

Week 10 - Predictive modeling

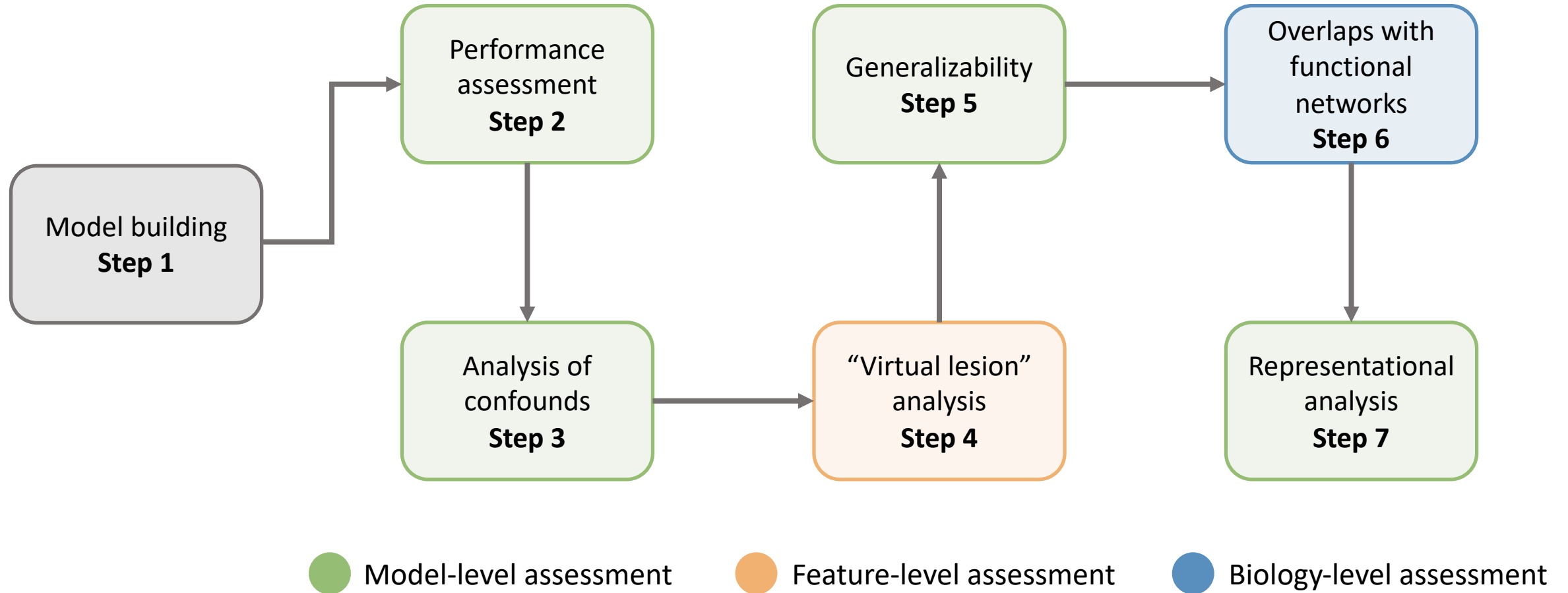
L10-03. Predictive modeling in practice – model interpretation

