



Mobile Computing Lab

Q&A

Olga Saukh
Institute of Technical Informatics, TU Graz
Complexity Science Hub Vienna

saukh@tugraz.at

19.05.2021

Outline

- **On-demand individual coaching (Q&A)**
- 2021-05-26 -- On-demand individual coaching (Q&A)
- 2021-06-02 – Progress review (Option 1 and 2)

Where We Are?

Option 1: Activity Monitoring

- **Activity Monitoring + Transfer Learning**

1. Activity Monitoring:
accelerometer + kNN (or other)
2. **Transfer Learning (WS4)**

- **Evaluation criteria:**

- Performance and creativity

Demos: 28.04.2021

Progress Review: 02.06.2021

Final demos: 23.06.2021

Option 2: Free Choice

- **Your own project** (requires approval)

1. Pitch your idea
2. **Implementation**

- **Evaluation criteria:**

- Technical depth and creativity

Progress Review: 05.05.2021

Progress Review: 02.06.2021

Final demos: 23.06.2021

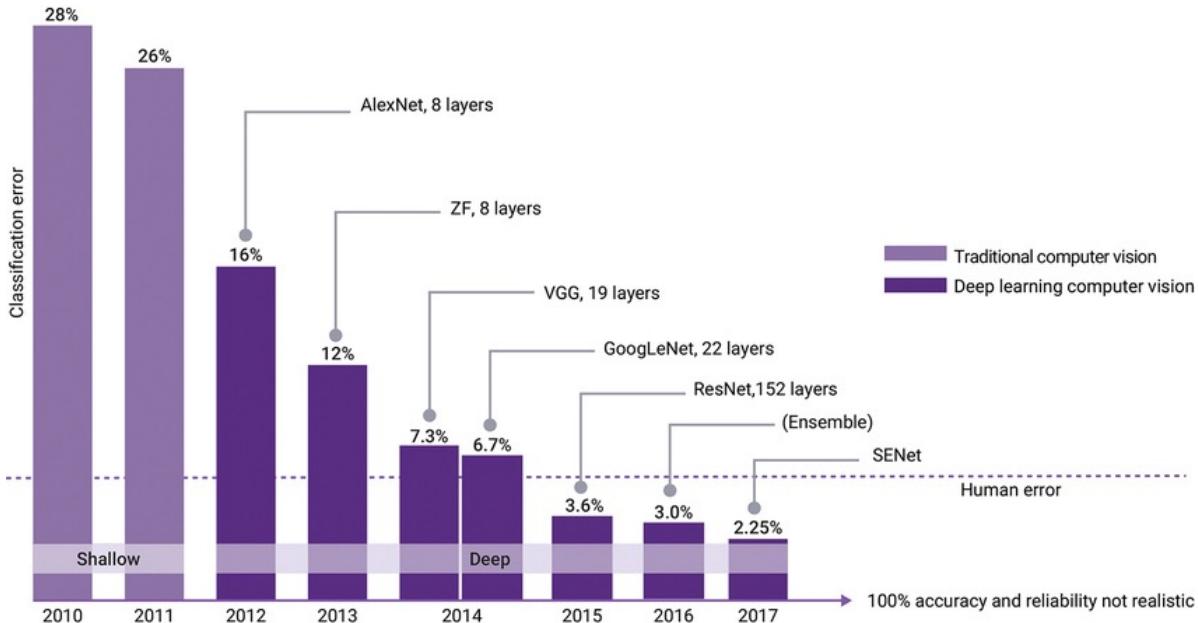
What is Performance?

What is Performance?

- **Many dimensions ...**
 - Accuracy
 - Robustness
 - Reliability
 - Speed / responsiveness
 - Energy consumption

- **All above is correlated with**
 - How much relevant experience do you have?
 - How much time did you spend on testing and improving your app?

