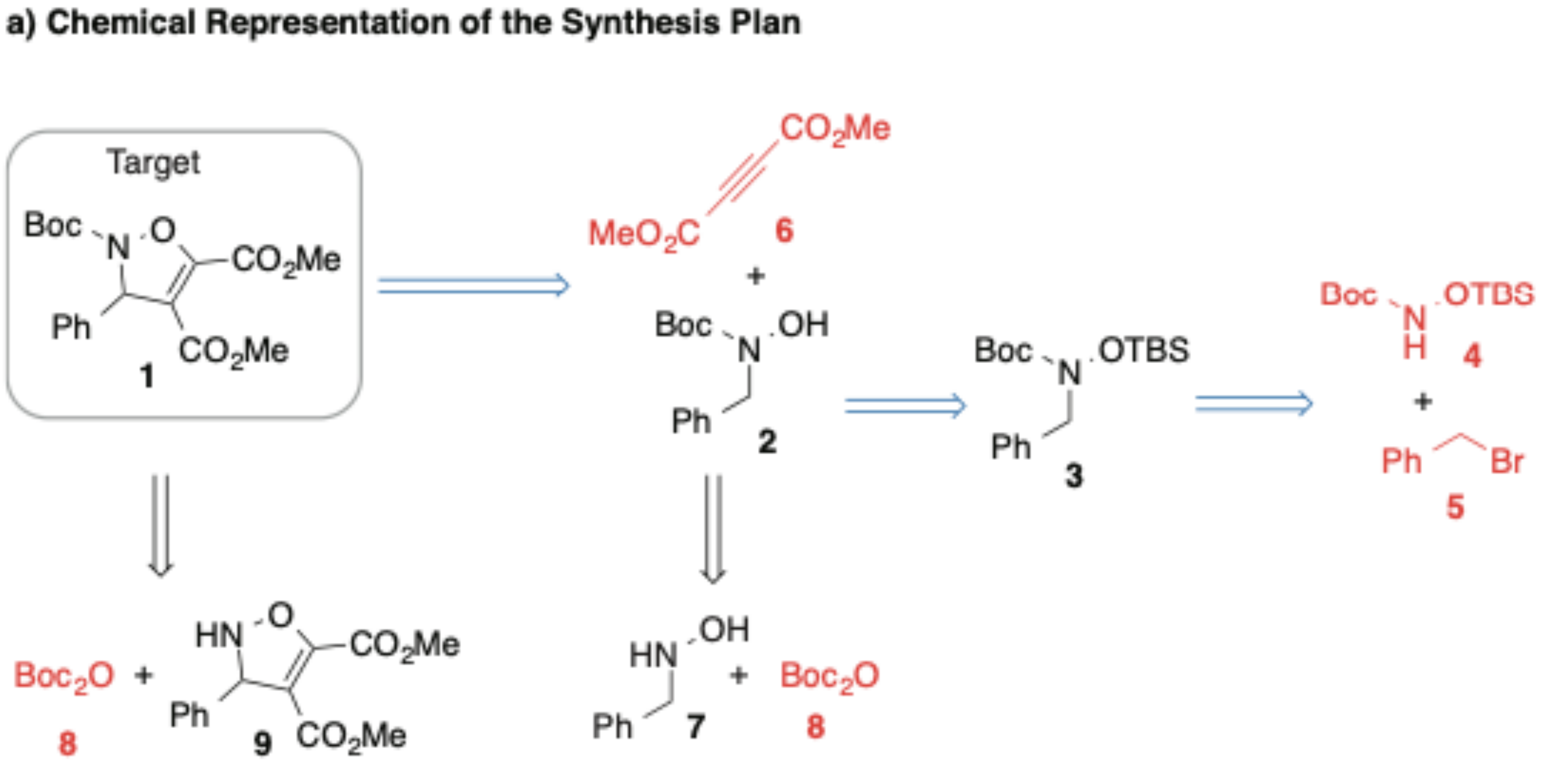
A TRANSFORMER MODEL FOR RETROSYNTHESIS

Karpov P., Godin G., Tetko I.V.: A Transformer Model for Retrosynthesis. In: Artificial Neural Networks and Machine Learning – ICANN 2019: Workshop and Special Sessions: 17th - 19th September 2019 2019; Münich. Springer International Publishing: 817-830.

