AI research in practice

Sonja Aits

20250924

Managing an AI project

Take inspiration from Agile project management

- Actively involve stakeholders from the start
- Interact with other team members
- Regularly reflect and recalibrate work processes
- Start small and then improve in small increments
- Welcome change

Keep track of important things

- Write a brief project plan and update it often
- Document data science experiments as thoroughly as laboratory experiments
- Store metadata, documentation and other relevant info together with models, data and code
- Use version control systems
- Use GitHub or similar tools to keep track of open/solved issues
- Standardize naming and file locations – Model1 is NOT a good model name

Invest time in automating recurrent tasks

imageanalysis / Malou_dir / Results / Model-Evaluations / **model1-evaluation.ipynb**

Preview | Code | Blame | 958 lines (958 loc) · 311 KB | Raw

Evaluation of performance

In [ ]:

In [1]:

```
import os

import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sb

import skimage.io
import skimage.morphology
import skimage.segmentation

import utils.evaluation
import utils.metrics
from config import config_vars
```

```
Using TensorFlow backend.
```

Configuration

In [2]:

```
# Partition of the data to make predictions (test or validation)
config_vars['path_files_training'] = '/home/maloua/Malou_Master/5_Models/2_Final_Models/data/4_filelists/1-2_training.txt'
config_vars['path_files_validation'] ='/home/maloua/Malou_Master/5_Models/2_Final_Models/data/4_filelists/VALIDATION.txt'
config_vars['path_files_test'] = '/home/maloua/Malou_Master/5_Models/2_Final_Models/data/4_filelists/TEST.txt'
```

Make use of other people's work…

- Data
- Models
- Code
- Visualizations
- Workflows
- Notebooks

- But always give credit!

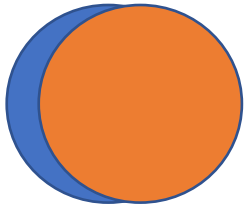
Share your own work

- Release models in model zoos
- Release code/notebooks with articles and on platforms like GitHub
- Release data in repositories – get DOIs
- Write articles about the methodology/data
- Share training material

- Don’t forget a license!
- Use FAIR principles: Findable, Accessible, Interoperable, and Reusable

Remember: Success criteria depend on the goal – don’t loose yourself in pointless optimization

Overlap	0.5	0.6	0.7	0.8	0.9	0.95
Model 1	92.1	91.0	87.5	84.7	68.8	13.6
Model 2	89.7	88.5	86.7	81.3	65.2	20.9

