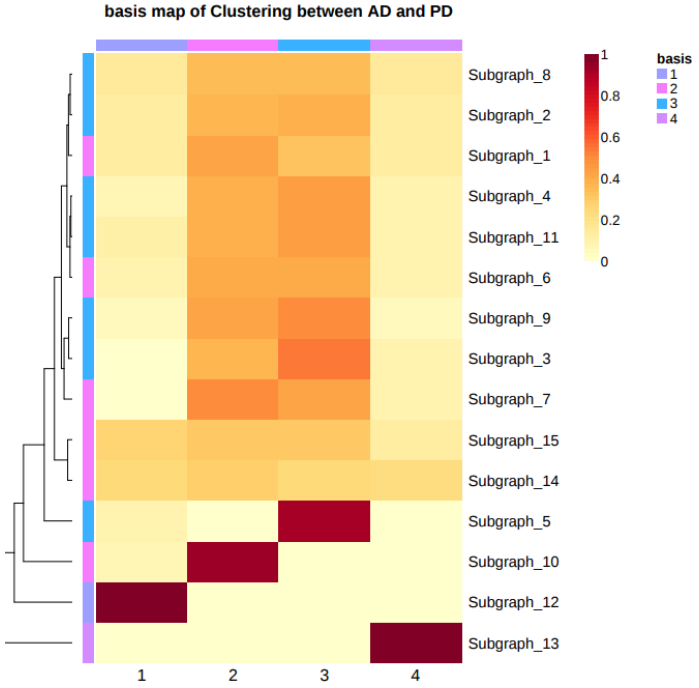
Clustering with 10 times shuffled dataset



Fraunhofer SCAI

Results:





Fraunhofer SCAI

Thanks!