Lada Kohoutová

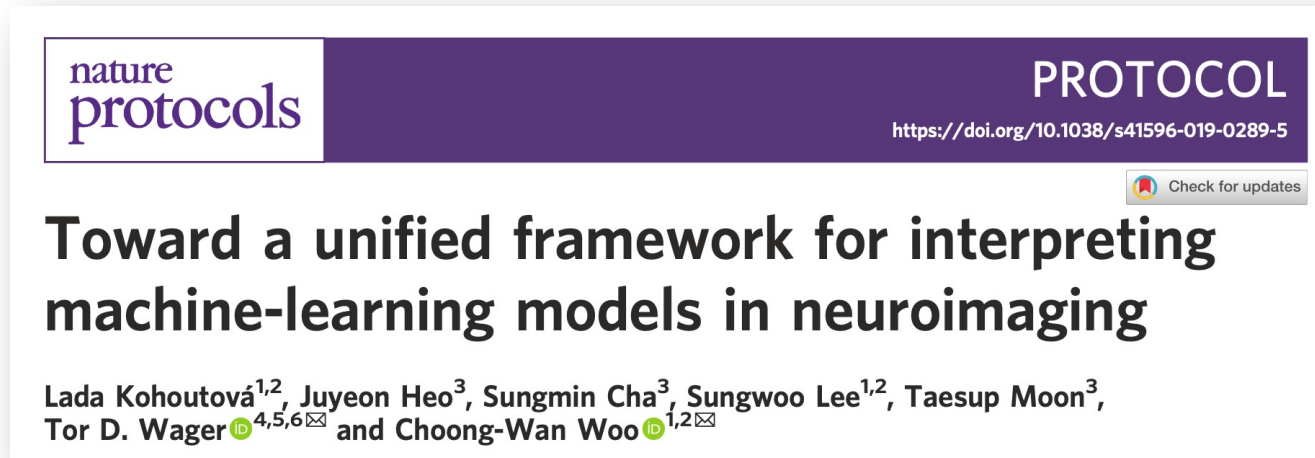
30 April 2021



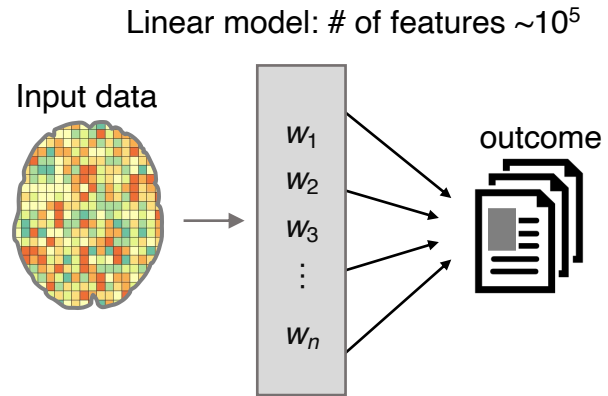
Tutorial workflow



What is interpretability and why we need it

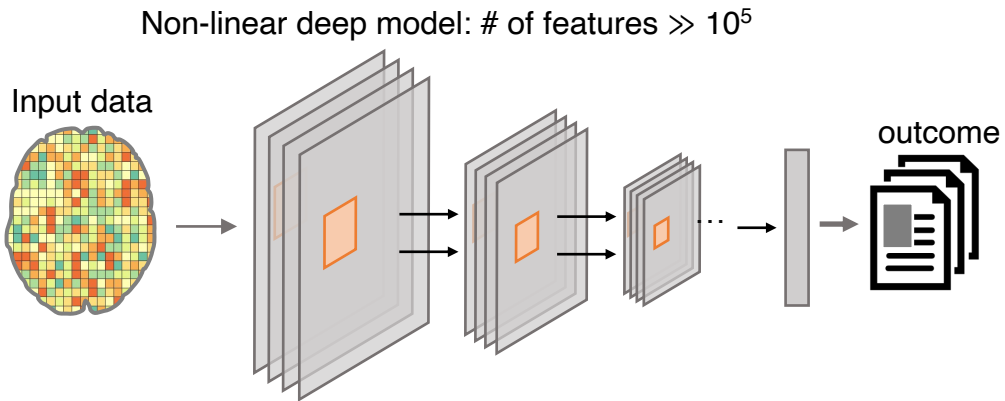


What is interpretability and why we need it



If we do not know *why* and *how* a model works, we cannot answer the following questions:

- When will the model fail?
- To which individuals or subgroups does it apply?
- How can it advance our understanding of the neurobiological mechanisms?



To be interpretable, the models should:

- A. Be **readable and understandable** to humans
- B. Provide **useful information about neural representations** and mechanisms of mental constructs and behaviours
- C. Show that **predictions are based on true signal**, not noise

