B.R.A.I.N. Lab
Biopsychological Research on Aging, Inflammation, and Neuropsychiatry
Studying Mental Illness and Wellness

Machine learning
Artificial intelligence
Neuro
Imaging
Focusing
On
Longevity &
Dementia
Laboratory

UCL

UC San Diego

Application of XAI to brain age prediction models

Hui Xin Ng

23 May 2024

Overview

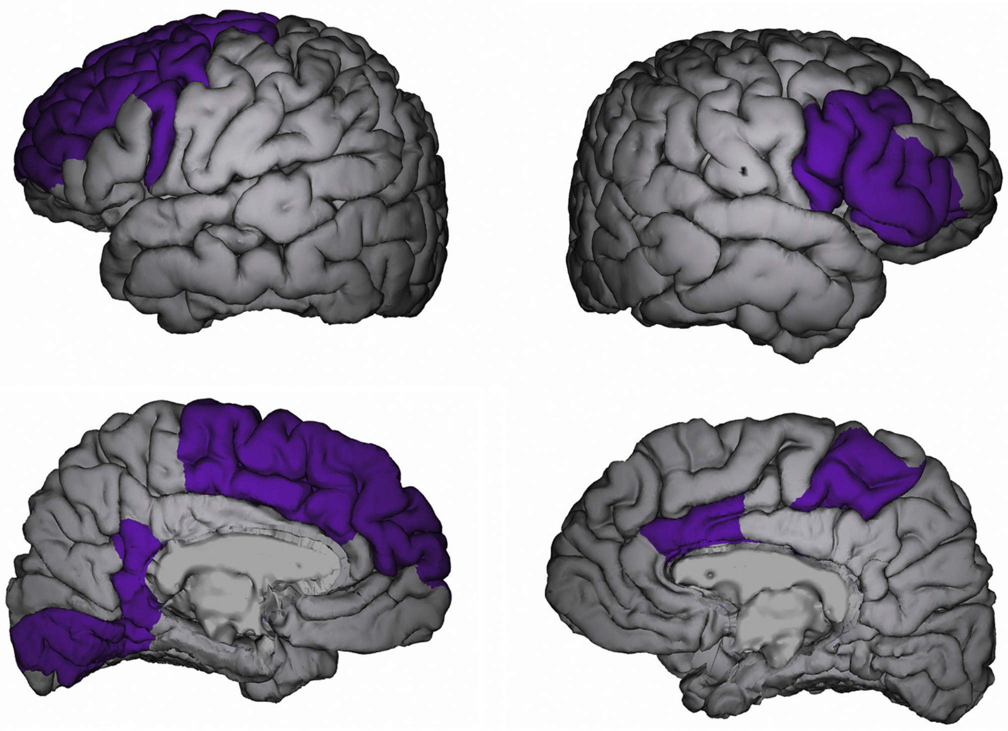
- Predicting age from neuroimaging data
- Interpretability and explainability methods
- Factors to consider when applying interpretability and explainability methods
- Code Demo
- Relevance to multimodal AI

Application of XAI to brain age prediction models

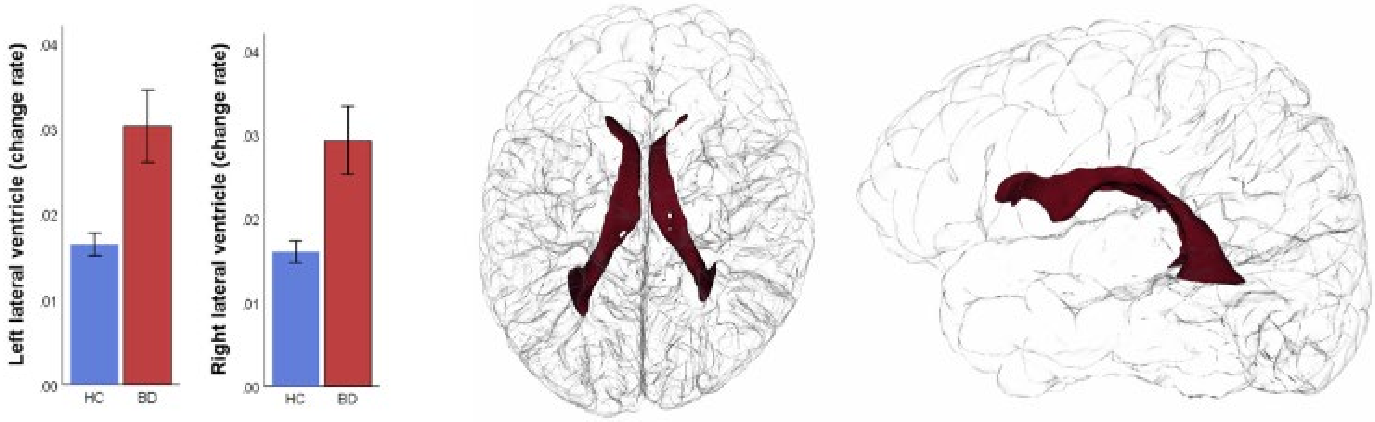
- Predicting age from neuroimaging data
- Interpretability and explainability methods
- Factors to consider when applying interpretability and explainability methods
- Code Demo
- Relevance to multimodal AI

