Class-Attention on Transmembrane Proteins

Adrian Henkel,

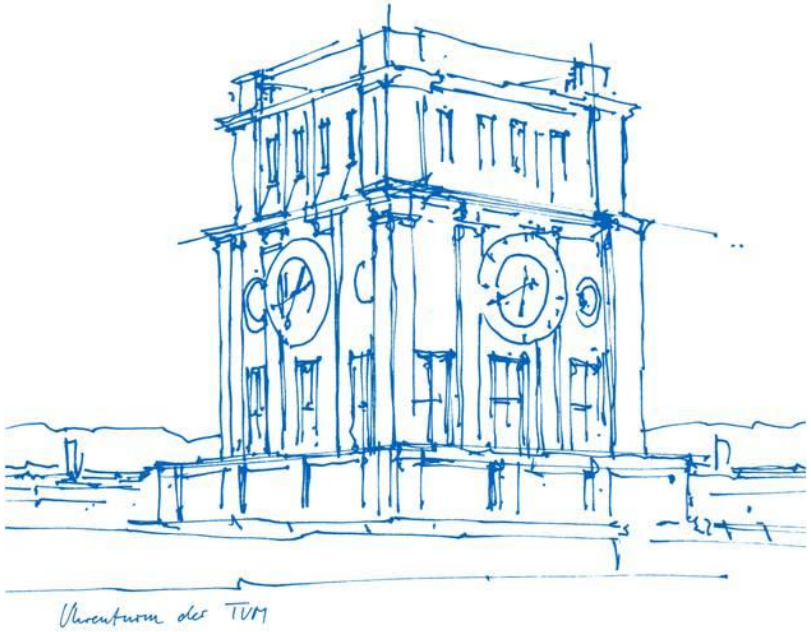
Finn Gaida,

Lis Arend,

Sebastian Dötsch,

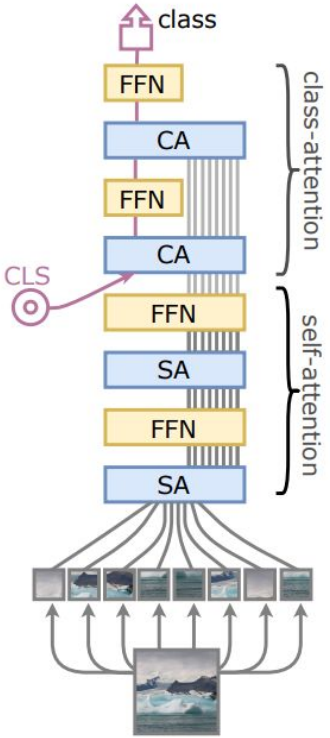
Shlomo Libo Feigin

Munich, 27.01.2022



Recap: Class-Attention

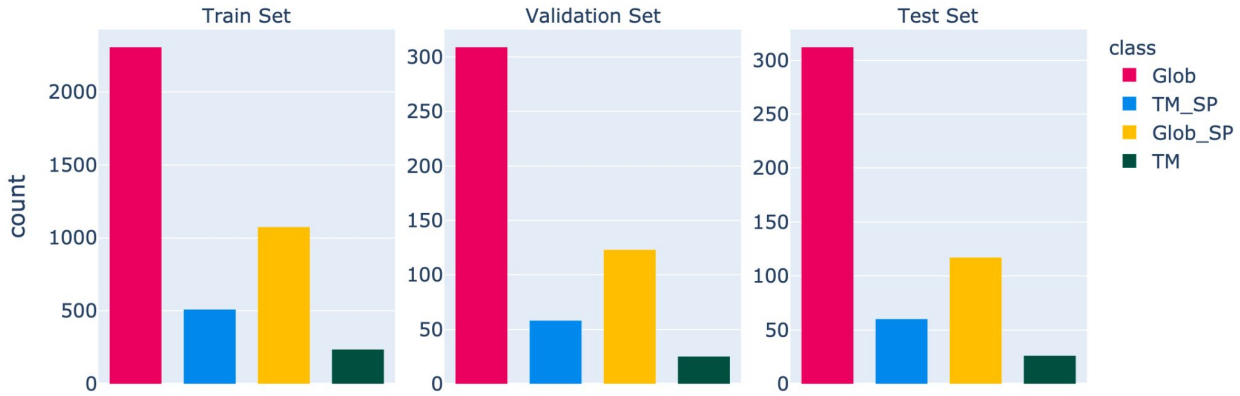
- Two stages to separate two objectives:
  - **Self-attention stage**
    - Input representation is calculated
    - CLS token doesn't appear yet
  - **Class-attention stage**
    - Input representation is fixed
    - Insertion of CLS token
    - → Update only CLS
    - CA identical to standard SA layer