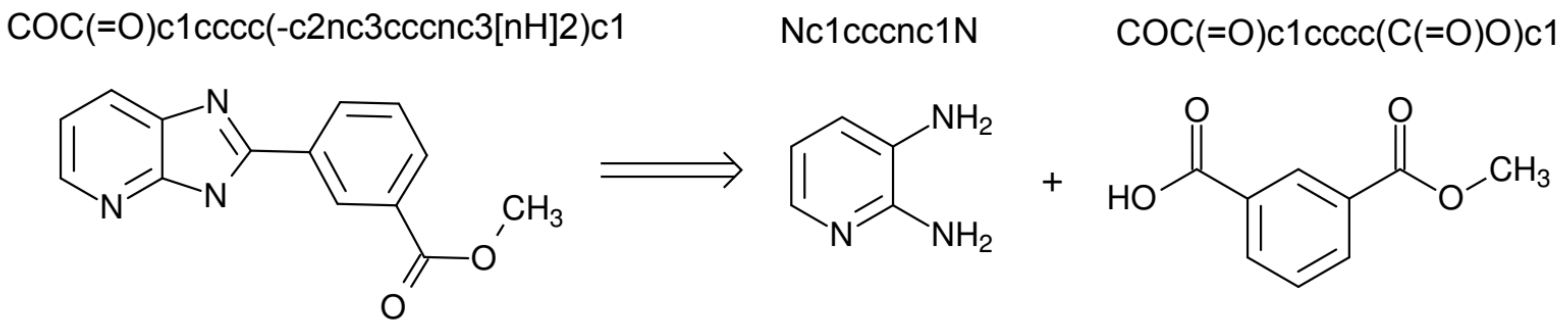
Hirah 31.07.2021

Retrosynthesis (逆合成路径设计)



Prediction of Chemical Synthesis 化学反应预测

- SMILES notations & Training Data



- COC(=O)c1cccc(-c2nc3cccnc3[nH]2)c1 <- Nc1cccnc1N.COC(=O)c1cccc(C(=O)O)c1

Anology between machine translation and retro synthesis

- Target Language 语言A <- Source Language 语言B
- Target Molecule 目标分子 <- Predecessors 反应前体
- Liu, B. et al. Retrosynthetic reaction prediction using neural sequence-to-sequence models. ACS Central Science 3, 1103–1113 (2017).
- Neural sequence-to-sequence (seq2seq) model