Age-like changes are found in Bipolar Disorder

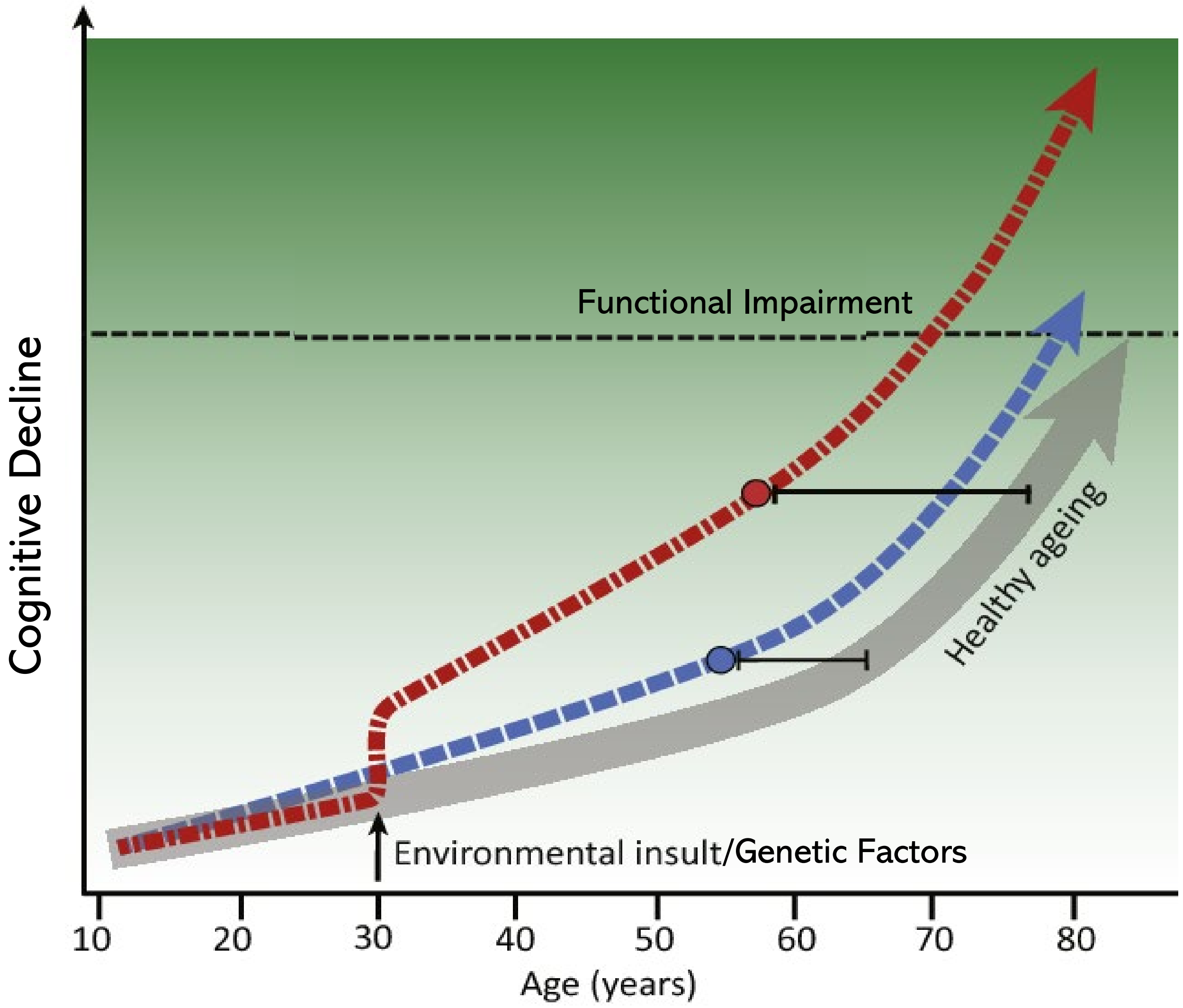
associations with (hypo)manic episodes



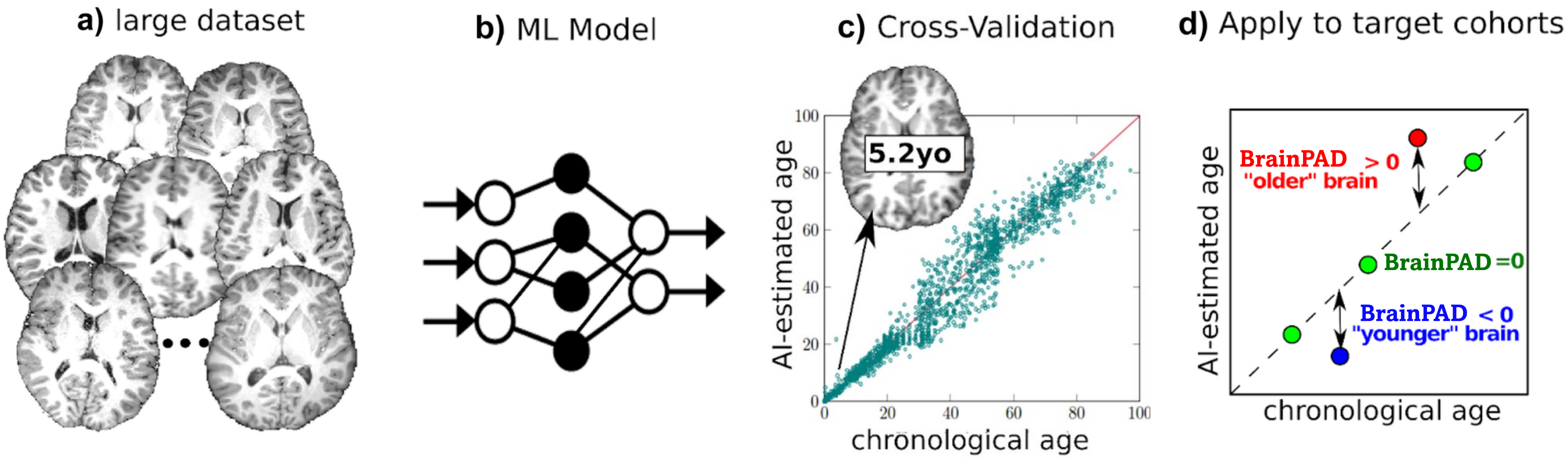
faster ventricular enlargement in BD



Onset of Bipolar Disorder may put individuals at an advanced aging trajectory



Brain Predicted Age Difference (Brain-PAD) as a marker of brain health



BrainPAD = Predicted age – Chronological age

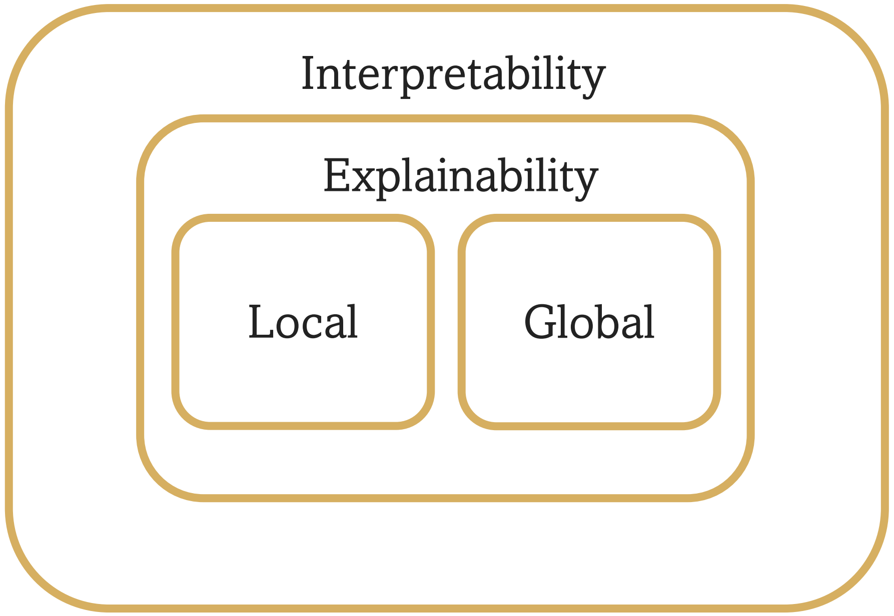
Application of XAI to brain age prediction models

- Predicting age from neuroimaging data
- Interpretability and explainability methods
- Interpretability and explainability challenges in regression tasks
- Code Demo
- Relevance to multimodal AI