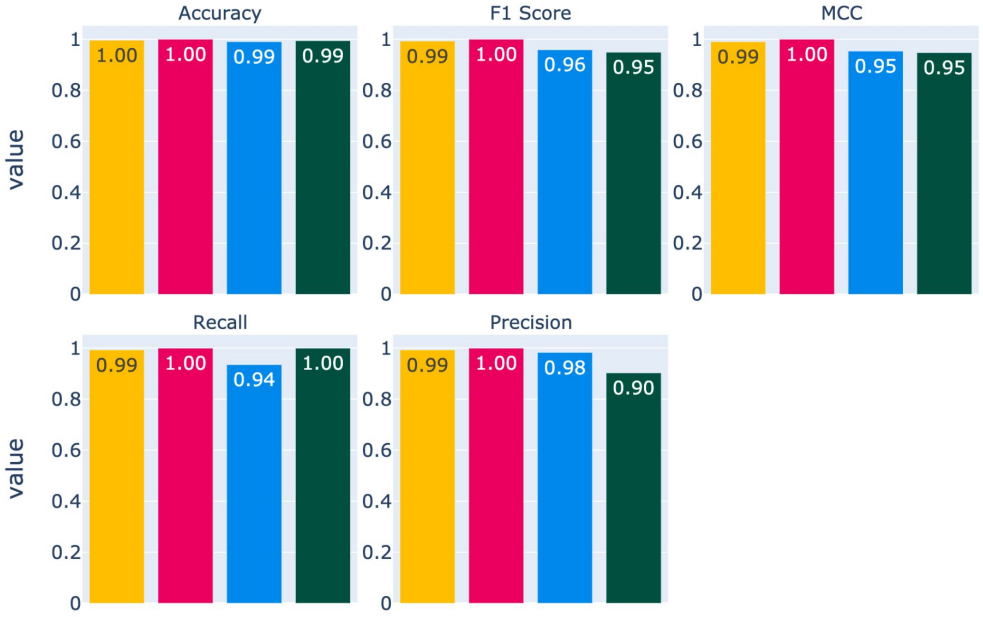
Touvron, H., Cord, M., Sablayrolles, A., Synnaeve, G., & Jégou, H. (2021). Going deeper with Image Transformers. CoRR, abs/2103.17239. https://arxiv.org/abs/2103.17239

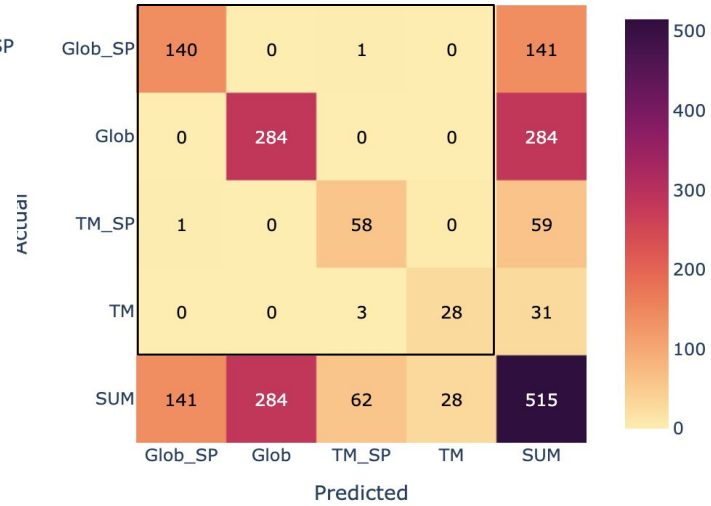
Transmembrane Dataset

- Predict the 4-state protein type:
  - **Glob_SP** ⇒ globular (non-membrane) proteins with signal peptides
  - **Glob** ⇒ globular (non-membrane) proteins without signal peptides
  - **TM_SP** ⇒ transmembrane proteins with signal peptides
  - **TM** ⇒ transmembrane proteins without signal peptides

