HELMHOLTZ AI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

Data Science in Earth Observation

AI4EO Hackathon - Regression

2022

Andrés Camero

Andres.CameroUnzueta@dlr.de

Helmholtz AI Consultant Team Leader

Department EO Data Science, DLR

DLR

Agenda

1. Helmholtz AI CountMeIn
2. The data
3. Hands-on
4. Open problems

Helmholtz AI CountMeIn

GOGREEN

USE AS LITTLE ENERGY AS POSSIBLE!

Given the CountMeIn problem and a maximum RMSE of 1111, participants should propose a solution that minimizes the environmental impact (while it achieves the target performance). Participants in this track need to use the HAICORE resources at Karlsruhe Institute of Technology (KIT) as the Computer HoreKa features a very accurate power measurement facilities. The impact will be computed using HAICORE logs, and your submission will include the Job ID.

GOFAST

USE AS LITTLE TIME AS POSSIBLE!

Given the CountMeIn problem and a RMSE of 1111, participants should propose a solution that minimizes the training and prediction time. Participants in this track can choose to use the HAICORE resources at KIT and the HAICORE resources at Forschungszentrum Jülich (FZJ). At KIT, a maximum of 56 A100 GPUs can be used, while at JUWELS Booster at FZJ all 3744 A100 GPUs are available. For the submission, the compute center, as well as the Job ID must be included.

HELMHOLTZAI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

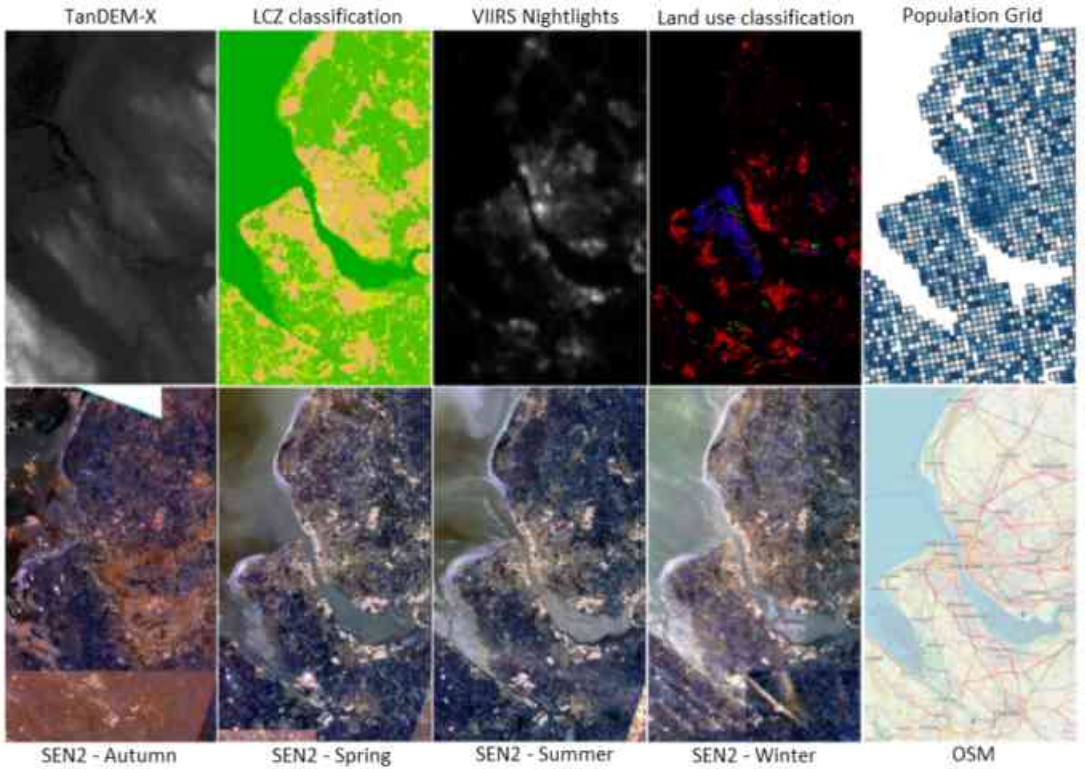
DLR JÜLICH Forschungszentrum KIT Karlsruher Institut für Technologie

DLR

So2Sat POP

98

European cities



6

Data sources

Doda, S., Wang, Y., Kahl, M., Hoffmann, E.J., Taubenböck, H. and Zhu, X.X., 2022. So2Sat POP--A Curated Benchmark Data Set for Population Estimation from Space on a Continental Scale. arXiv preprint arXiv:2204.08524.

