

# DeepNOG: fast and accurate protein orthologous group assignment

by Roman Feldbauer, Lukas Gosch, Lukas Lüftinger, Patrick Hyden, Arthur Flexer and Thomas Rattei

Bioinformatics, Volume 36, Issue 22-23, 1 December 2020, Pages 5304–5312

**Group 16: Jeff Gower & Nico Bohlinger**



# Content

- Introduction
- Research question
- Methodological approaches
- Results
- Conclusion

# Introduction

Task

Classification of proteins into orthologous groups

Current Algorithms

HMMER, DIAMOND

Challenge

Massive amounts of data

Problem

Slow inference

Deep Learning

New Algorithm

DeepFam

Challenge

Scaling and restrictions

New Algorithm 2.0

DeepNOG

# Research question

How can current deep learning architectures for the assignment of orthologous groups (precisely DeepFam) be enhanced...

... to scale towards huge datasets

... without restrictions such as a maximum sequence length

... while still keeping good enough performance compared to alignment-based methods?

# Methodological approaches

## Architecture

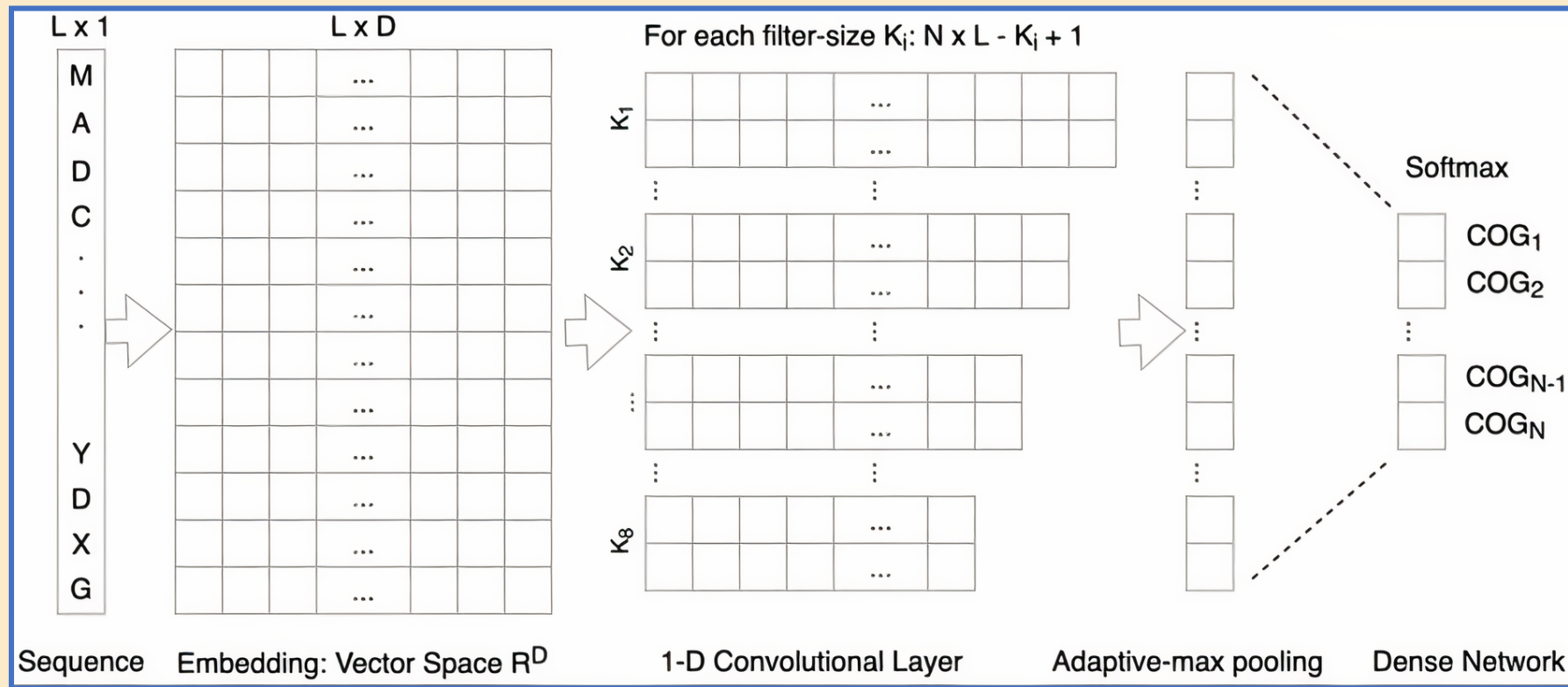


Fig. 1. DeepNOG network architecture

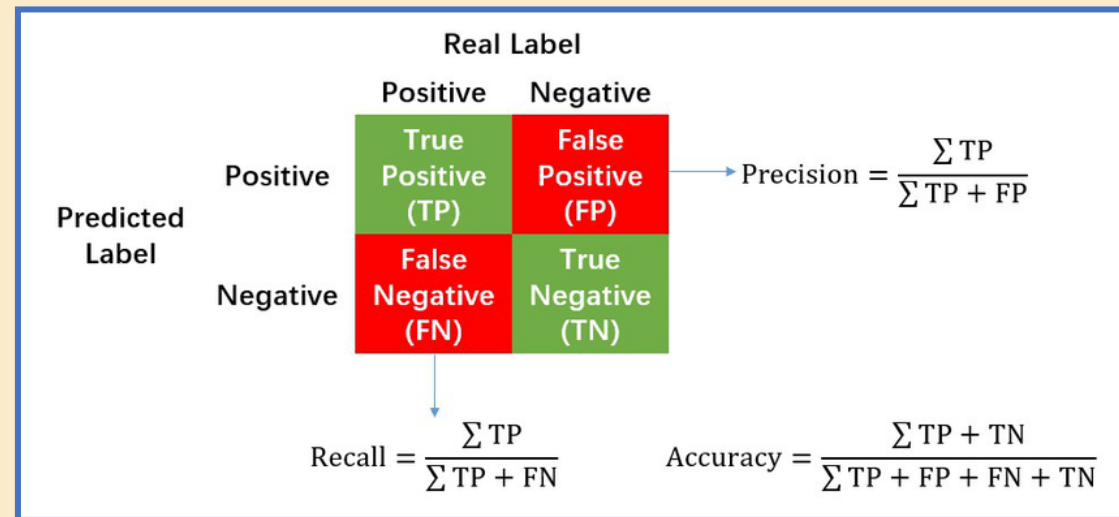
# Methodological approaches

## Databases:

- COG
- eggNOG

## Performance indicators:

- Accuracy, Precision, Recall
- Inference time



<https://www.researchgate.net/publication/336402347/figure/fig3/AS:812472659349505@1570719985505/Calculation-of-Precision-Recall-and-Accuracy-in-the-confusion-matrix.ppm>

