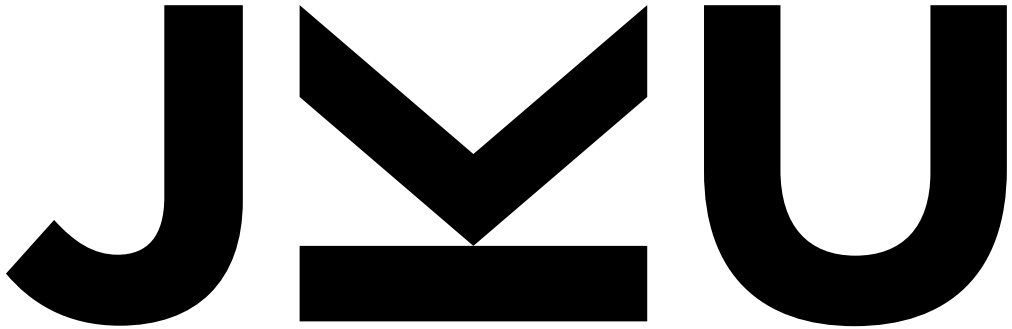
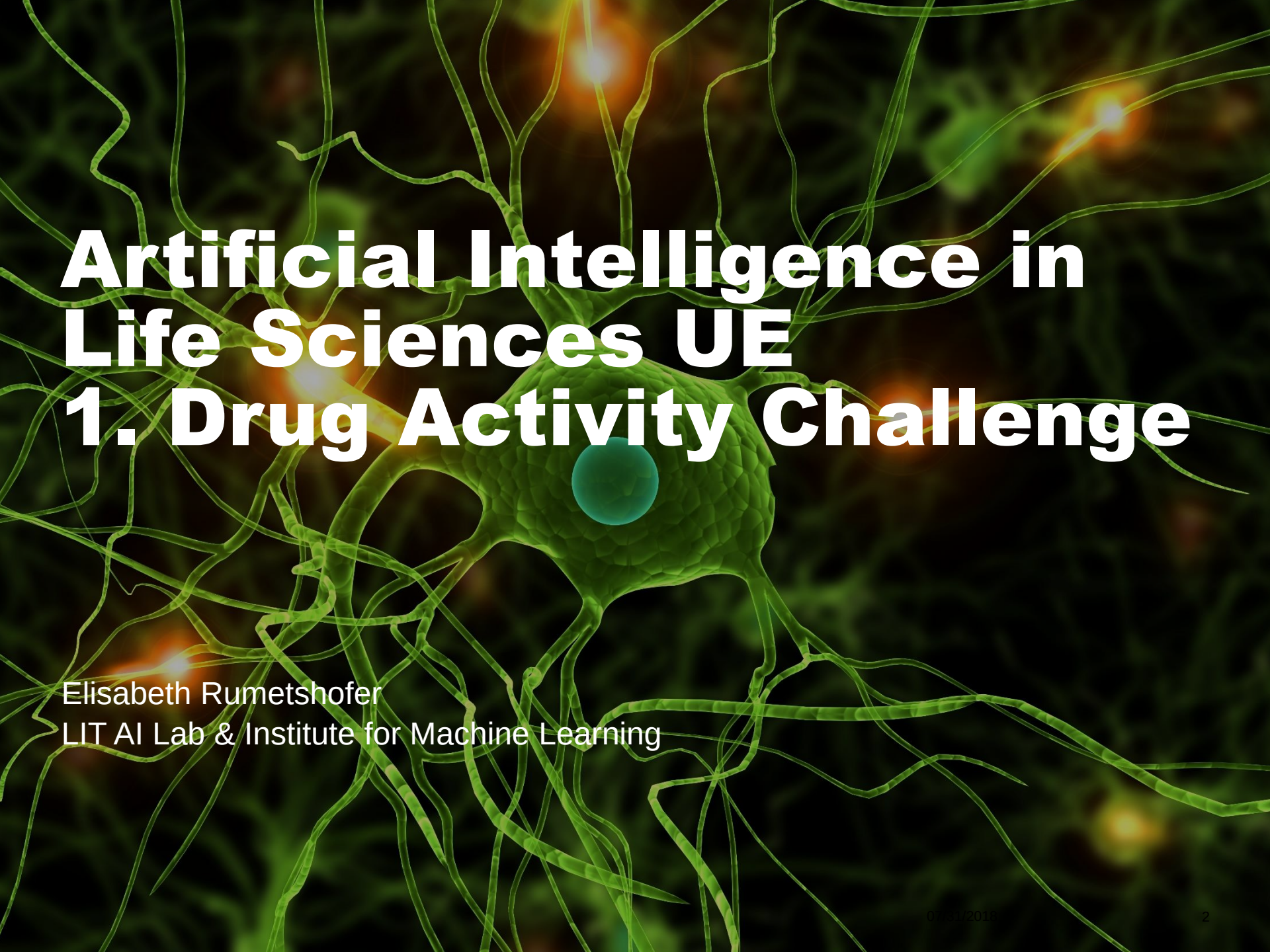