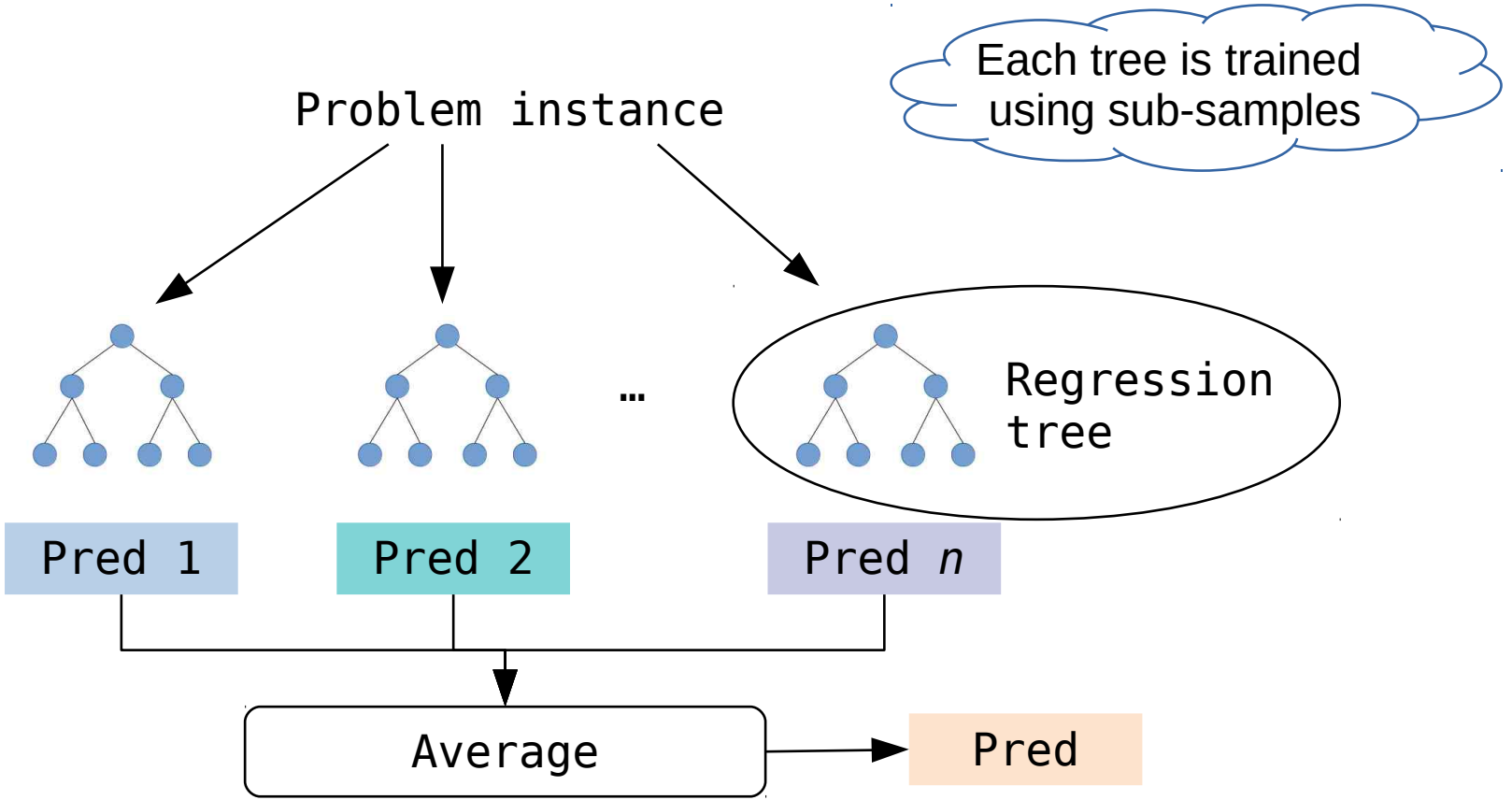
https://github.com/acamero/data-science-eo-regression