Importance of interpretability: Classifying dogs vs. wolves

Definitions of interpretability and explainability

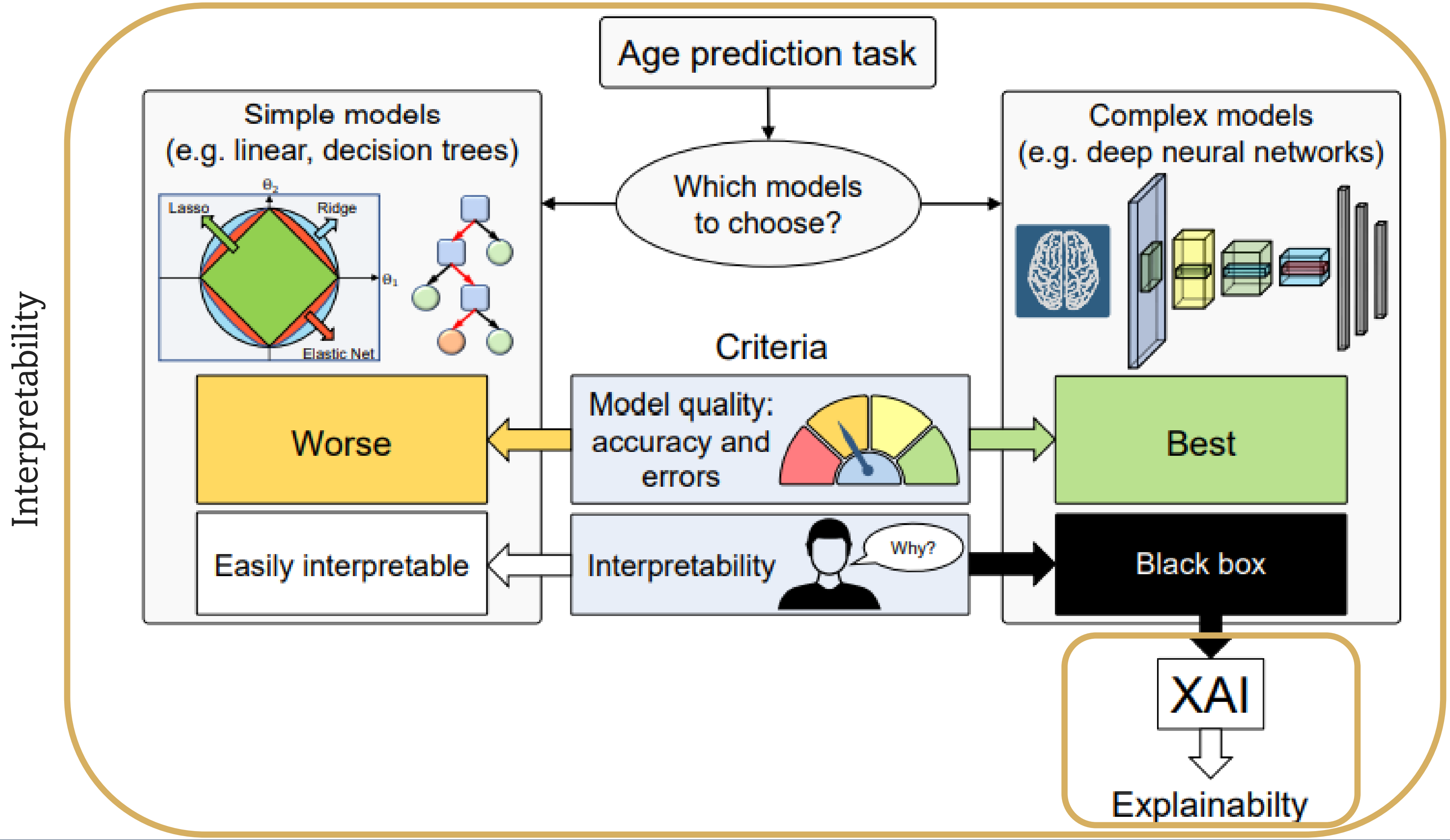
UC San Diego



- Interpretability: Overall ability to understand and make sense of the model's decision-making process.
- Explainability: Model mechanisms and relative importance of features to the model
- Local: Why a model made a specific prediction for a single data point
- Global: Why a model made its predictions across all instances and scenarios?

Performance and interpretability as criteria for model selection in age prediction tasks

UC San Diego



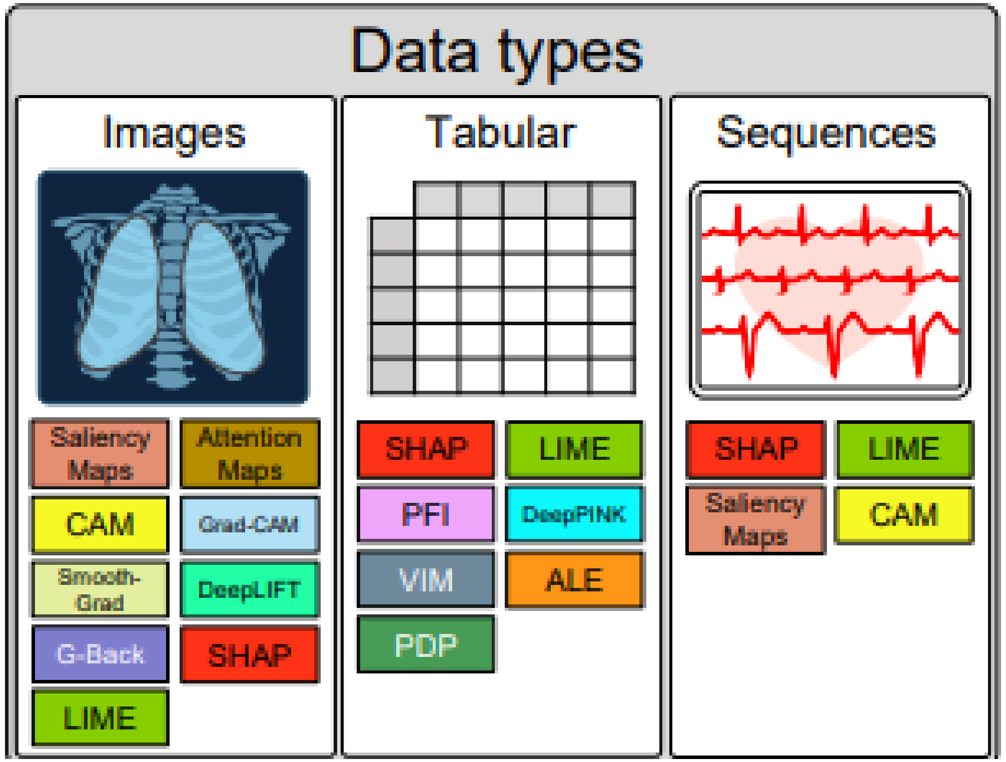
UC San Diego

Application of XAI to brain age prediction models

- Predicting age from neuroimaging data
- Interpretability and explainability methods
- Factors to consider when applying interpretability and explainability methods
- Code Demo
- Relevance to multimodal AI