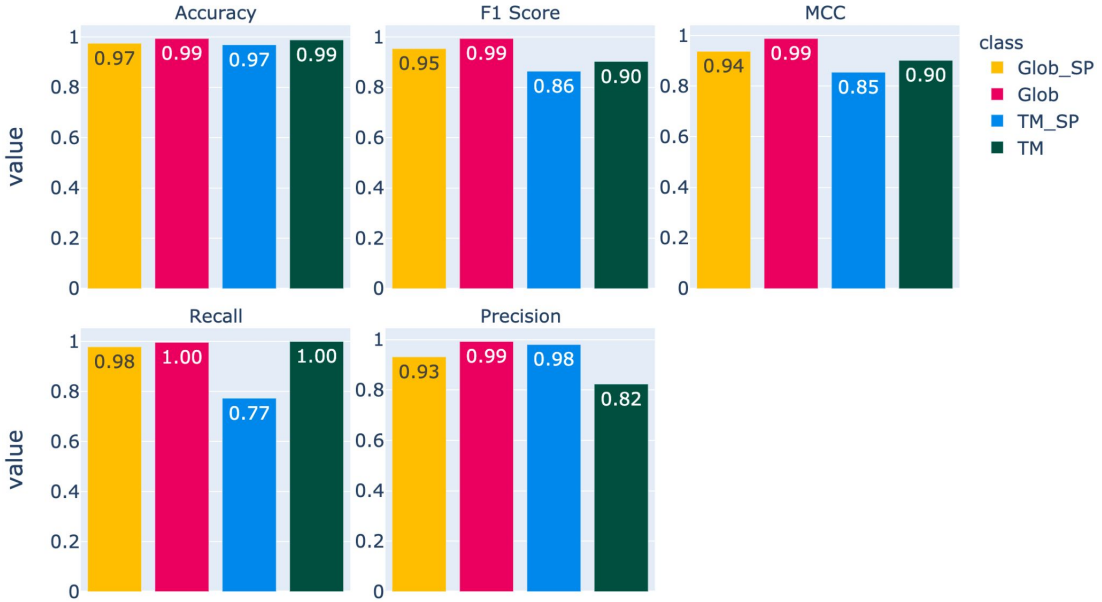


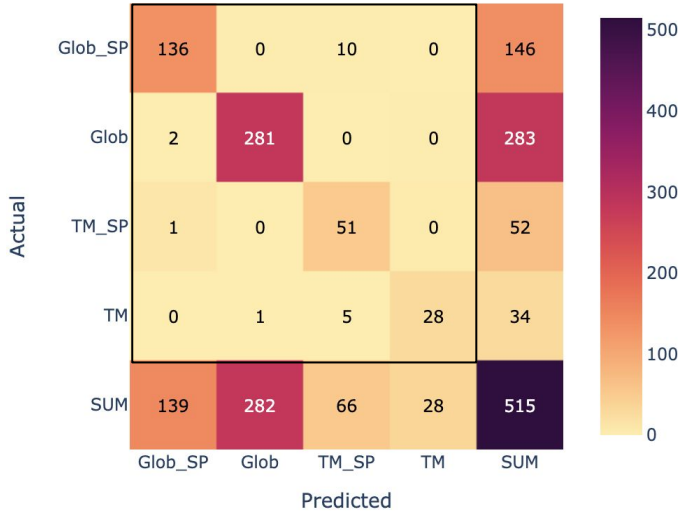
Simple MLP on Protein-Mean Embeddings





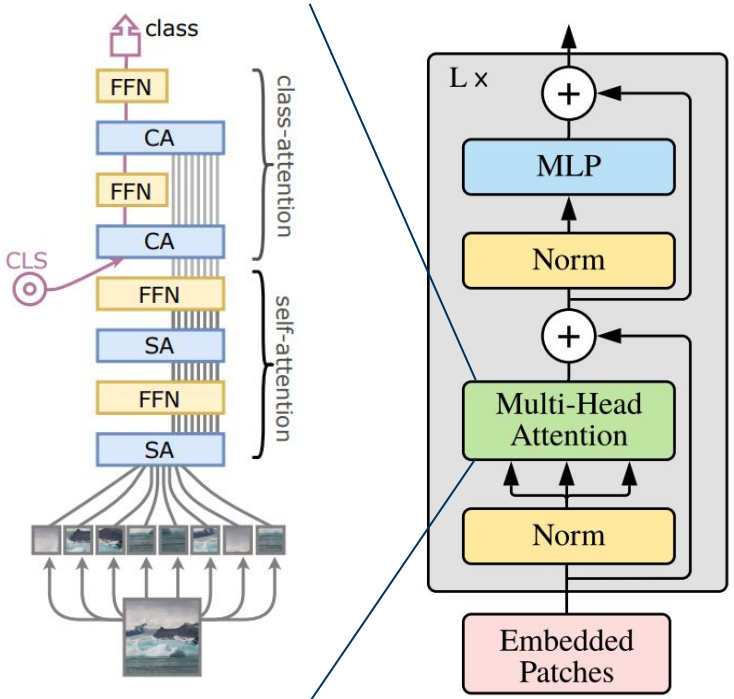
CNN on Protein-Mean Embeddings





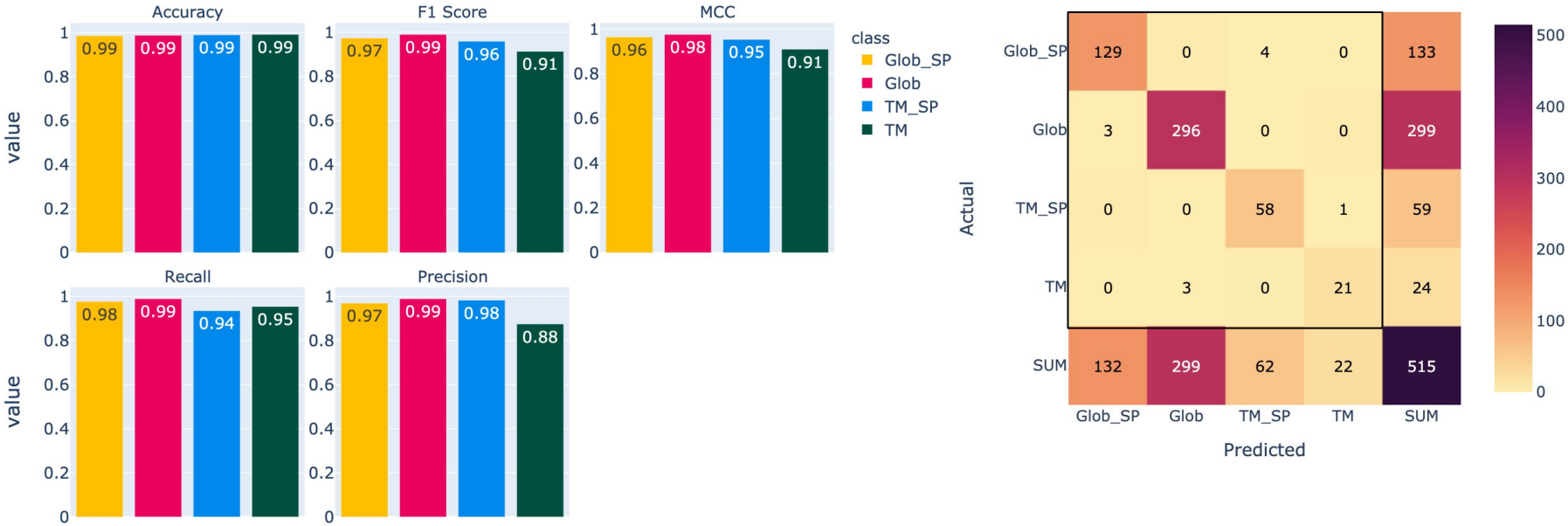
Class-Attention Model variations

Size	Embedding	Attention Heads	SA Blocks	CA Blocks	Acc
CaiT-XS	Mean	1	4	1	0.98
	Reduced	1	4	1	0.97
	Full	1	4	1	0.79
CaiT-S	Reduced	2	12	1	0.99
	Full	2	12	1	0.64
CaiT-M	Full	4	24	2	0.63
CaiT-L	Full	8	24	2	0.63