Machine learning

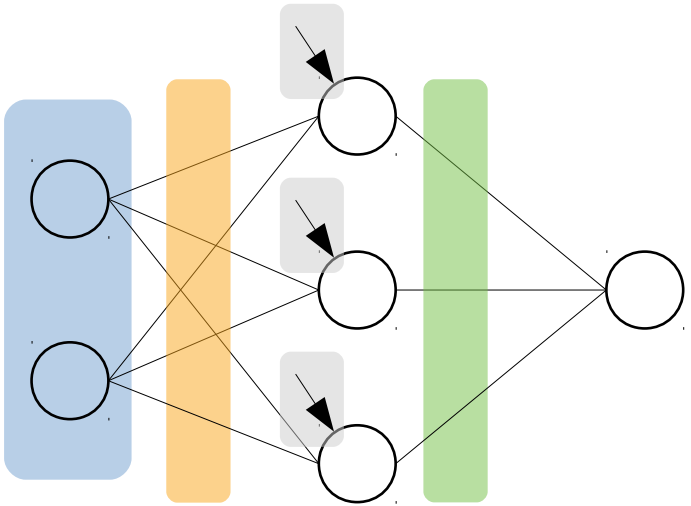
A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

Tom M. Michell, 1997

Random forest regressor



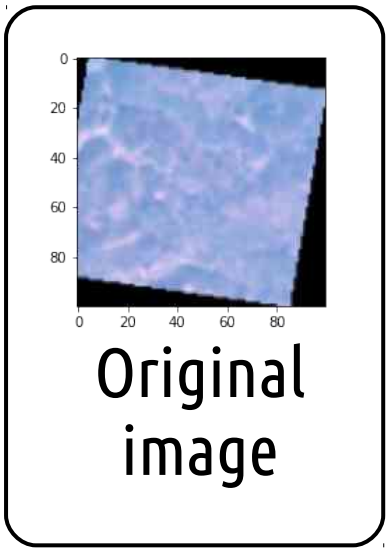
Artificial neural networks

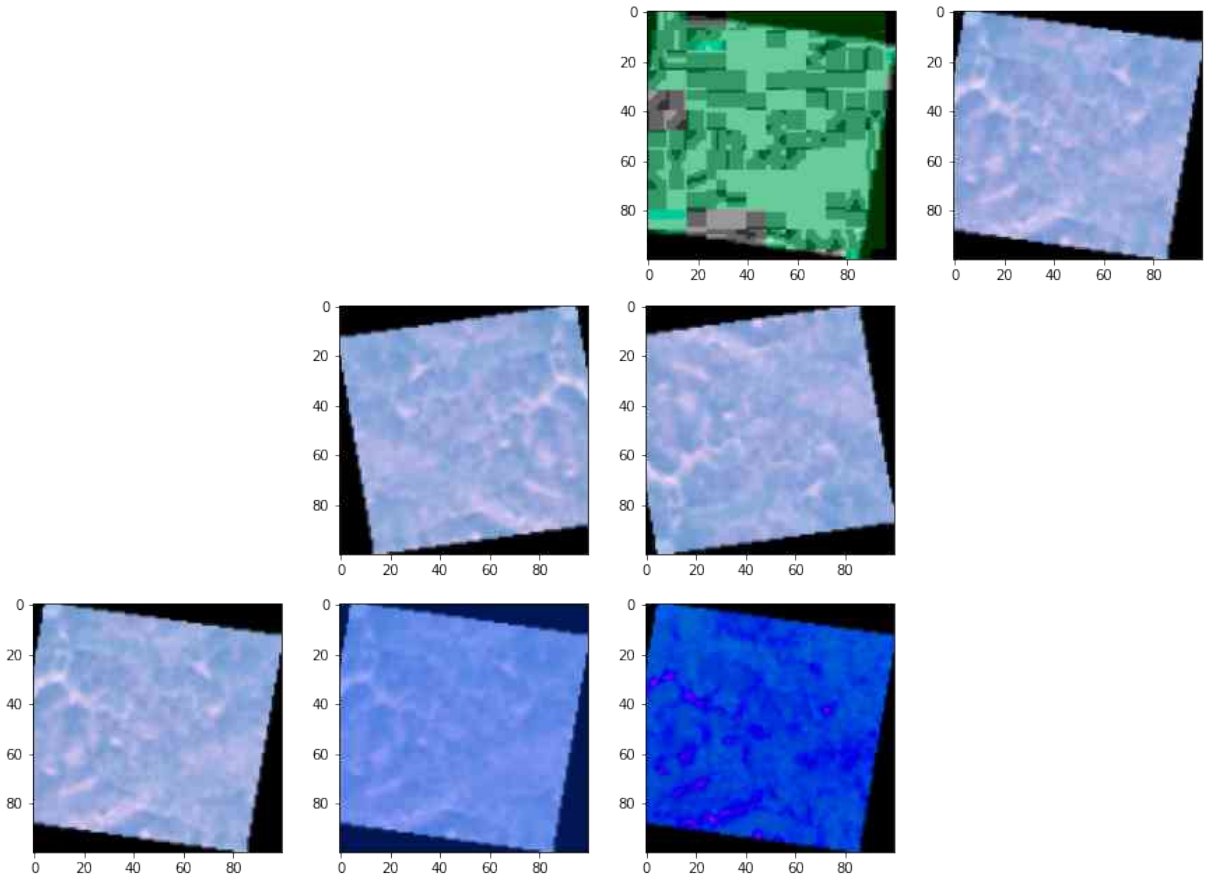


$$\hat{y} = f\left(W \cdot x + b\right)$$

