

AIDD Program School

Son Ha



czodrowskilab.org

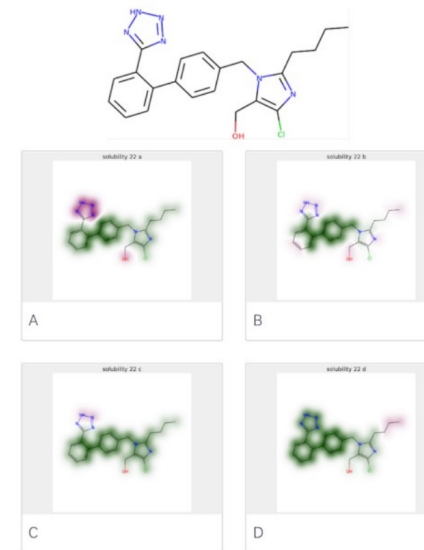
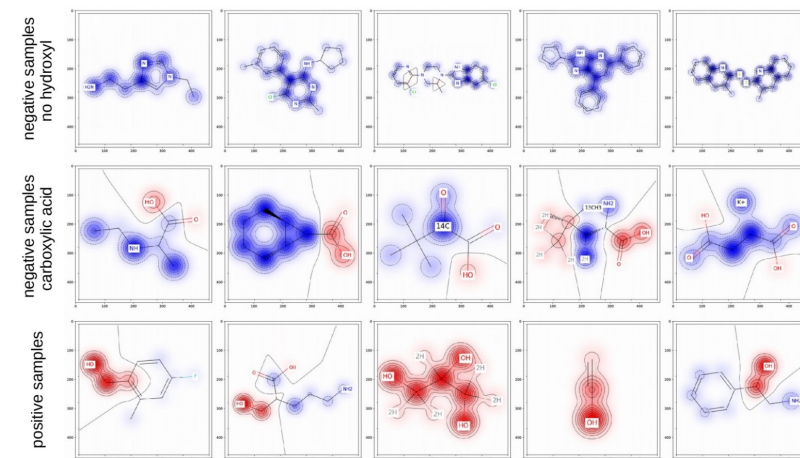


github.com/czodrowskilab

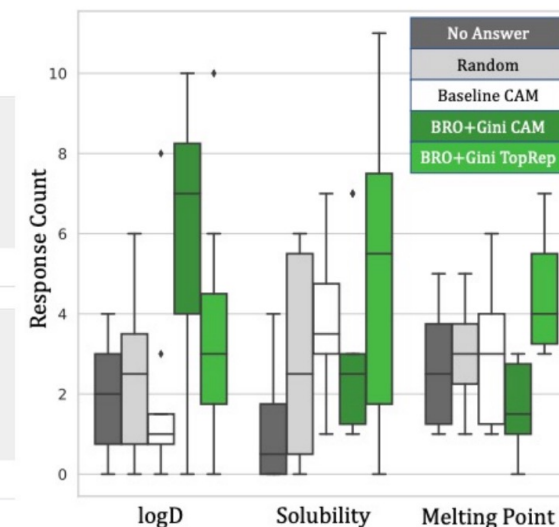
- **AIDD**: Advance machine learning for Innovative Drug Discovery EU program.
- School: Every 6 months, over 1 week of talks from research groups/companies.
- **Everyone** can join via Zoom.
- **Powerpoints** available online: <https://ai-dd.eu/lectures>


















- **Explainability for molecular neural network**, Floriane Montanari
- **Explainable AI (XAI):**
 - Understands and interpret Machine Learning models.
 - Why? ML models are **good at prediction**, but notoriously **hard to interpret**
- **Aim (in Drug Discovery):**
 - Model debugging (is model learning what it is supposed to?)
 - How to modify the structure to improve desired properties.
- **Example 1:** Interpreting FNN on ECFP (upper figure).
 - Prediction: Alcohol Group
 - Integrated Gradient helps identify **which ecfp bits are important**
 - Red: Responsible for positive prediction. Blue: negative
- **Example 2:** How Bayer checks model produce a **good interpretation** (lower figures)
 - Comparing 4 models and their interpretation.
 - **Ask chemists** to choose which interpretation makes sense by a questionnaire.
 - Feedbacks from experts will be used to improve models.
- (Slides: https://ai-dd.eu/sites/default/files/school-2/Montanari_XAI.pdf)