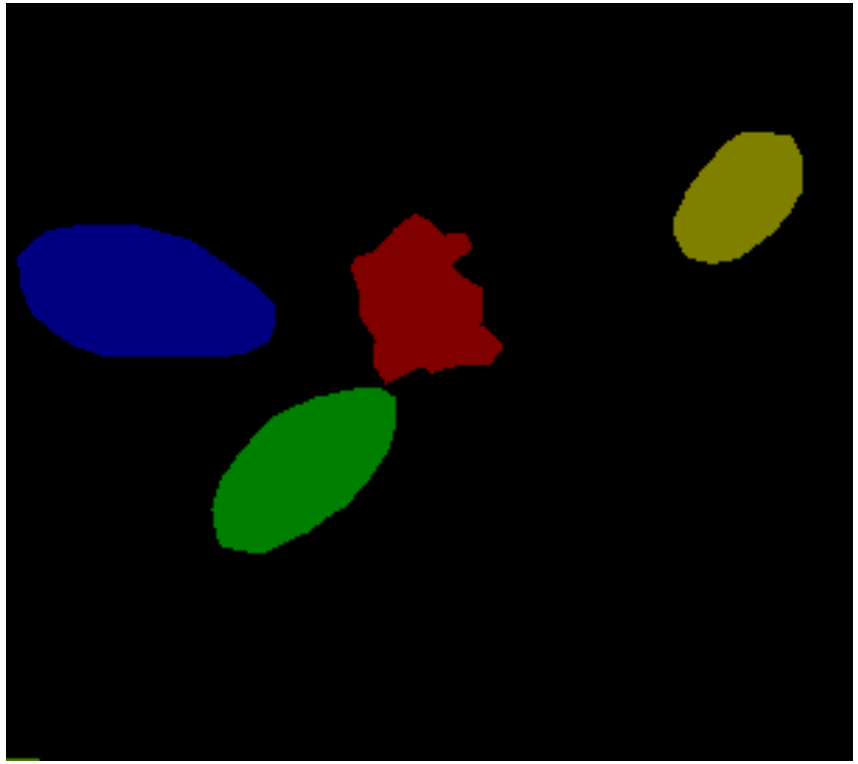


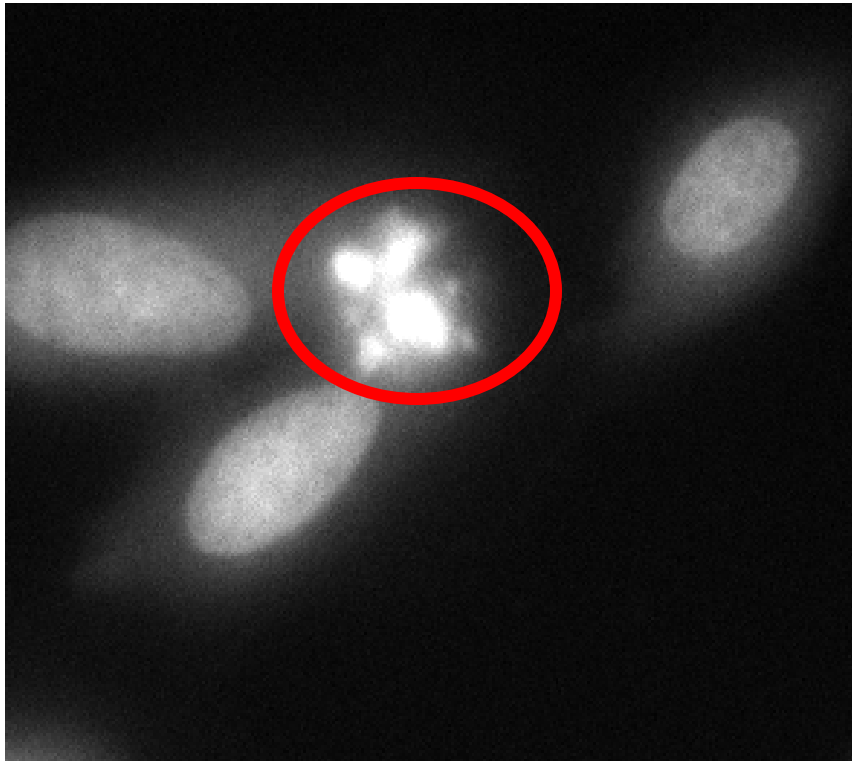
Common problems

Every AI project has problems!