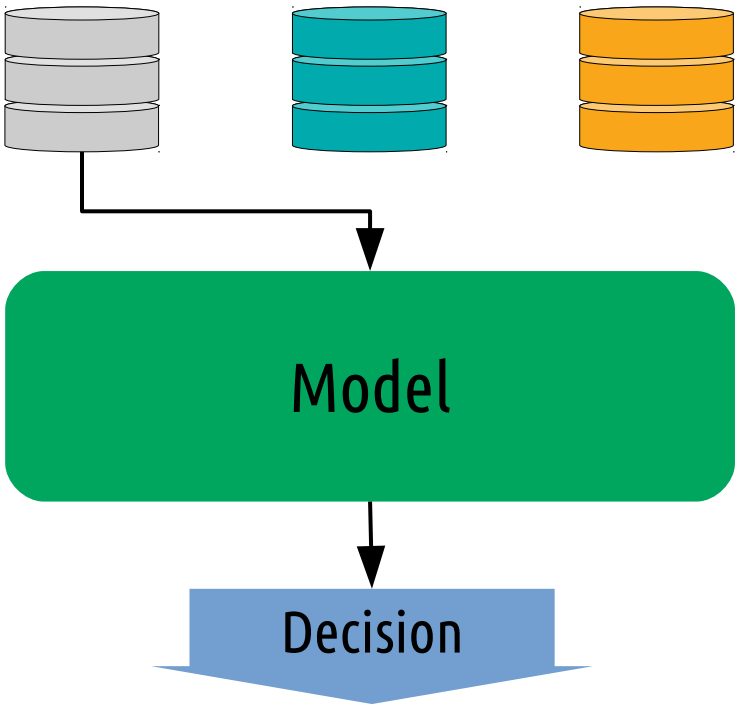
activation function

Data augmentation

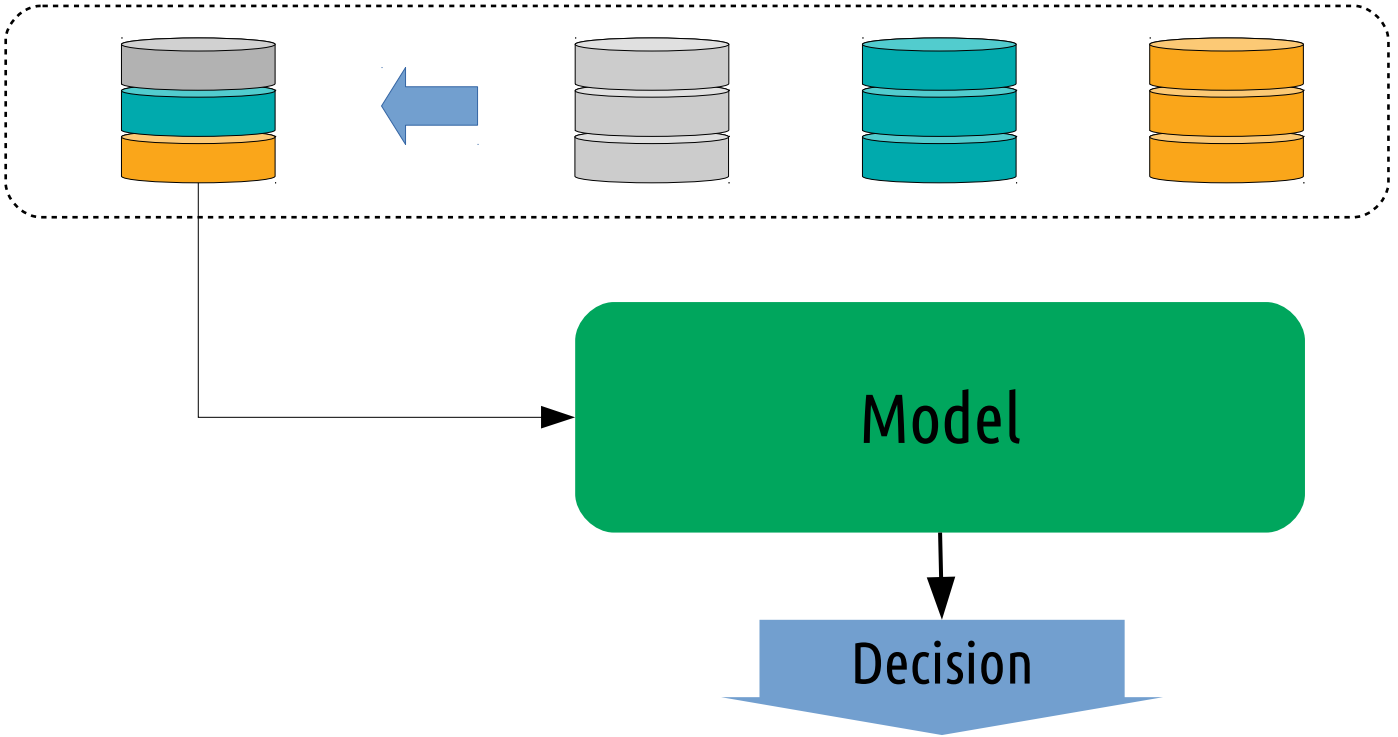




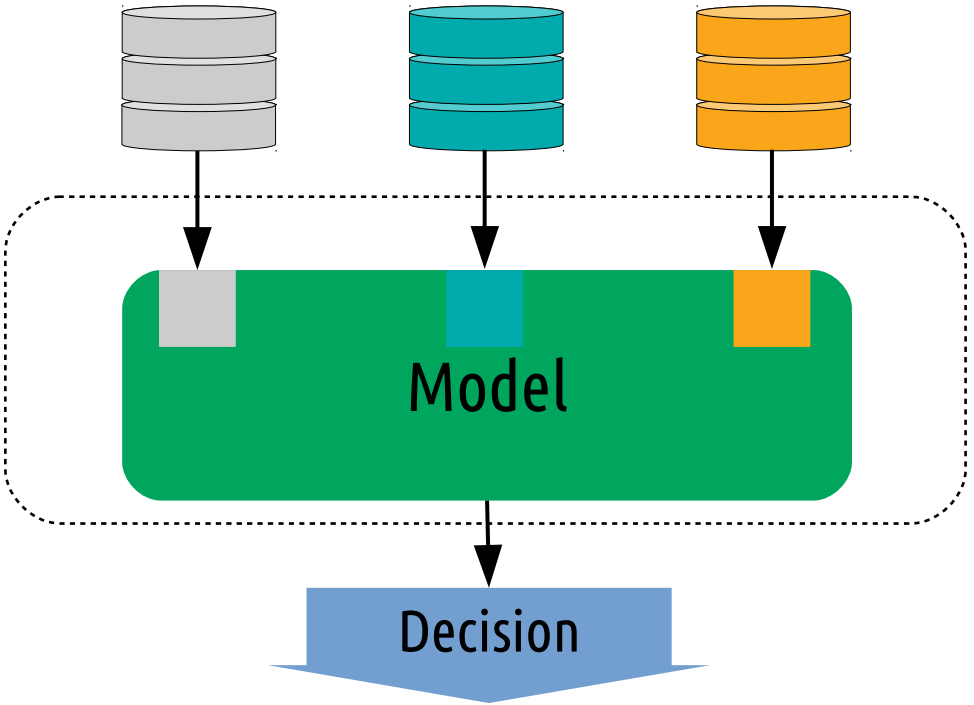
Data fusion



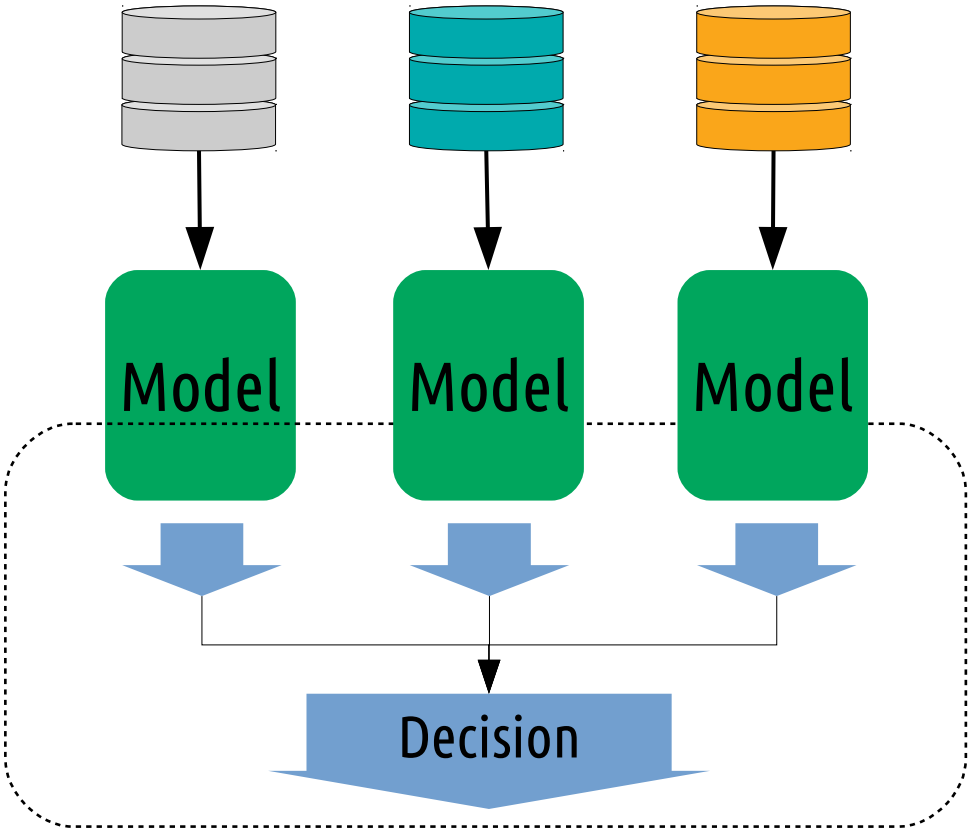
Data fusion



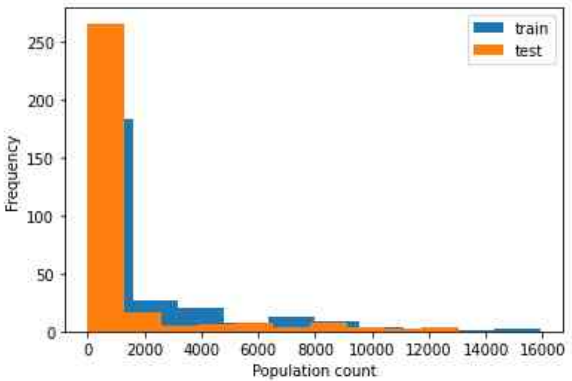