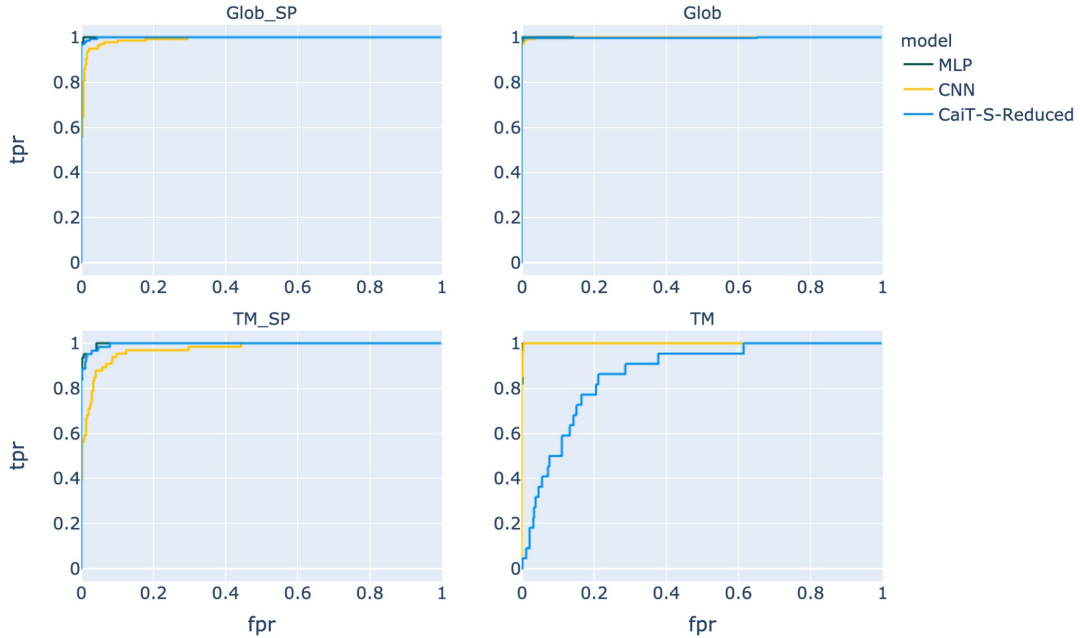


[1] https://wandb.ai/fga/pp2/reports/PP2-Transmembrane-Class-attention-experiments--VmlldzoxNDgyMzMx
[2] Touvron, H., Cord, M., Sablayrolles, A., Synnaeve, G., & Jégou, H. (2021). Going deeper with Image Transformers. CoRR, abs/2103.17239. https://arxiv.org/abs/2103.17239
[3] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

CaiT S - Reduced Embeddings



Comparison of Models



Comparison of Models

Discussion: CaiT for TMH Prediction

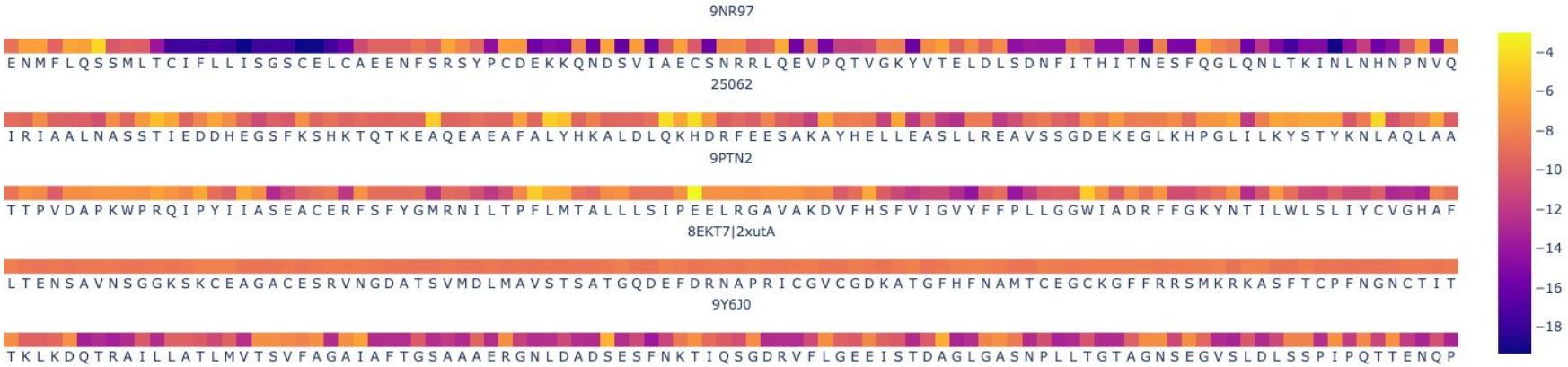
Advantages	Disadvantages
● More insight into learned rules ● Option to predict directly from sequence ● More capable to train on huge amounts of data	● Longer training time ● More energy consumption ● Bigger model ● Easier to overfit

Thanks for your **Attention**

Conclusion

- no cross-validation => not better than almost 1.0 accuracy

Sequence Attention Visualization



Amino Acid Attention Distribution