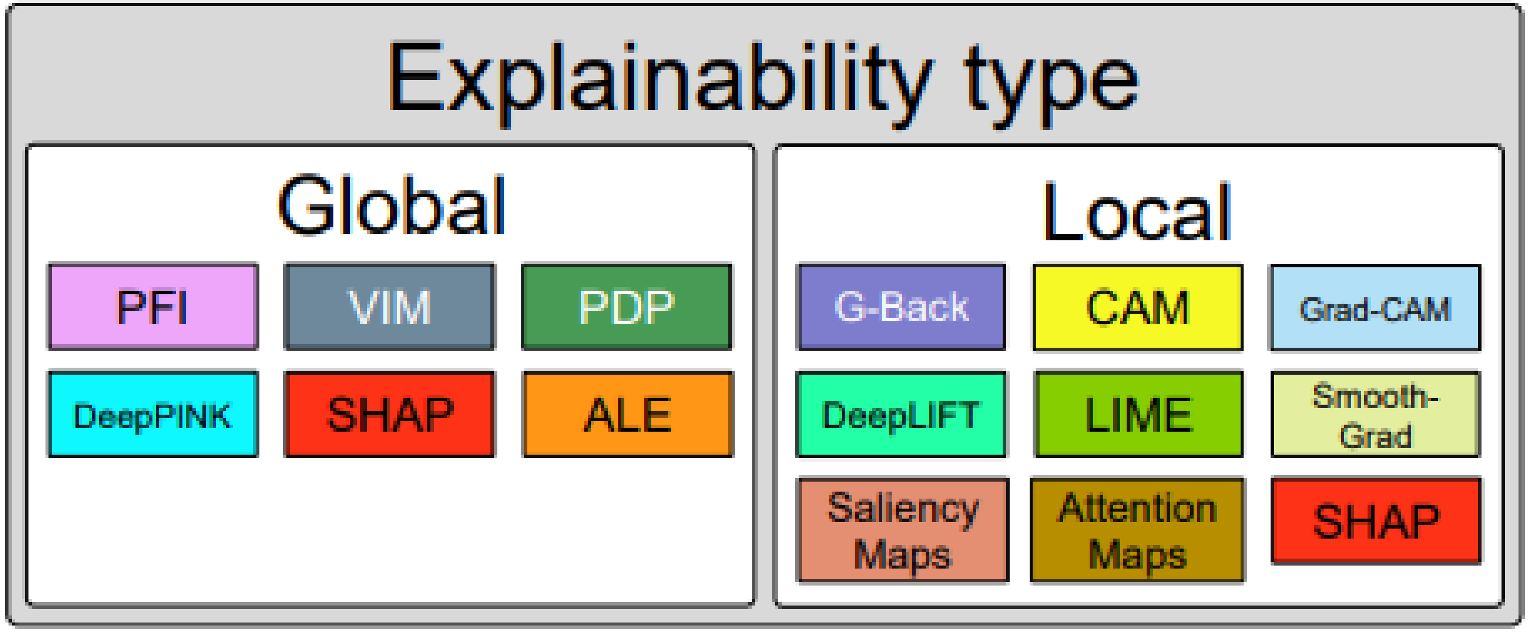
Factors to consider when applying interpretability and explainability methods to medical imaging

- Is the data in the form of **images** or pre-processed **tabular** or **time-series** data?



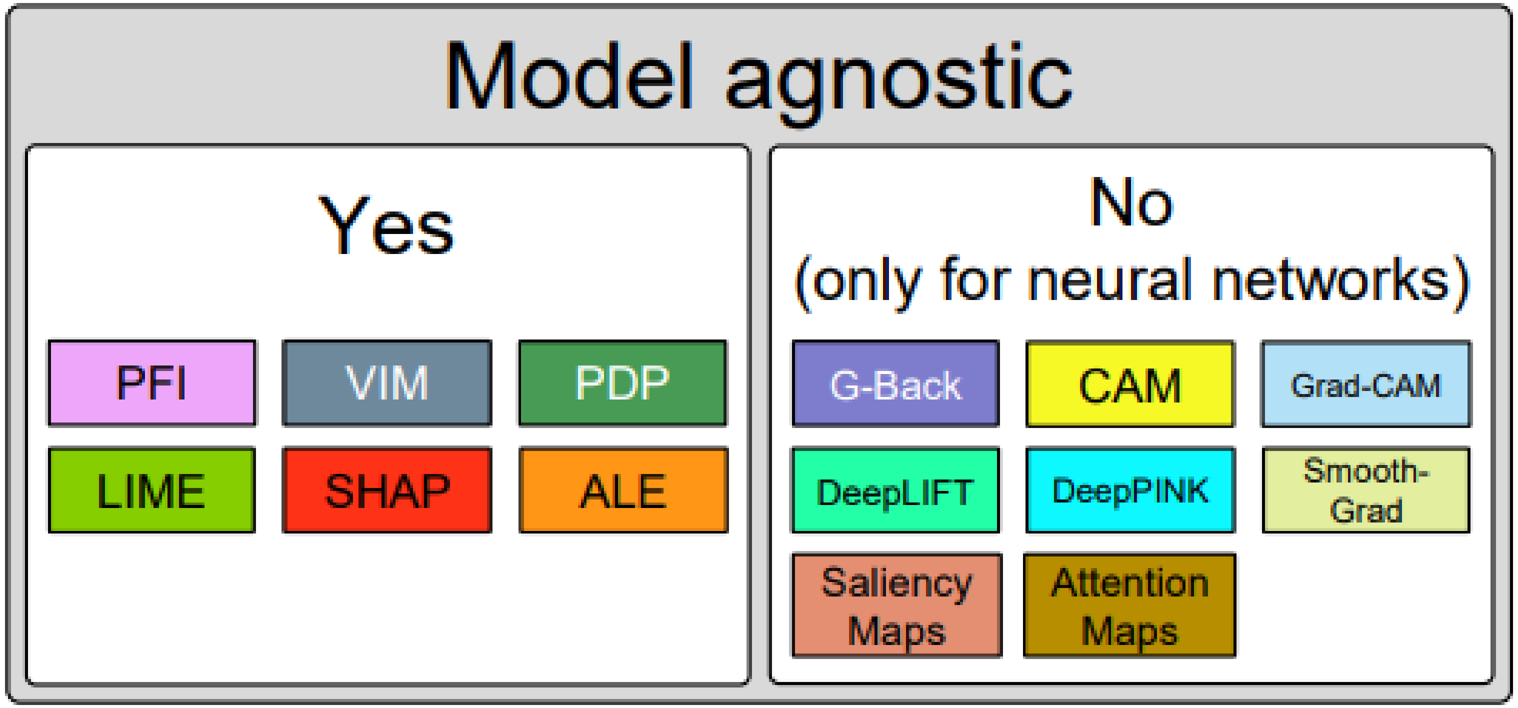
Factors to consider when applying interpretability and explainability methods to medical imaging

- Can the method explain the model's **overall behavior** or explain a **single prediction**?