Data fusion



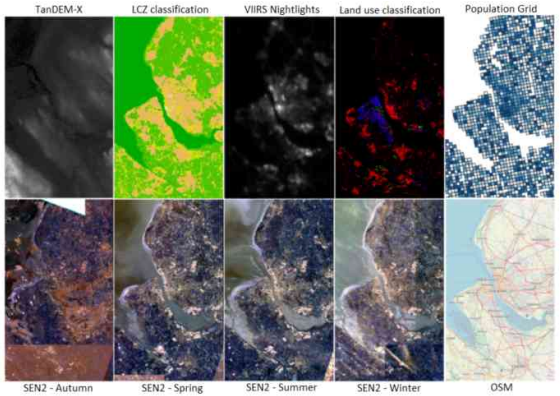
Data fusion



So far…



Highly imbalanced data



Multi-modal data

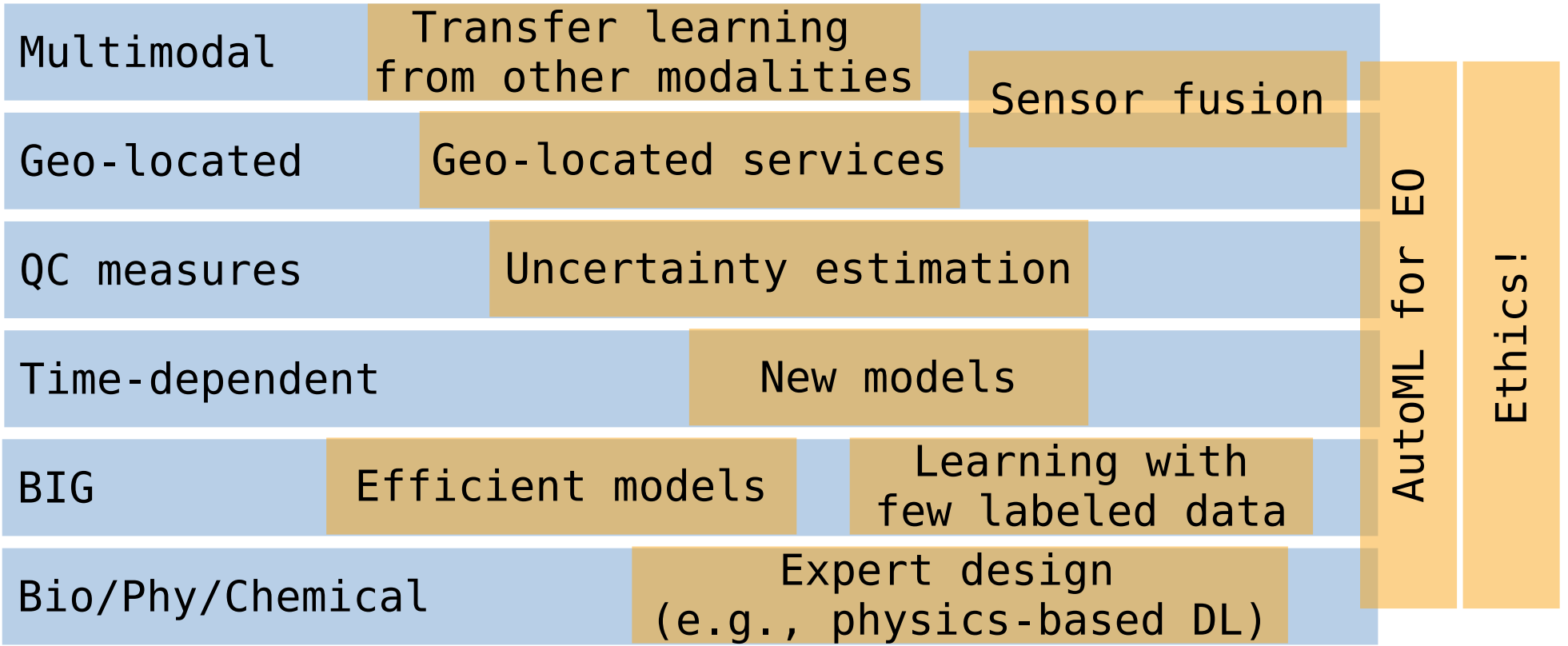
Geographic diversity

DLR

Remote sensing data is…

- Multimodal
- Geo-located
- QC measures
- Time-dependent
- BIG
- Bio/Phy/Chemical

Thus, requires…



References

1. Doda, S., Wang, Y., Kahl, M., Hoffmann, E.J., Taubenböck, H. and Zhu, X.X., 2022. So2Sat POP--A Curated Benchmark Data Set for Population Estimation from Space on a Continental Scale. arXiv preprint arXiv:2204.08524.
2. LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), pp.436-444.
3. McCarthy, J., 2007. What is artificial intelligence?
4. Mohri, M., Rostamizadeh, A. and Talwalkar, A., 2018. Foundations of machine learning. MIT press.
5. Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F. and Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), pp.8-36.