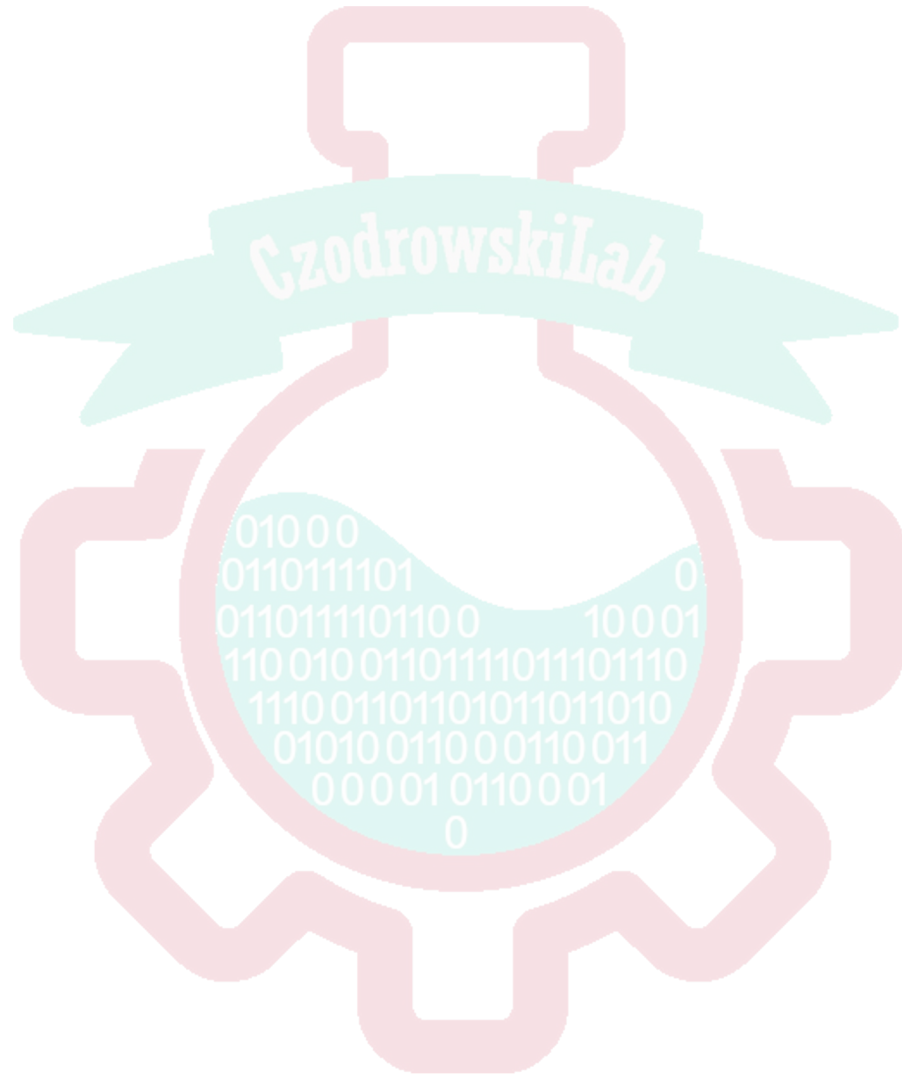
☐ No answer convinces me





#	Team	Members	Score	Entries	Last	Code
1	Sparse People	   	0.73336	6	15d	
2	DAMPP	  	0.73324	17	15d	
3	Winner, Winner, Chicken Dinner	  	0.72604	5	15d	
4	AnaVincenzoMikhailPaula	   	0.72450	6	15d	
	Naive MLP		0.70933			

- Goal: Domain adaptation: Building a model that withstands domain shift
- Domain shift: Compounds in the train set are from a completely different chemical space than test set.
- Fun fact: the best model is a simple Logistic Regression (with no tricks!)
- *(Lesson: Don't count out 'simple' models!)*



Thanks you for your attention!



czodrowskilab.org



github.com/czodrowskilab