Explosive Progress in Machine Learning



ImageNet Large Scale Visual Recognition Challenge results show that deep learning is surpassing human levels of accuracy

<https://www.synopsys.com>

Talk by L. Schmidt, 2021

V. Shankar et al., Evaluating Machine Accuracy on ImageNet, ICML 2020

Good dataset

ImageNet: 1000 classes,
1.2M examples

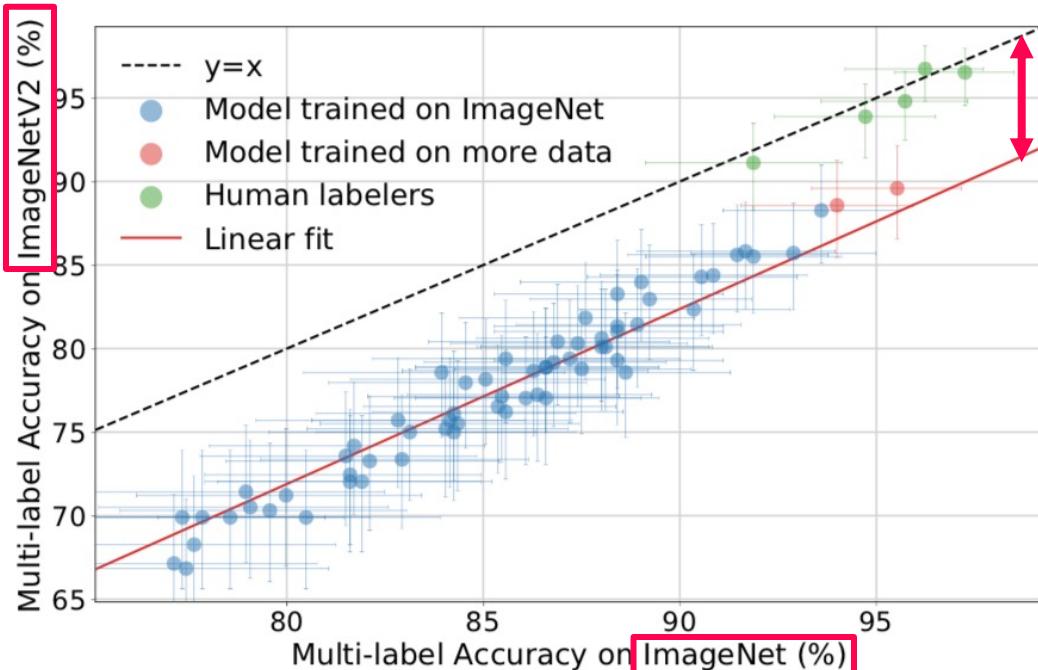


Human error

What does good performance on a test set actually imply?

Generalization

- The classifier should perform similarly well on a new data from the **same source**



- No overfitting due to the test set reuse. Progress is real!
- The best models on the original test set stay the best on the new test set.
- All models see a substantial drop in accuracy
- Human vs machine comparison are unstable

Data distribution shift!

Generalization: Pollen Sensors

Human expert: ~80% accuracy

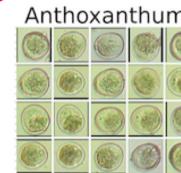
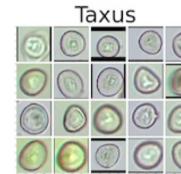
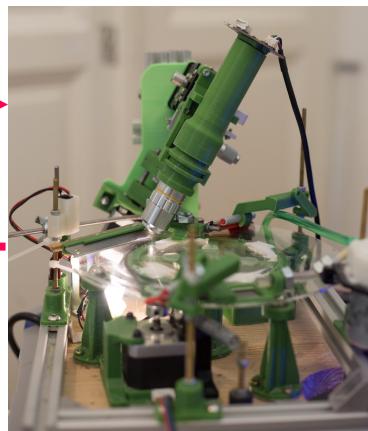
- At least the classifier should perform similarly well on a new data from the **same source**