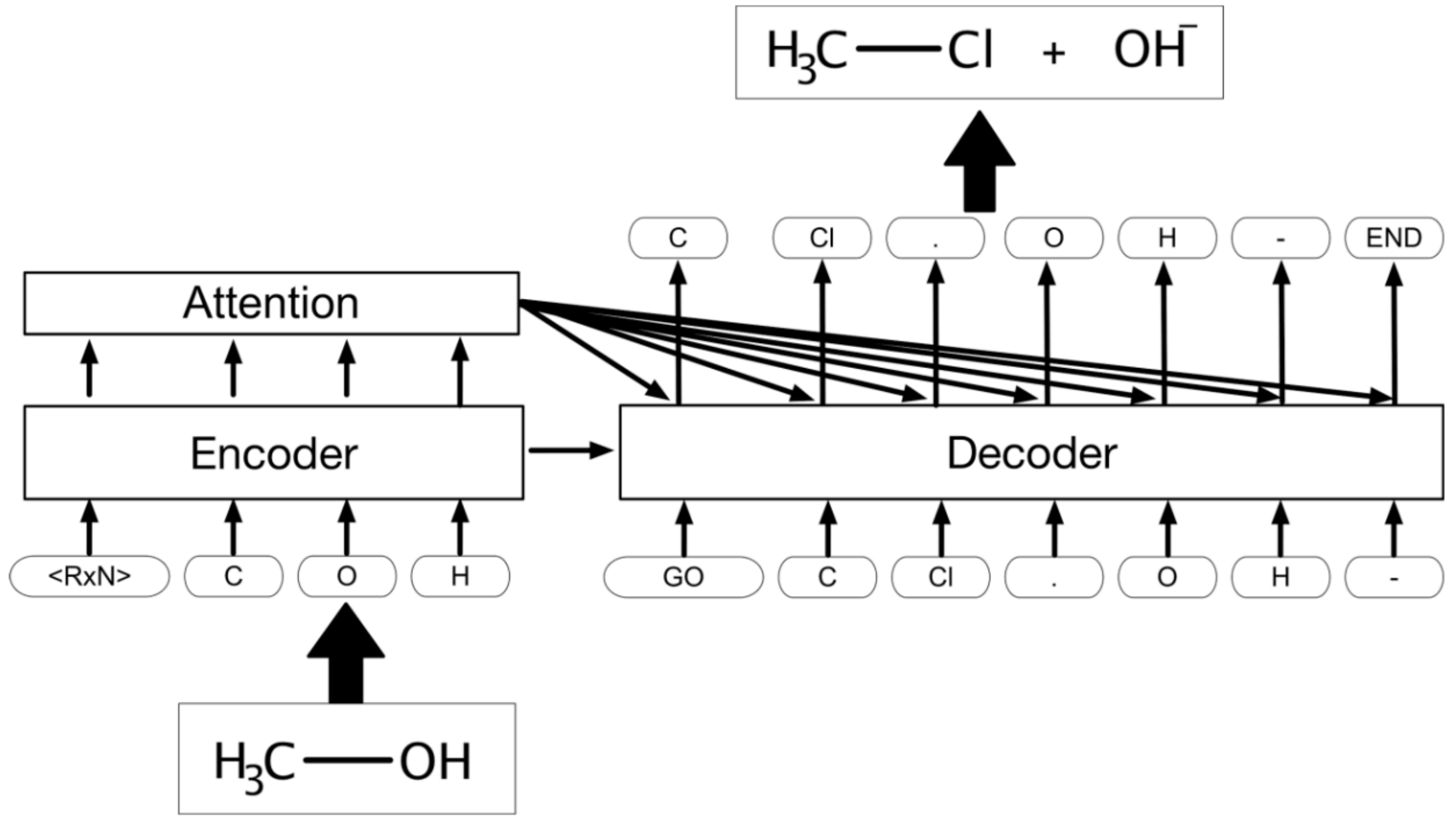


Figure 3: seq2seq model architecture

Transformer model inference

- Explore the internal representation of data
- Estimates the probability of the next symbol over the model's vocabulary (66) given all previous symbols in the string.
- Calculate the logits:

$$z_i = Transformer(\{x_1, x_2, x_3, ..., x_L\}, \{y_1, y_2, y_3, ..., y_{i-1}\})$$

- Convert to probabilities:

$$q_i = \frac{exp(z_i/T)}{\sum_{j=0}^{V} exp(z_j/T)}$$

- Greedy - $y_{i} = q_{I}$
- Beam Search 集束搜索：每次保留Top - K个模型推荐的结果

Heuristics for Training A Transformer

- 使用较大的Batch Size
- 在开始时增加Learning Rate
- Applying Cyclic Learning Rate：在优化后期增大学习率避免local minima

Results

Table 1: Accuracy (%) of the models on test set when all reactants were correctly predicted.

Model	Greedy	Top-1	Top-3	Top-5	Description
Seq2Seq		37.4	52.4	57.0	Literature result from [17] based on Seq2Seq architecture.
T1	34.4	37.9	57.3	62.7	Transformer Model trained with validation control set (early stopping, ~200 epochs).
$T1_1$	37.3	39.8	59.1	63.9	The same as T1, but without early stopping (1000 epochs).
T2	39.3	41.8	61.3	67.2	Transformer Model trained on both training and validation sets for 1000 epochs.
T3	**40.6**	**42.7**	**63.9**	**69.8**	Transformer Model trained with cyclic learning rate schedule for 1000 epochs. Averaging cycles 6,7,8,9, and 10.