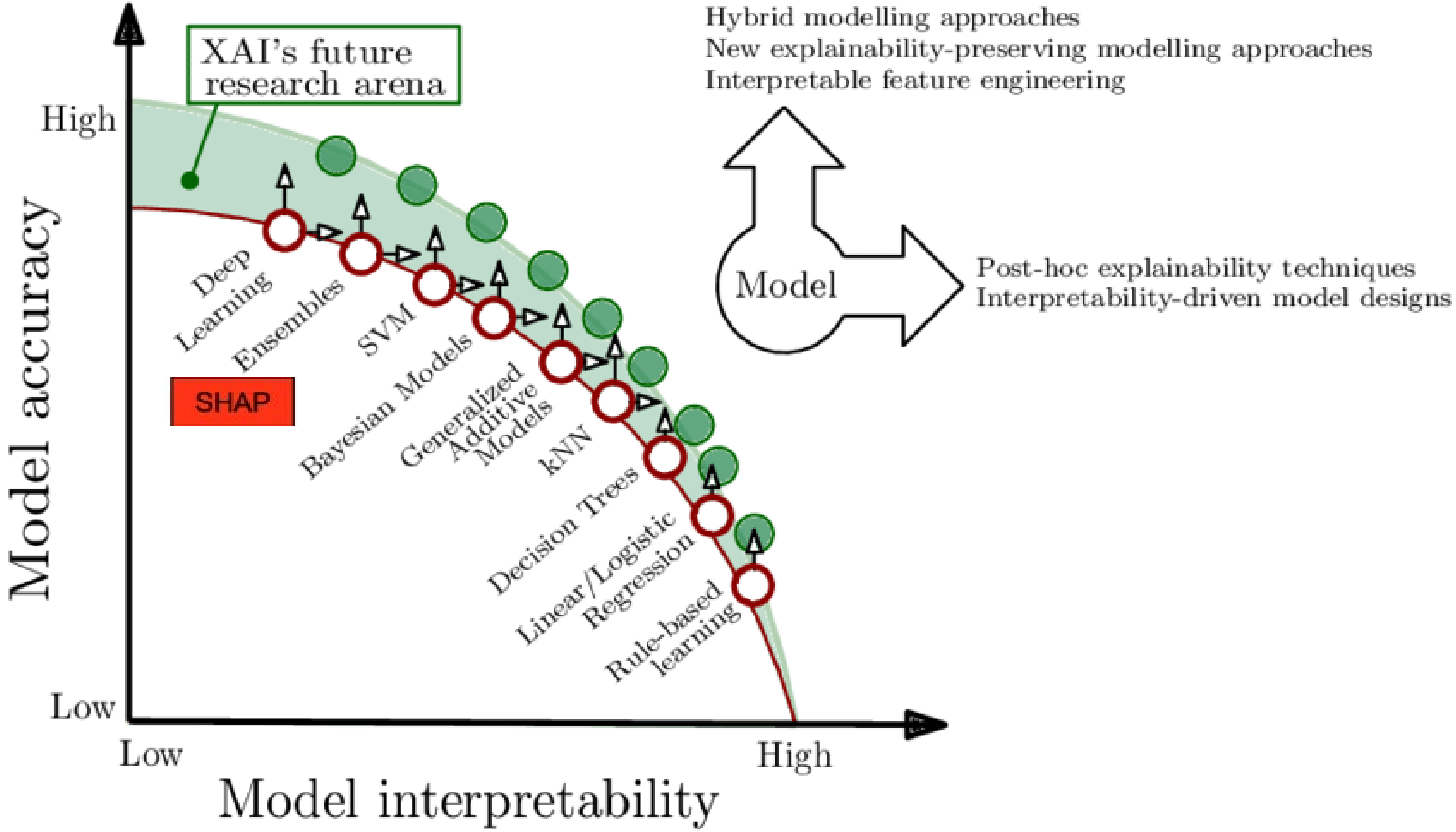
Factors to consider when applying interpretability and explainability methods to medical imaging

- Can this method be applied **independently** of the specific model being used?



Factors to consider when applying interpretability and explainability methods to medical imaging

- Should I use models that are **intrinsically** interpretable or use **post-hoc** interpretability methods?

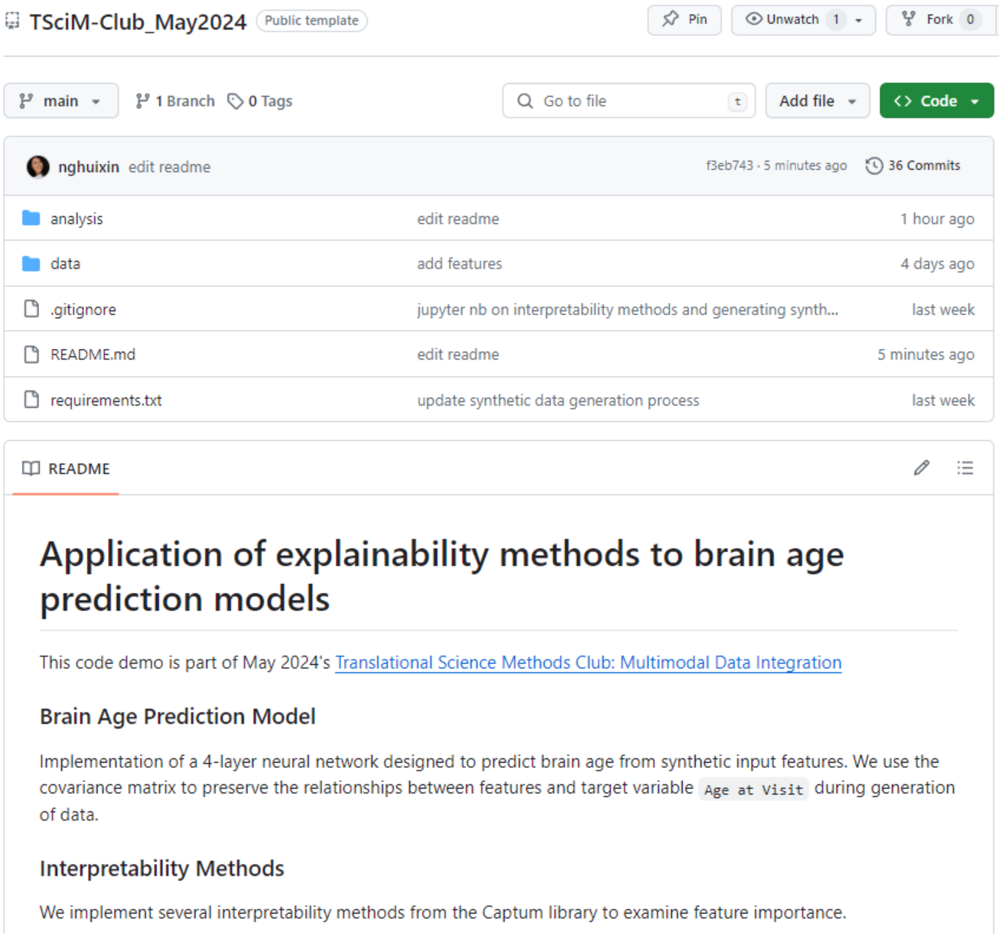


Application of XAI to brain age prediction models

- Predicting age from neuroimaging data
- Interpretability and explainability methods
- Factors to consider when applying interpretability and explainability methods
- Code Demo
- Relevance to multimodal AI