Library: 16 pollen types



Real world



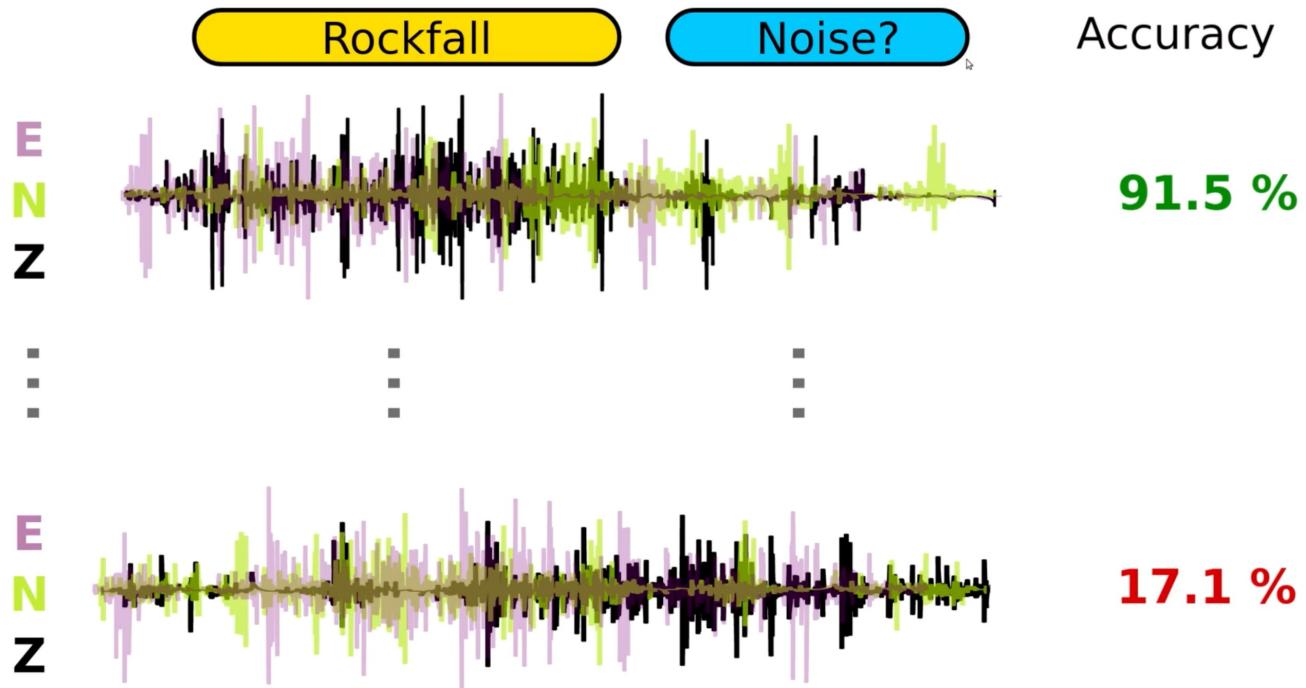
90-95%
classification
accuracy



2-20%
classification

N. Cao et al., Automated Pollen Detection with an Affordable Technology, EWSN 2020
N. Cao et al., Dataset: Pollen Video Library for Benchmarking Detection, Classification, Tracking and Novelty Detection Tasks, DATA 2020

Generalization: Seismometers



M. Meyer et al., Using system context information to complement weakly labelled data, WeaSuL 2021

Datasets as a Research Area

- Data augmentation during training is a good idea
- Evaluating model performance on multiple test sets is a good idea
- Direct machine vs human comparison leads to overoptimistic results
- Achieving truly reliable machine learning requires a deeper understanding of the input changes a model should be robust to
- Robustness to small, naturally occurring distribution shifts is a performance dimension not addressed by current benchmarks, occurs in the real world and easily handled by humans

Questions?