JOHANNES KEPLER
UNIVERSITY LINZ



Artificial Intelligence in Life Sciences UE
1. Drug Activity Challenge

Elisabeth Rumetshofer
LIT AI Lab & Institute for Machine Learning

Drug Activity Challenge

- Predict activity of unseen compounds for 11 bioassays
- **Multitask:** activity for each bioassay individually
- **Unbalanced labels:** different amount of actives/inactives
- **Sparse labels:** not all compound-assay pairs have measurements
- **Masking:** unknown samples need to be masked out for loss/evaluation

	Assay 1	Assay 2	Assay 3	Assay 4	Assay 5	Assay 6	Assay 7
Compound 1	active	inactive	unknown	inactive	inactive	active	unknown
Compound 2	inactive	unknown	unknown	unknown	inactive	inactive	unknown
Compound 3	unknown	unknown	inactive	unknown	Inactive	unknown	active

Data

- Training set
  - Compounds as SMILES strings
  - Activity value per compound-assay pair
  - Activity values: 1 (active), 0 (unknown), -1 (inactive)

- Test set
  - Compounds as SMILES strings

- Additional files
  - Sample submission to show required format
    - Values between 0 and 1 for every compound-assay pair
  - Server evaluation script for main metric

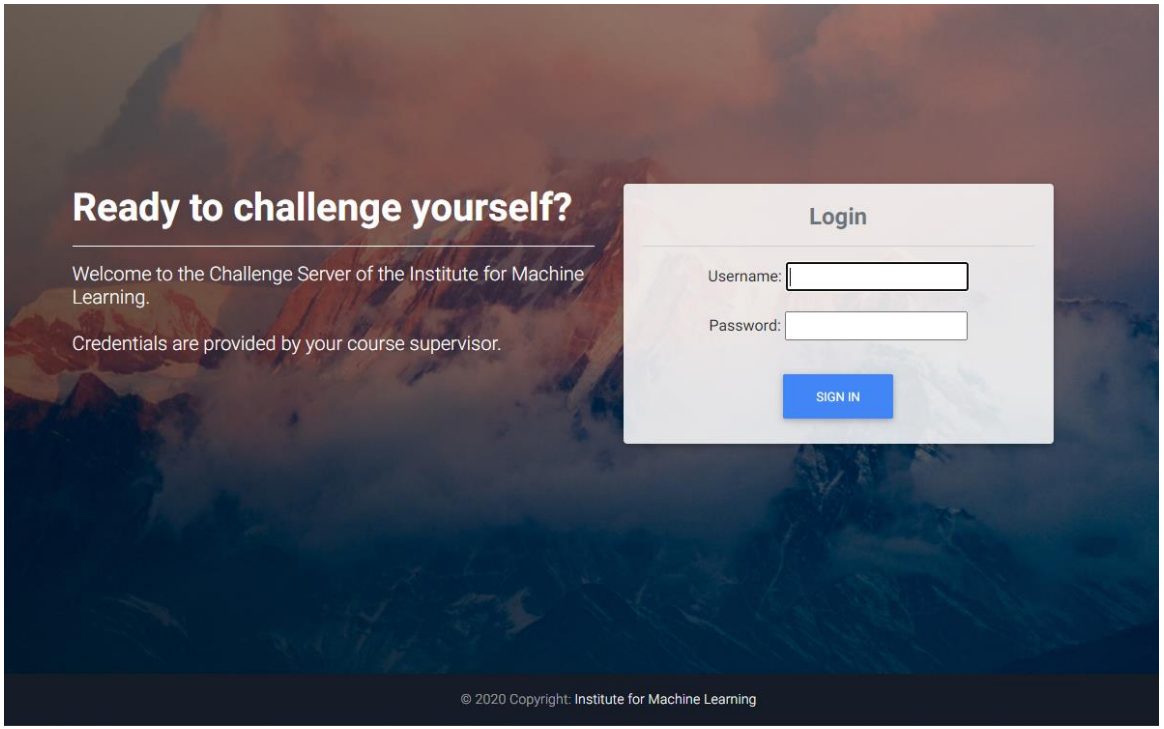
Assessment

- Main metric: **mean area under ROC curve (AUC)**
  https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html
- AUC per task
  - for better insights into model performance (not relevant for grading)
- Max. 10 submissions

- Public LB results
  - perfomance indicator
  - best model of public LB will be evaluated for private LB
- Private LB results
  - will be calculated after challenge is closed
  - determine your final challenge performance

Challenge Server

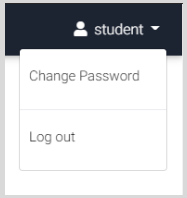
- https://apps.ml.jku.at/challenge



User and password

- K + 8 digit student ID number
- e.g. `K01234567`
- If you already have an account: your PW did not change

Please **change** your **password** when you login the first time!



Final Remarks

- Challenge start: 19th March, 12:00
- Deadline for predictions (challenge server): Wed. **28 April, 12:00**
- Deadline for report/code (Moodle): Thu. **29 April, 12:00**
- Predictions, report and code are **required for grading**

- Max. 1 page report
  - concise description of your experiments
    - pre-processing/input data, tried models/hyperparameters, final approach, post-processing,…
  - add following info about your code:
    - if you don’t use a notebook, state path to your main file
    - you can use code from the internet, but state fundamental code sources (for plagiarism reasons)