



CEITEC

Central European Institute of Technology
BRNO | CZECH REPUBLIC

MUNI

Genomic Benchmarks:

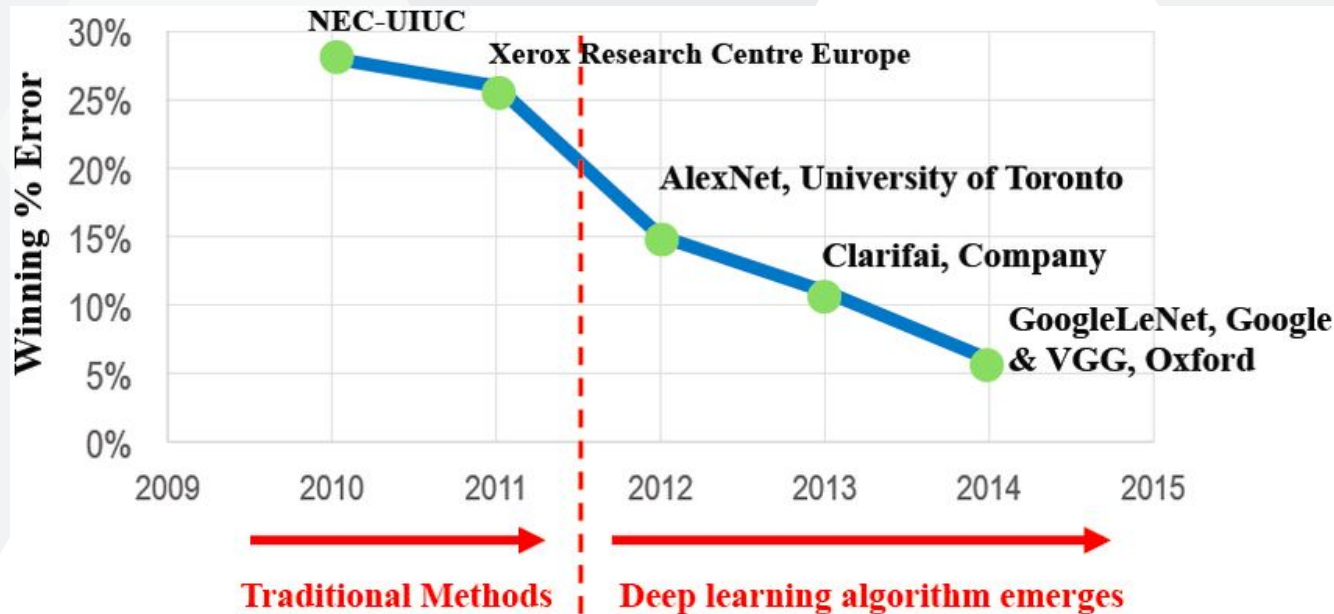
A Collection of Datasets for Genomic Sequence Classification

Katarina Gresova, Vlastimil Martinek, David
Cechak, Petr Simecek, Panagiotis Alexiou

Benchmarks as drivers of innovation

- Benchmark = a collection of transparent and well-curated datasets in a given field that are recognised by a community and used to evaluate methods
- Defining a problem
- Providing an objective measure
- Creating a competition

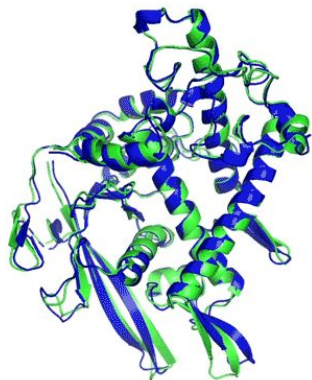
ImageNet as a driver of innovation



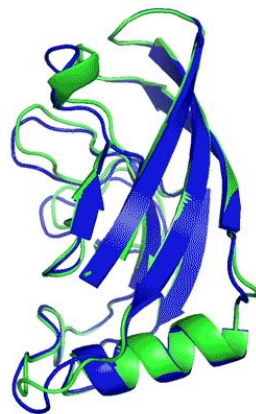
Shawahna, Ahmad, Sadiq M. Sait, and Aiman El-Maleh. "FPGA-based accelerators of deep learning networks for learning and classification: A review." *IEEE Access* 7 (2018): 7823-7859. <https://medium.com/syncedreview/sensetime-trains-imagenet-alexnet-in-record-1-5-minutes-e944ab049b2c>

CASP as a driver of innovation

C
A
S
P
14



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)



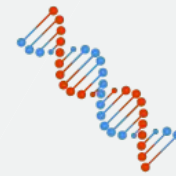
T1049 / 6y4f
93.3 GDT
(adhesin tip)

- Experimental result
- Computational prediction

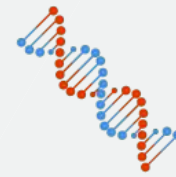
<https://predictioncenter.org/casp14/index.cgi>
<https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>

Genomic Benchmarks as a driver of innovation

- Big goal: Genome annotation
- Defined task: Functional elements classification
- Starting point: 8 datasets + CNN
- Values: accessible, reproducible and comparable



Genomic Benchmarks as a driver of innovation



And now it's your turn!

Acknowledgement



Katarína Grešová



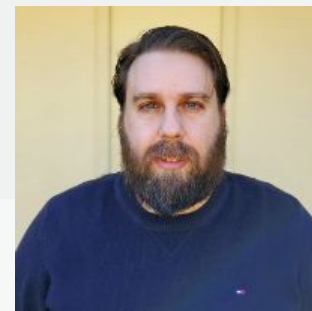
Vlastimil Martinek



David Čechák



Petr Šimeček



Panagiotis Alexiou



Thank you for your Attention!
katarina.gresova@ceitec.muni.cz

<https://predictioncenter.org/casp14/index.cgi>
<https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>