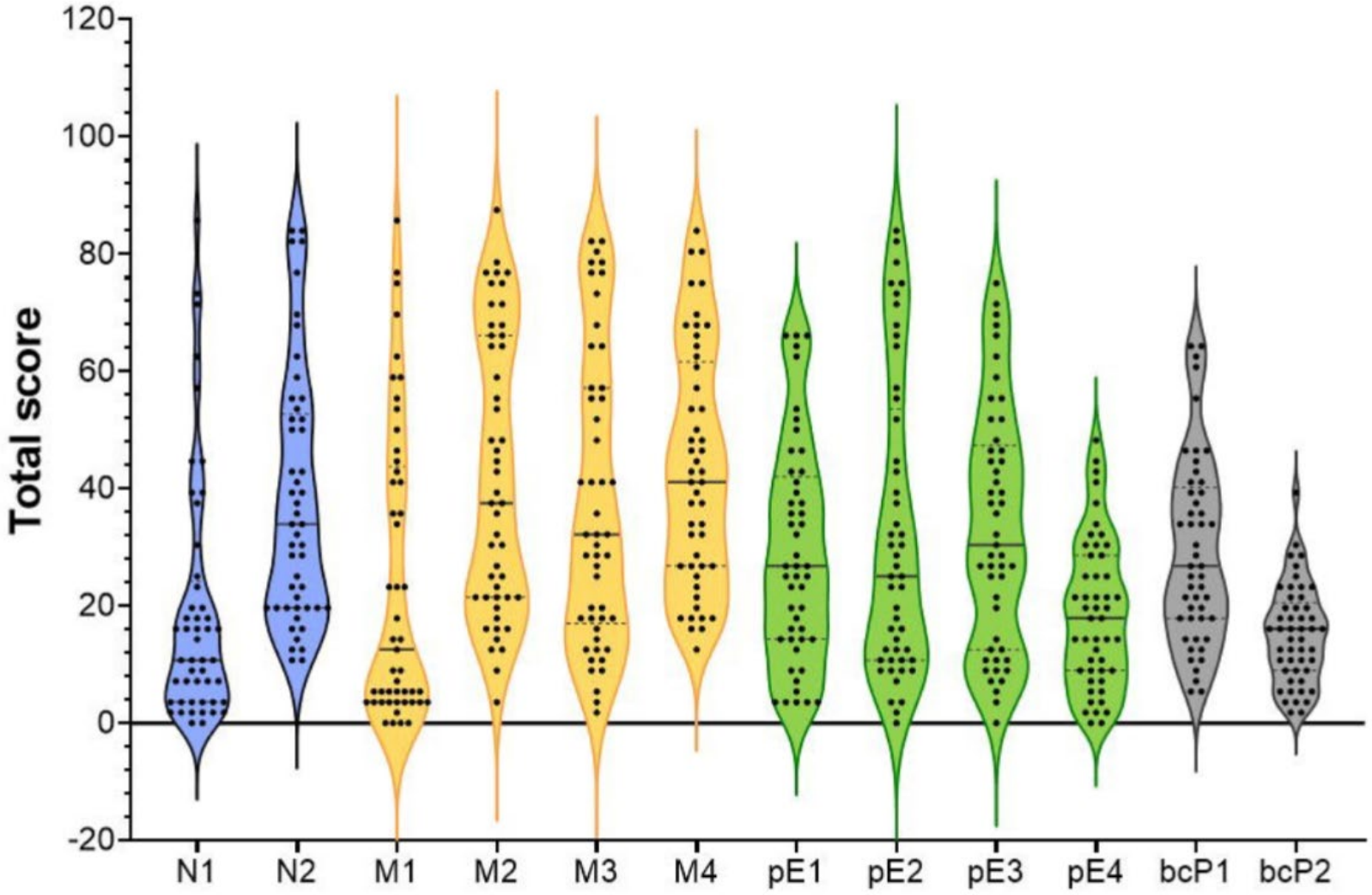
- Problems with data access and handling
- Limited dataset size
- High dimensionality
- Biased data
- Poor annotation quality