Figure 1a in the paper



Model interpretation framework

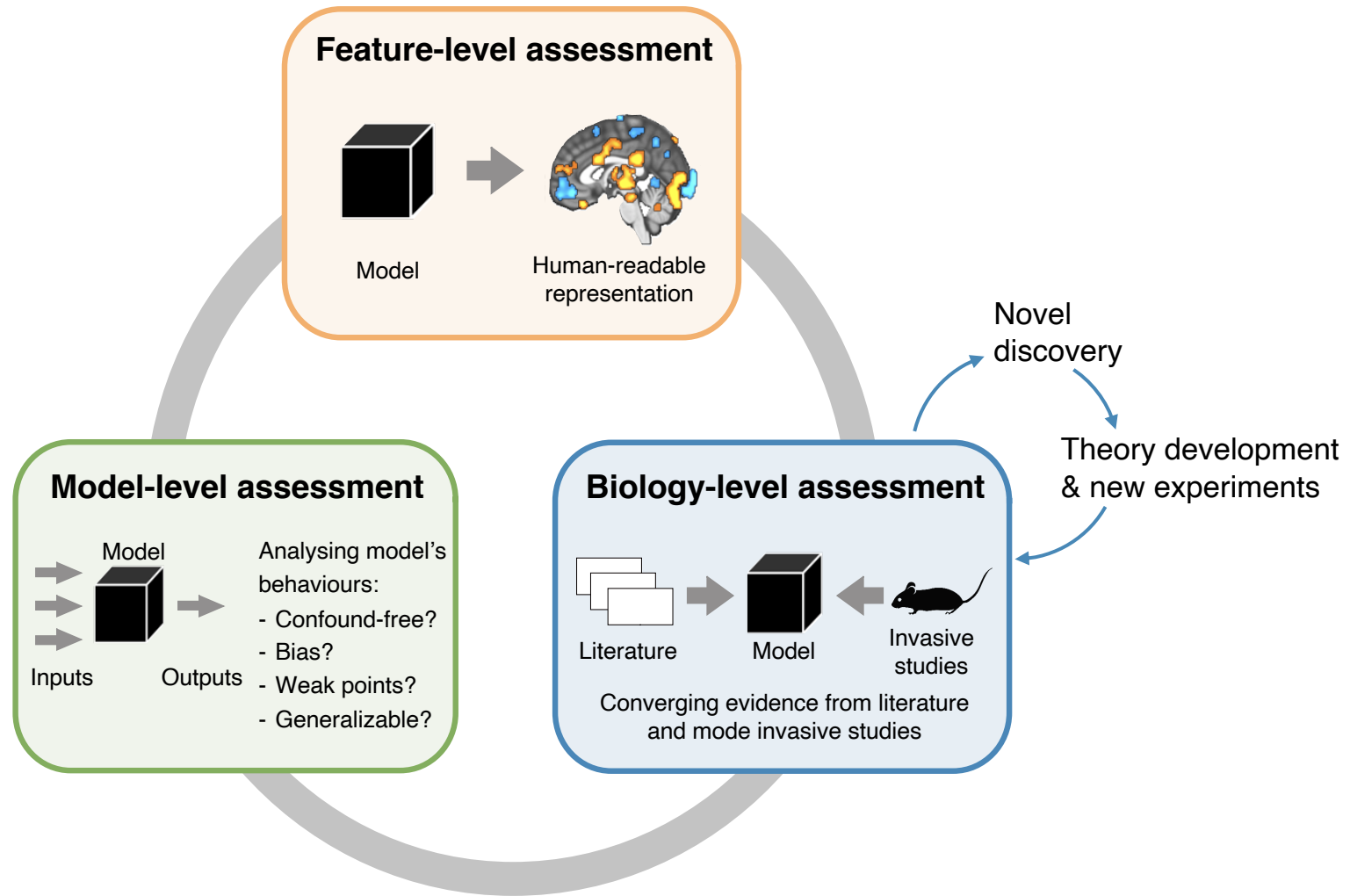
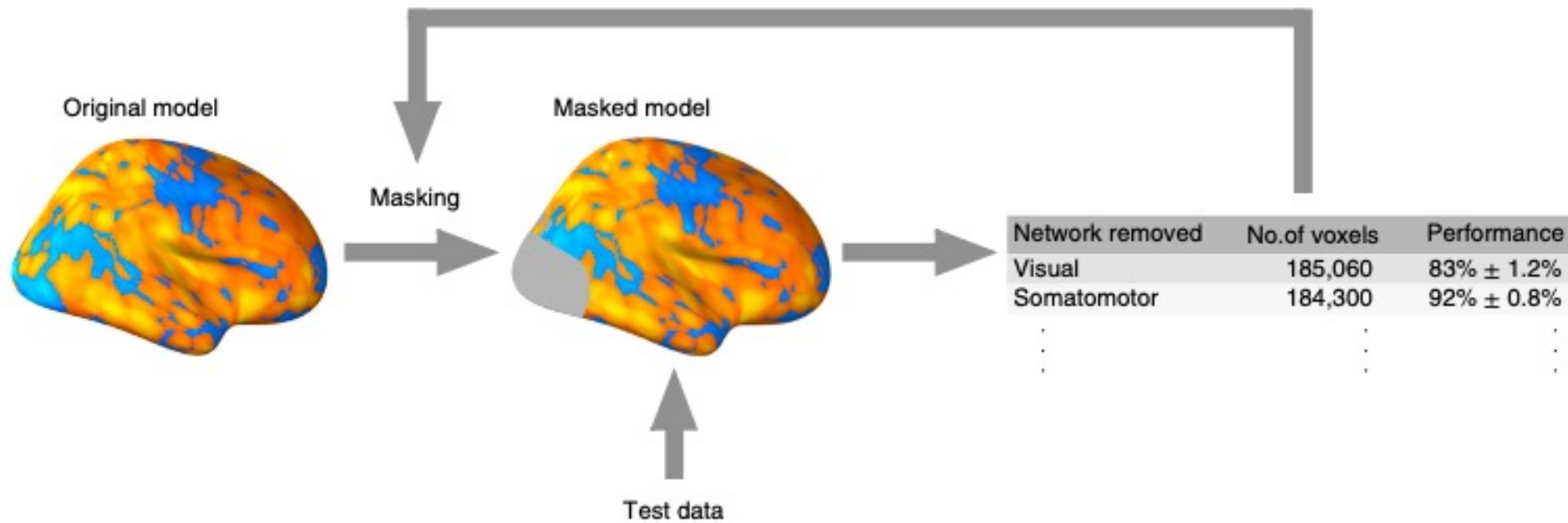


Figure 1b in the paper



“Virtual lesion” analysis

- Predicting with a reduced model (Chang et al., 2015)



Representational analysis

- Comparing the performance of different models on the same datasets