Application of Interpretability methods to a simple neural network model

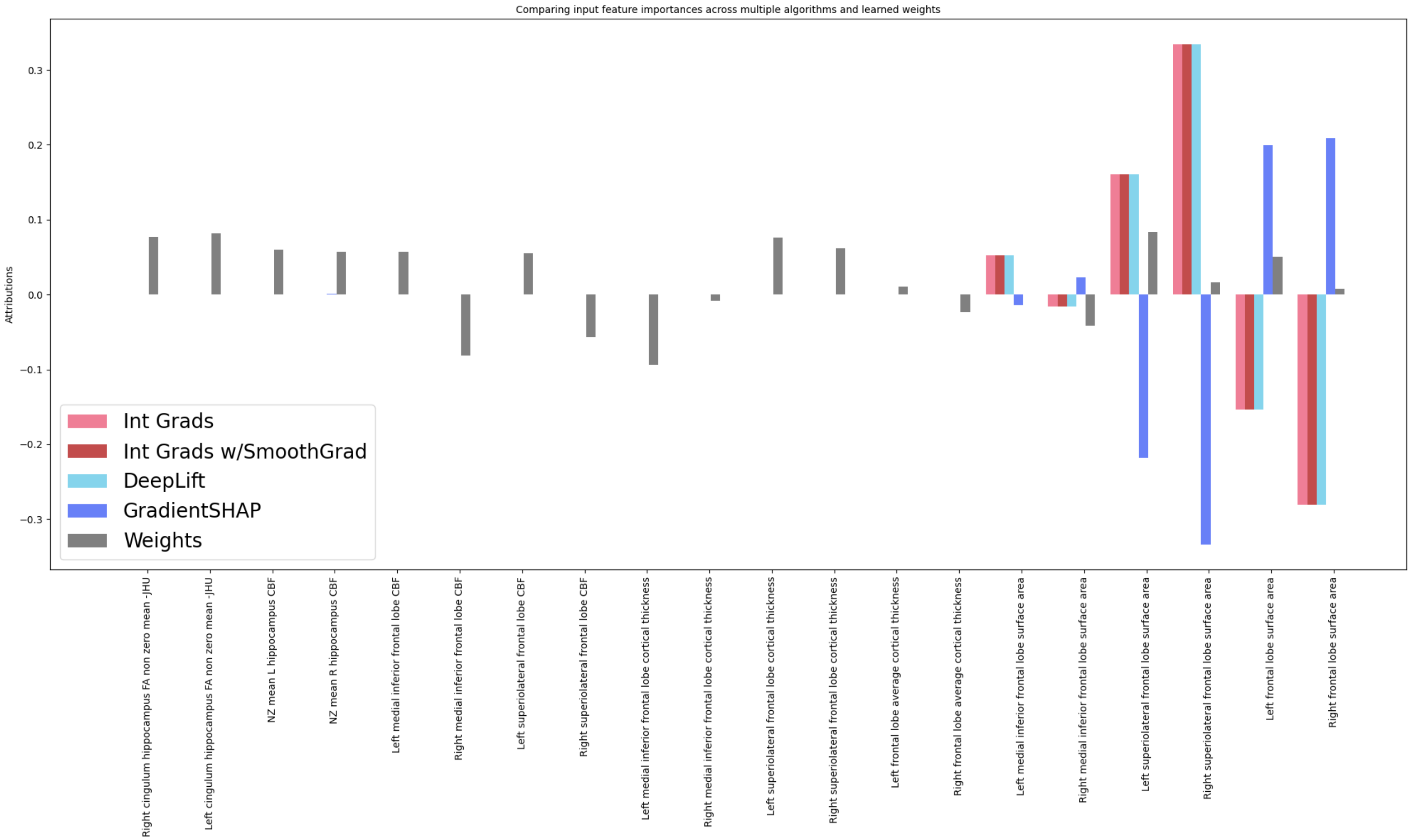
https://github.com/nghuixin/TSciM-Club_May2024



- Link to github repository in the chat
- Captum library helps standardize interpretability methods across projects/papers
- Dataset used:
  - Simulated dataset of healthy comparisons and individuals with BD
  - Brain features and age
- Attribution methods compared:
  - Integrated gradients
  - Integrated gradients with noise tunneling
  - Deep Learning Important FeaTures (DeepLift)
  - GradientSHAP
  - Model weights

Comparison of feature importance across algorithms and learned weights

https://github.com/nghuixin/TSciM-Club_May2024



Application of XAI to brain age prediction models

- Predicting age from neuroimaging data
- Interpretability and explainability methods
- Factors to consider when applying interpretability and explainability methods
- Code Demo
- Relevance to multimodal AI

New tools facilitate our ability to evaluate algorithms on multimodal data

UC San Diego

fusilli: multimodal data fusion Python library

DATA

Image + Tabular

OR

Tabular + Tabular

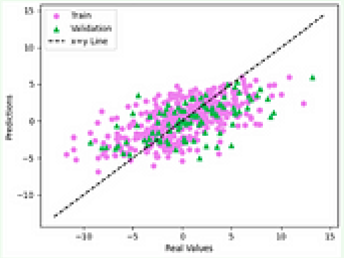
FUSION MODELS

Deep-learning based models:

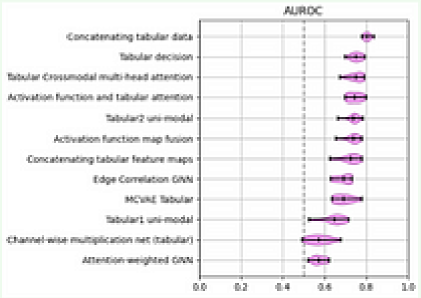
- Unimodal benchmarks
- Attention-based
- Graph Neural Networks
- Autoencoders
- and much more...

EVALUATE

For one model:

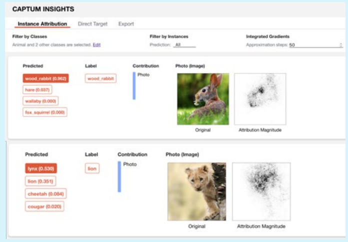


Comparing models:

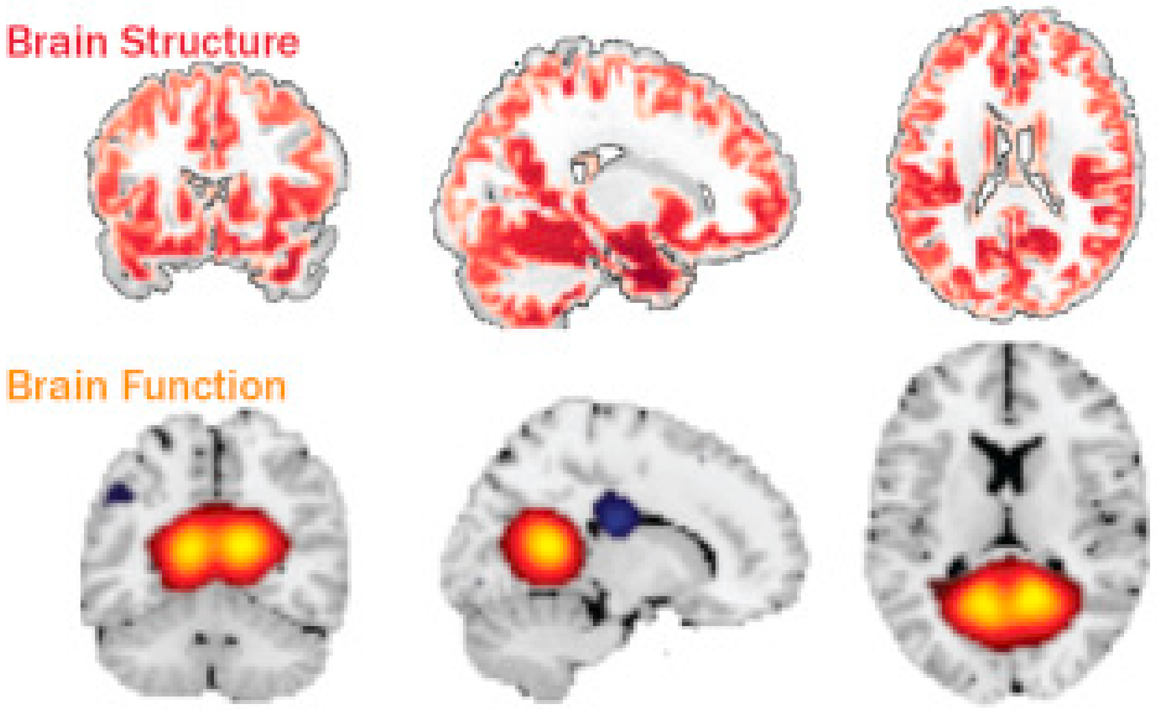


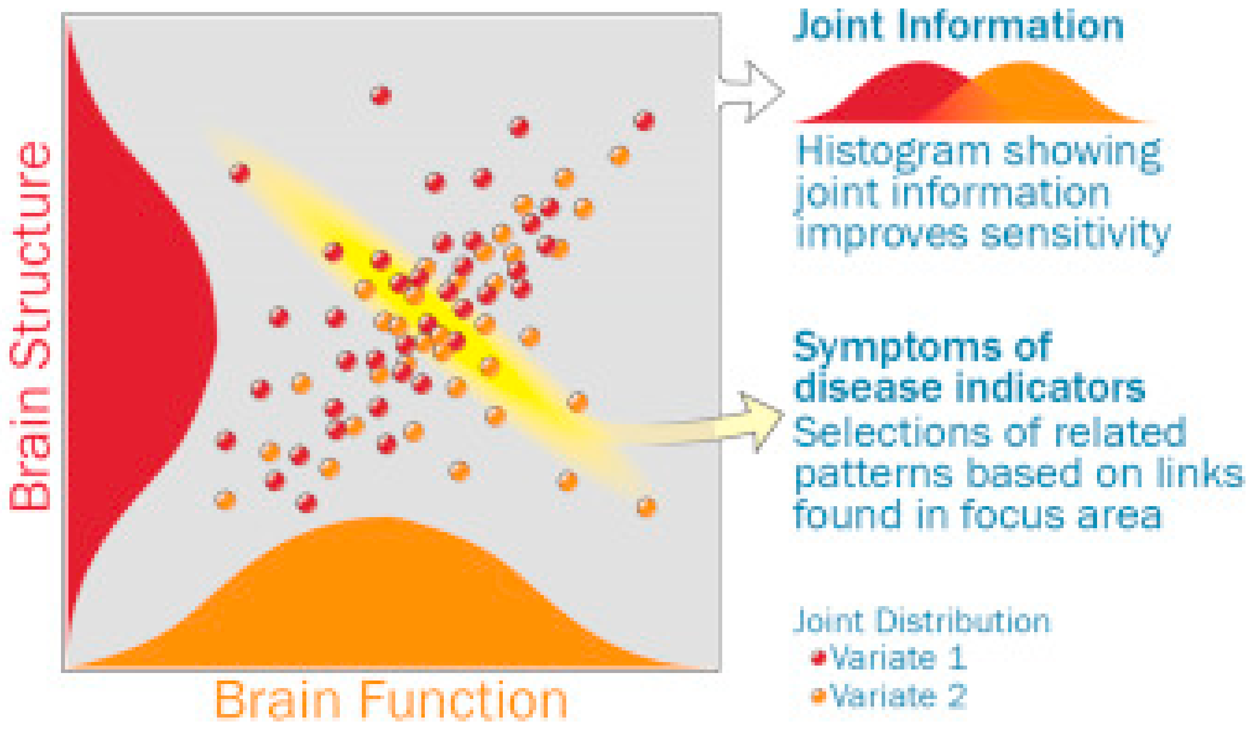
EXPLAIN

Captum

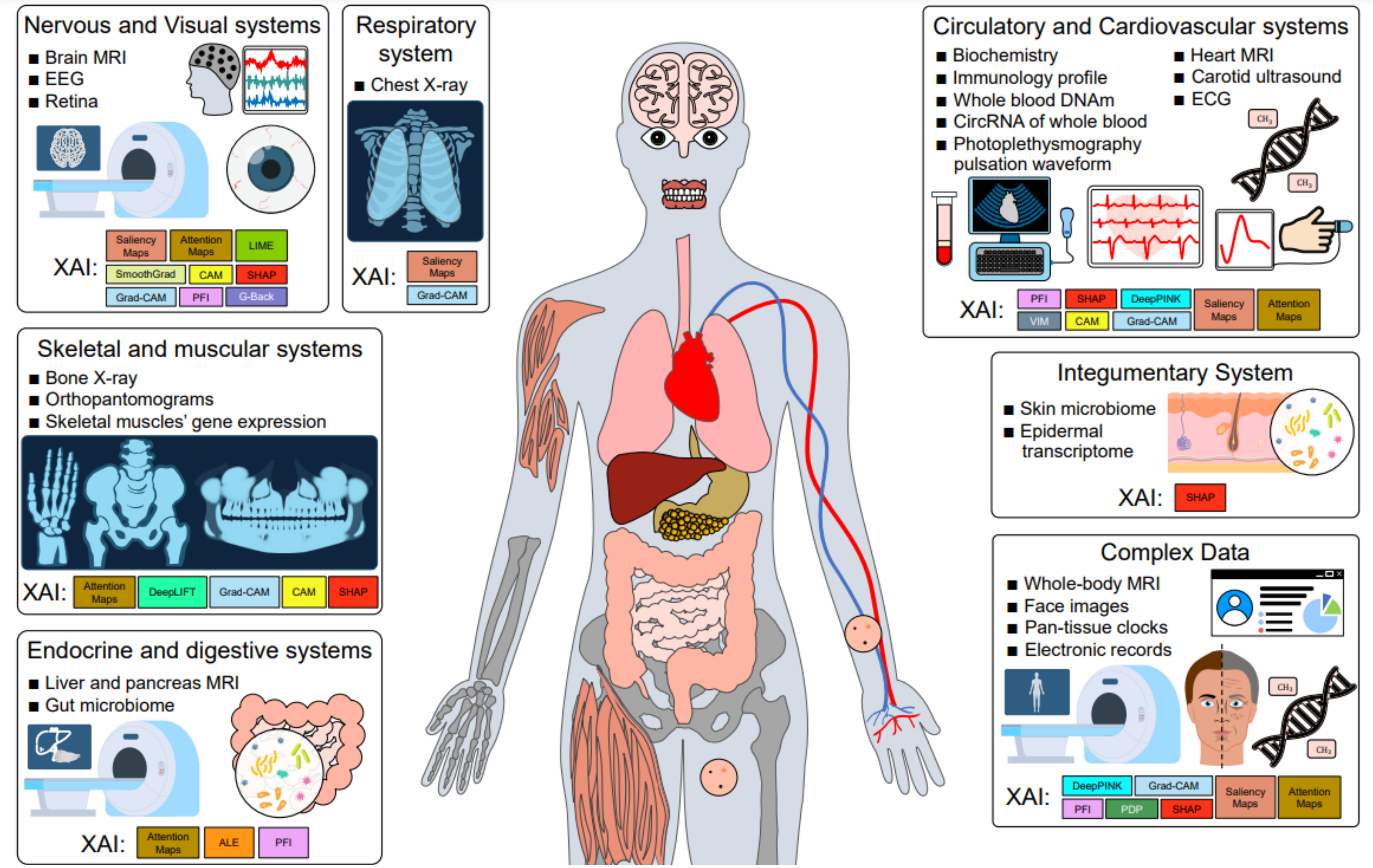


Utilization of multimodal data enhance our ability to distinguish neuropsychiatric illnesses