Annotations are subjective

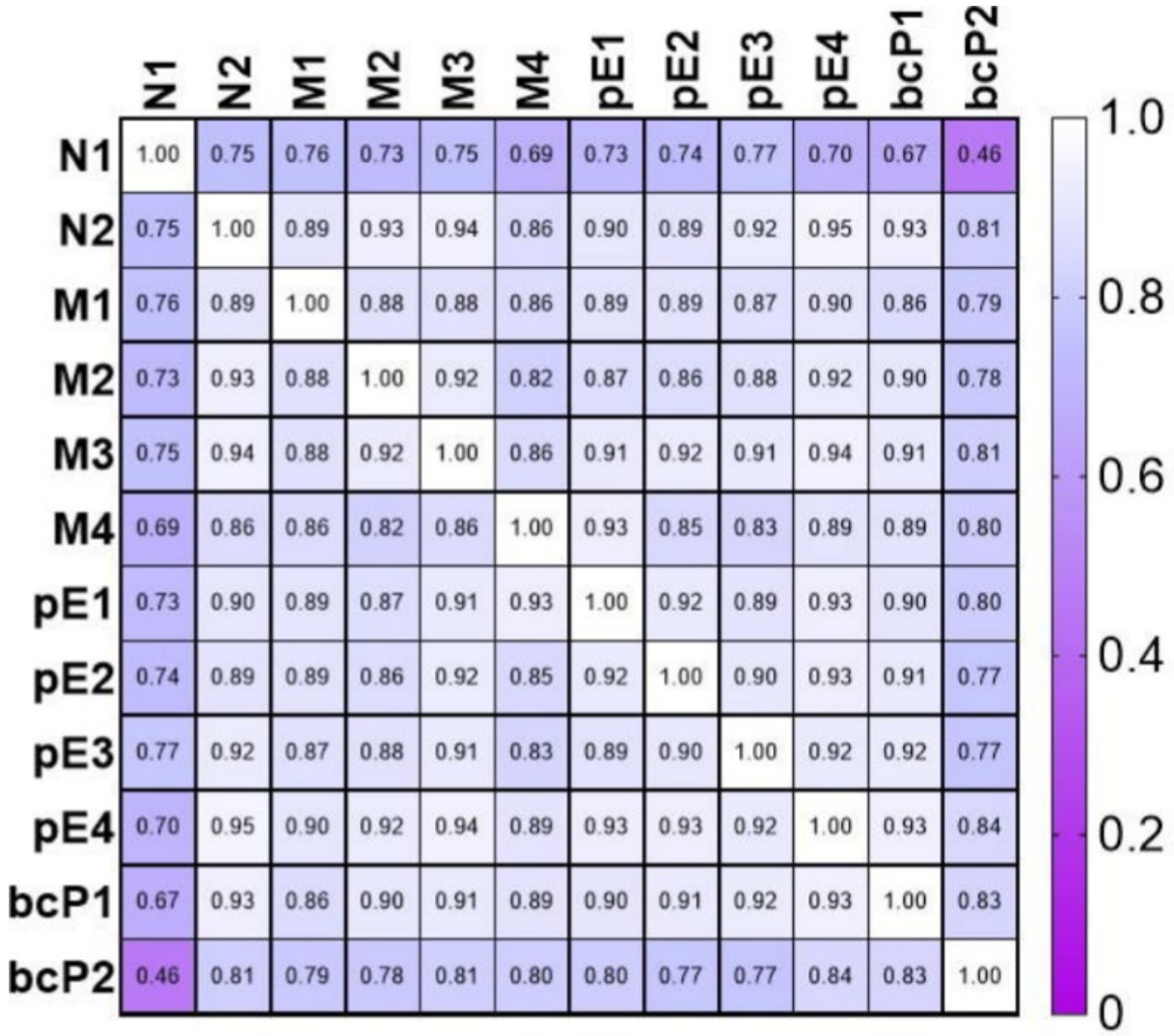




Annotations can be unreliable



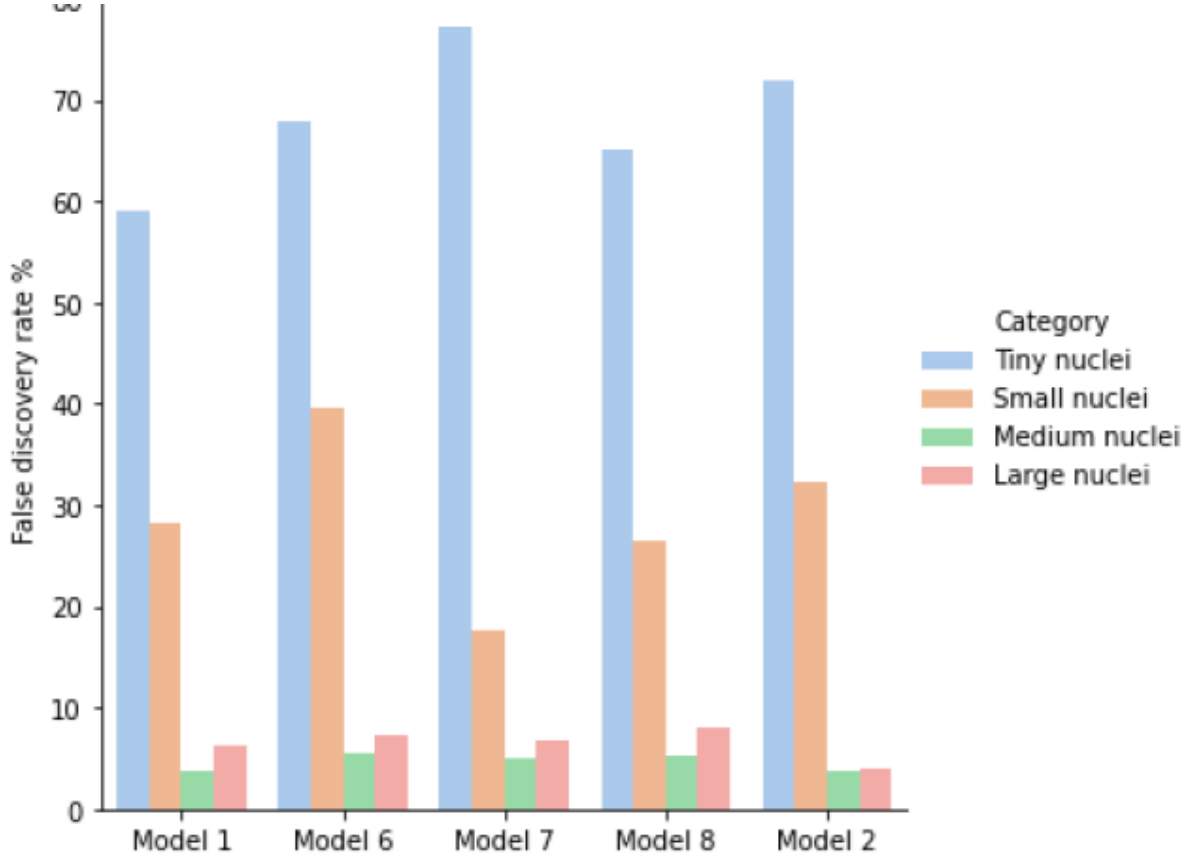
Annotations can be unreliable



Every AI project has problems!

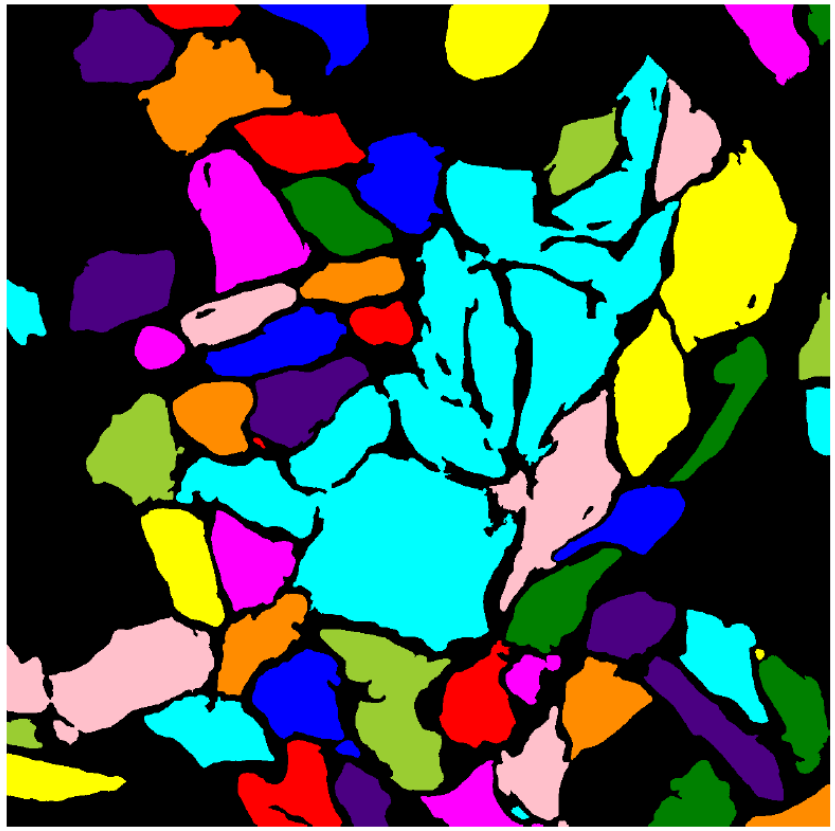
- Problems with data access and handling
- Limited dataset size
- High dimensionality
- Biased data
- Poor annotation quality
- Missing values
- Unsuccessful model training
- The “Black Box” issue
- Deployment obstacles
- Interdisciplinary barriers

Look beyond the primary metrics!



→ Guidance for model improvement

Look beyond the primary metrics!



→ Some problems may be solved with postprocessing

Next steps

Connect with others

- Stay in touch with course mates
- Go to events
  - AI Lund
  - COMPUTE
  - LU/LTH profile areas
- Reach out to colleagues at LU and elsewhere

What to learn next

- Python - Pandas and Matplotlib for EDA
- ML/DL library (PyTorch, HuggingFace, Scikit-Learn)
- Version control with Git
- Infrastucture: NAISS and Colab
- Data/project management
  - Documentation with Jupyter notebooks
- Explainable AI

- Small projects in your domain of interest