Results

https://chemrxiv.org/engage/chemrxiv/article-details/60c7417af96a00cd49286446

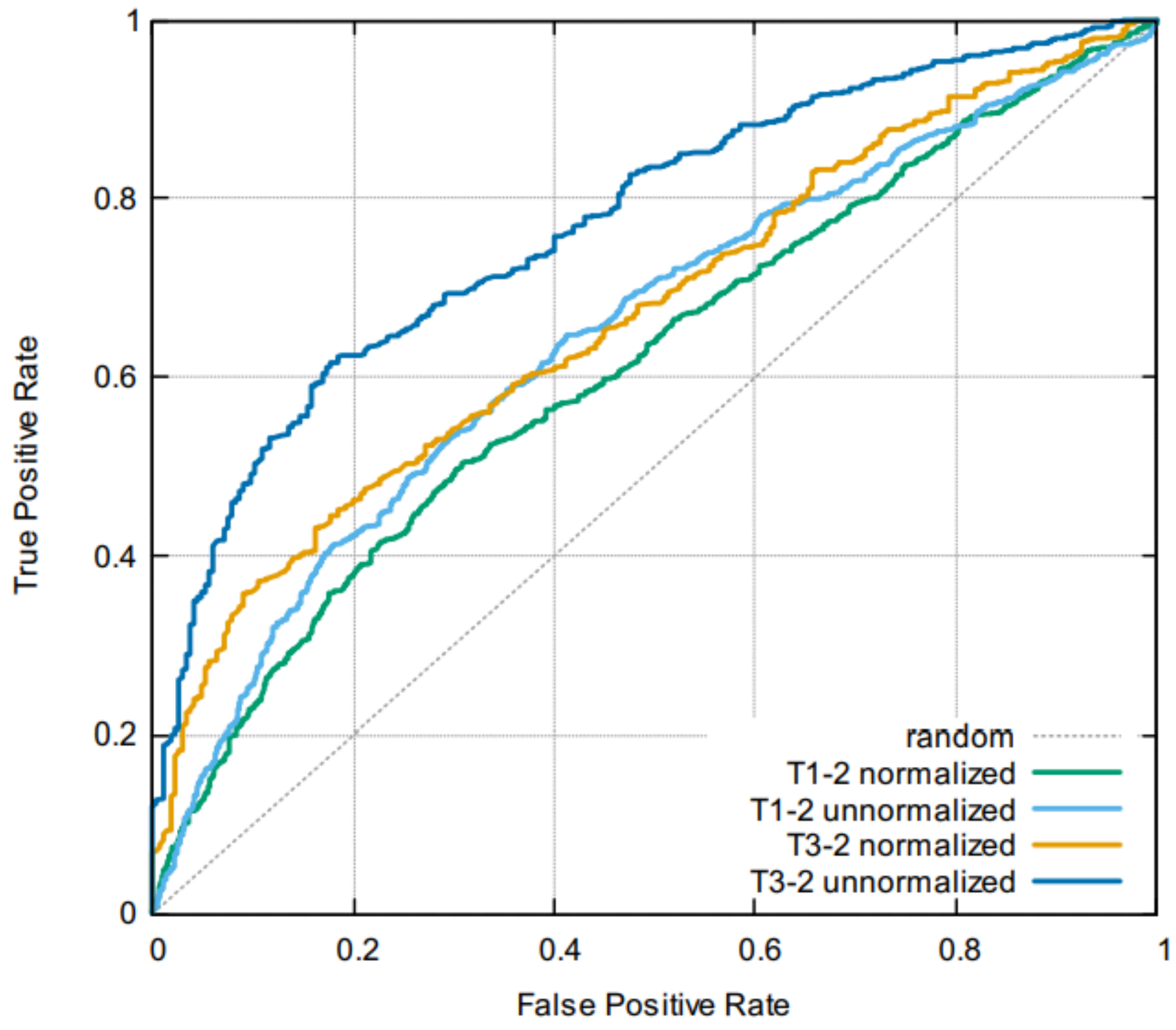


Figure 4: Internal classification performance.