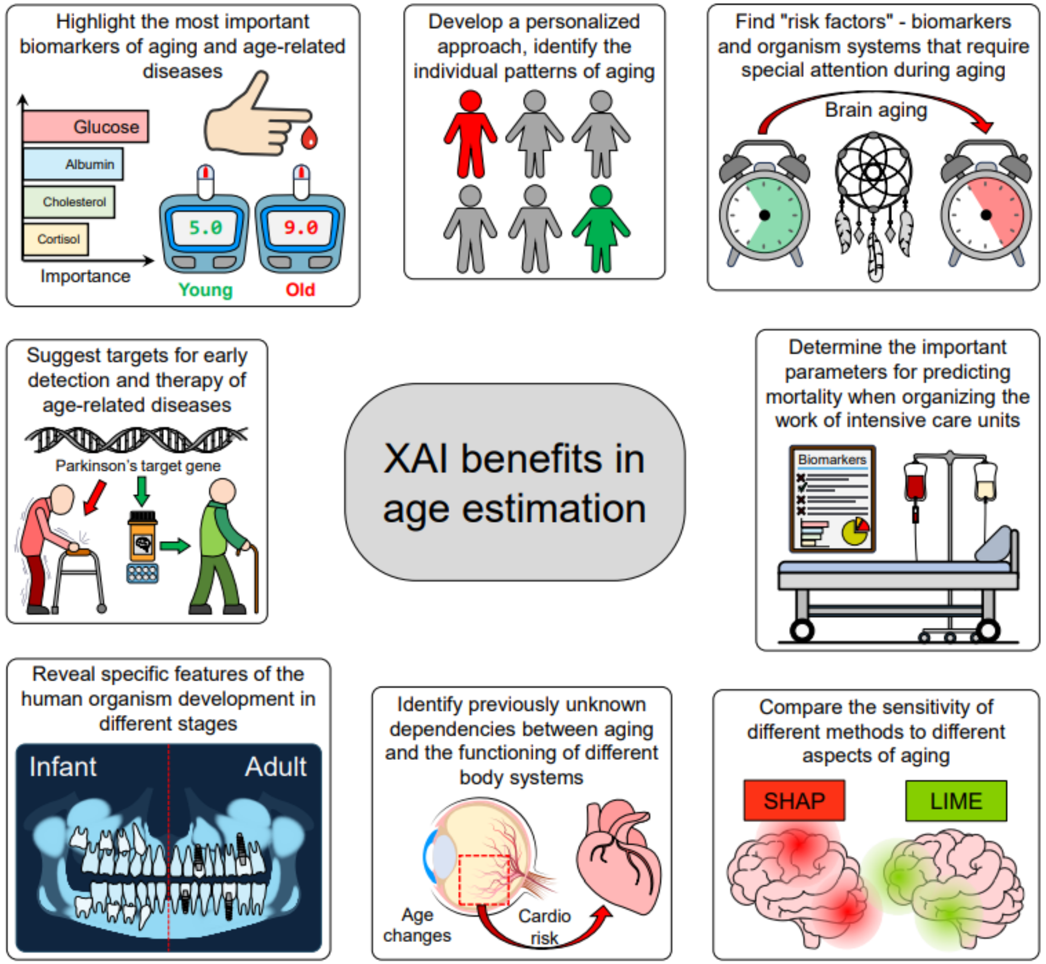


XAI methods applied to age prediction models across organ systems



Added value of applying XAI methods to age prediction tasks

UC San Diego



References

- Abé, C., Ching, C. R. K., Liberg, B., Lebedev, A. V., Agartz, I., Akudjedu, T. N., Alda, M., Alnæs, D., Alonso-Lana, S., Benedetti, F., Berk, M., Bøen, E., Bonnin, C. del M., Breuer, F., Brosch, K., Brouwer, R. M., Canales-Rodríguez, E. J., Cannon, D. M., Chye, Y., … Landén, M. (2022). Longitudinal Structural Brain Changes in Bipolar Disorder: A Multicenter Neuroimaging Study of 1232 Individuals by the ENIGMA Bipolar Disorder Working Group. *Biological Psychiatry*, *91*(6), 582–592. https://doi.org/10.1016/j.biopsych.2021.09.008
- Besse, P., Castets-Renard, C., Garivier, A., & Loubes, J.-M. (2018). *Can Everyday AI be Ethical? Machine Learning Algorithm Fairness (english version)*. https://doi.org/10.13140/RG.2.2.22973.31207
- Calhoun, V. D., & Sui, J. (2016). Multimodal Fusion of Brain Imaging Data: A Key to Finding the Missing Link(s) in Complex Mental Illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(3), 230–244. https://doi.org/10.1016/j.bpsc.2015.12.005
- Cole, J. H., Poudel, R. P. K., Tsagkrasoulis, D., Caan, M. W. A., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, *163*, 115–124. https://doi.org/10.1016/j.neuroimage.2017.07.059
- Florence Townend, Patrick J. Roddy, & Philipp Goebl. (2024). florencejt/fusilli: Fusilli v1.1.0 (v1.1.0). Zenodohttps://doi.org/10.5281/zenodo.10463697
- Kalyakulina, A., Yusipov, I., Moskalev, A., Franceschi, C., & Ivanchenko, M. (2024). eXplainable Artificial Intelligence (XAI) in aging clock models. *Ageing Research Reviews*, *93*, 102144. https://doi.org/10.1016/j.arr.2023.102144
- Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., Popovic, D., Oeztuerk, O., Haas, S. S., Weiske, J., Ruef, A., Kambeitz-Ilankovic, L., Antonucci, L. A., Neufang, S., Schmidt-Kraepelin, C., Ruhrmann, S., Penzel, N., Kambeitz, J., Haidl, T. K., … PRONIA Consortium. (2021). Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression. *JAMA Psychiatry*, *78*(2), 195–209. https://doi.org/10.1001/jamapsychiatry.2020.3604
- Lombardi, A., Diacono, D., Amoroso, N., Monaco, A., Tavares, J. M. R. S., Bellotti, R., & Tangaro, S. (2021). Explainable Deep Learning for Personalized Age Prediction With Brain Morphology. *Frontiers in Neuroscience*, *15*, 674055. https://doi.org/10.3389/fnins.2021.674055
- Singh, N. M., Harrod, J. B., Subramanian, S., Robinson, M., Chang, K., Cetin-Karayumak, S., Dalca, A. V., Eickhoff, S., Fox, M., Franke, L., Golland, P., Haehn, D., Iglesias, J. E., O'Donnell, L. J., Ou, Y., Rathi, Y., Siddiqi, S. H., Sun, H., Westover, M. B., … Gollub, R. L. (2022). How Machine Learning is Powering Neuroimaging to Improve Brain Health. *Neuroinformatics*. https://doi.org/10.1007/s12021-022-09572-9
- Yang, G., Ye, Q., & Xia, J. (2022). Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *An International Journal on Information Fusion*, *77*, 29–52. https://doi.org/10.1016/j.inffus.2021.07.016

Acknowledgements

The Alan Turing Institute

Sarah Buehler

Yuhan Wang

Prof. Lisa Eyler

Prof. Jim Hollan

Prof. Deanna Greene

Prof. Jeremy Elman

Prof. Terry Jernigan

UC San Diego School of Medicine

Department of Radiology, Center for Functional MRI

Machine learning Artificial intelligence Neuro Imaging Focusing On Longevity & Dementia Laboratory

B.R.A.I.N. Lab

Biopsychological Research on Aging, Inflammation, and Neuropsychiatry

Studying Mental Illness and Wellness

UCL

Dr. Federica Klaus

Ashley Sutherland

Prof. James Cole

Dr. Richard Merkin

ENIGMA