# Results

## Accuracy on root level

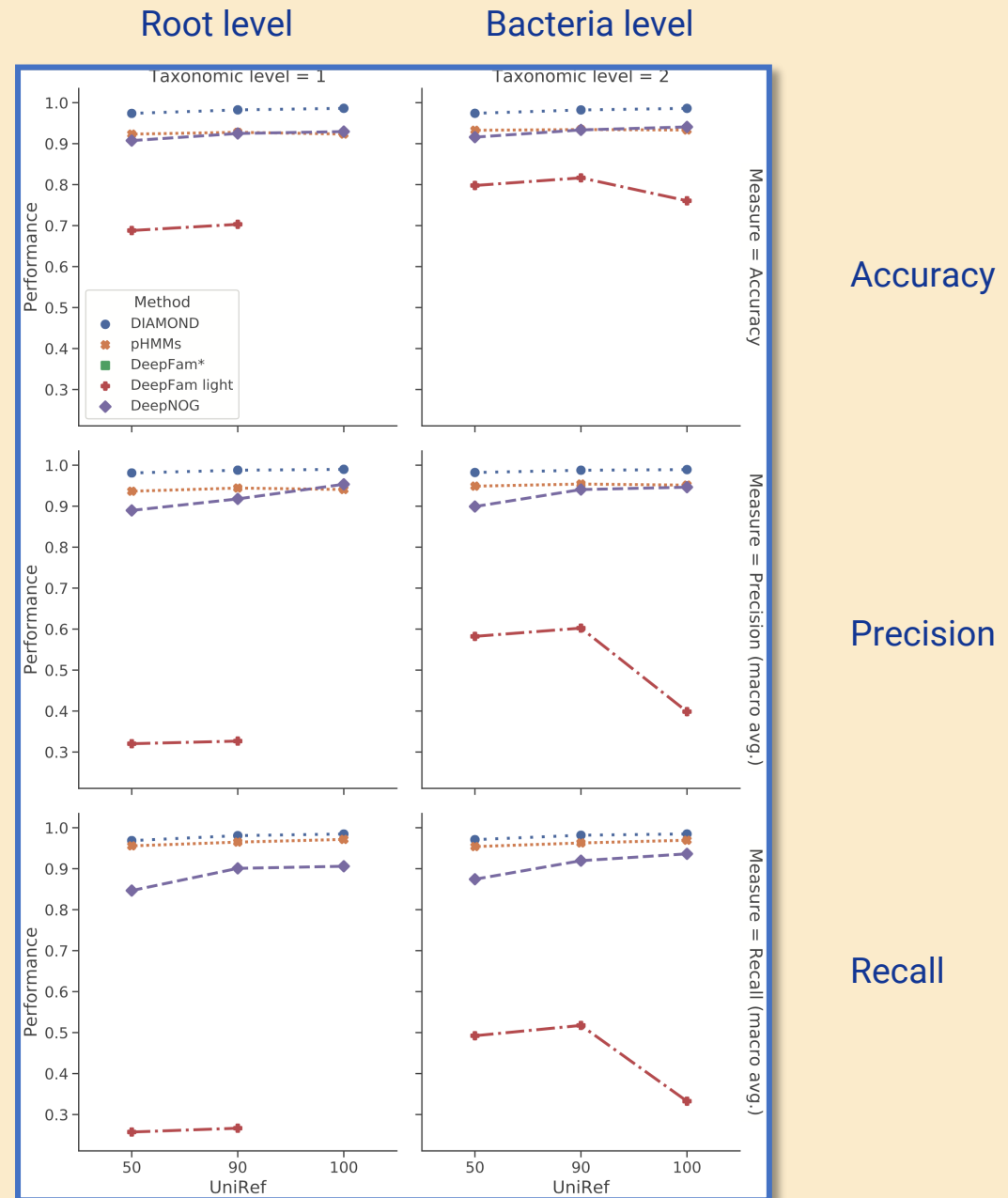
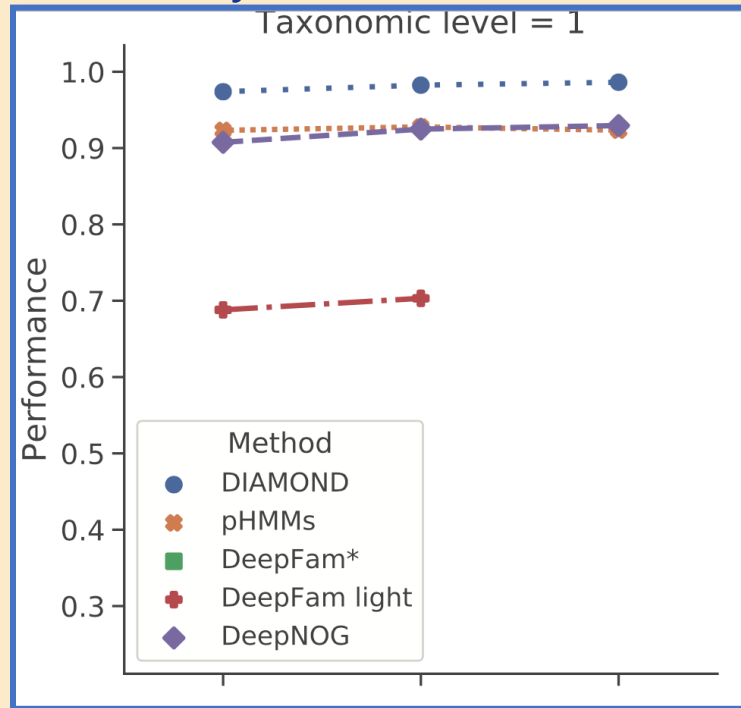


Fig. 4. Assignment accuracy

# Results

**Table 2.** Inference time (seconds/1000 sequences) for COG and eggNOG 5 (bacteria level)

	COG-500	COG-100	NOG <sub>2</sub> <sup>5</sup> -500	NOG <sub>2</sub> <sup>5</sup> -100
DIAMOND	161.7	214.5	781.6	810.0
pHMMs	96.3	207.0	218.9	253.7
DeepFam	49.0	50.2	n/a	n/a
DeepFam light	32.7	35.0	34.9	38.7
DeepNOG (CPU)	<b>24.3</b>	<b>26.0</b>	<b>26.4</b>	<b>28.9</b>
pHMMs (parallel)	4.8	5.1	9.5	14.4
DeepNOG (GPU)	0.6	0.6	0.6	0.6



# Conclusion

How can current deep learning architectures for the assignment of orthologous groups (precisely DeepFam) be enhanced...

... to scale towards huge datasets

→ Scales far better than DeepFam

... without restrictions such as a maximum sequence length

→ More flexible: Variable sequence length, Integration with other homology tools, ...

... while still keeping good enough performance compared to alignment-based methods?

→ Performance is close but not on the same level