GlycanML: A Multi-Task and Multi-Structure Benchmark for Glycan Machine Learning

Minghao Xu

Glycans are ubiquitous

Starch

Bread

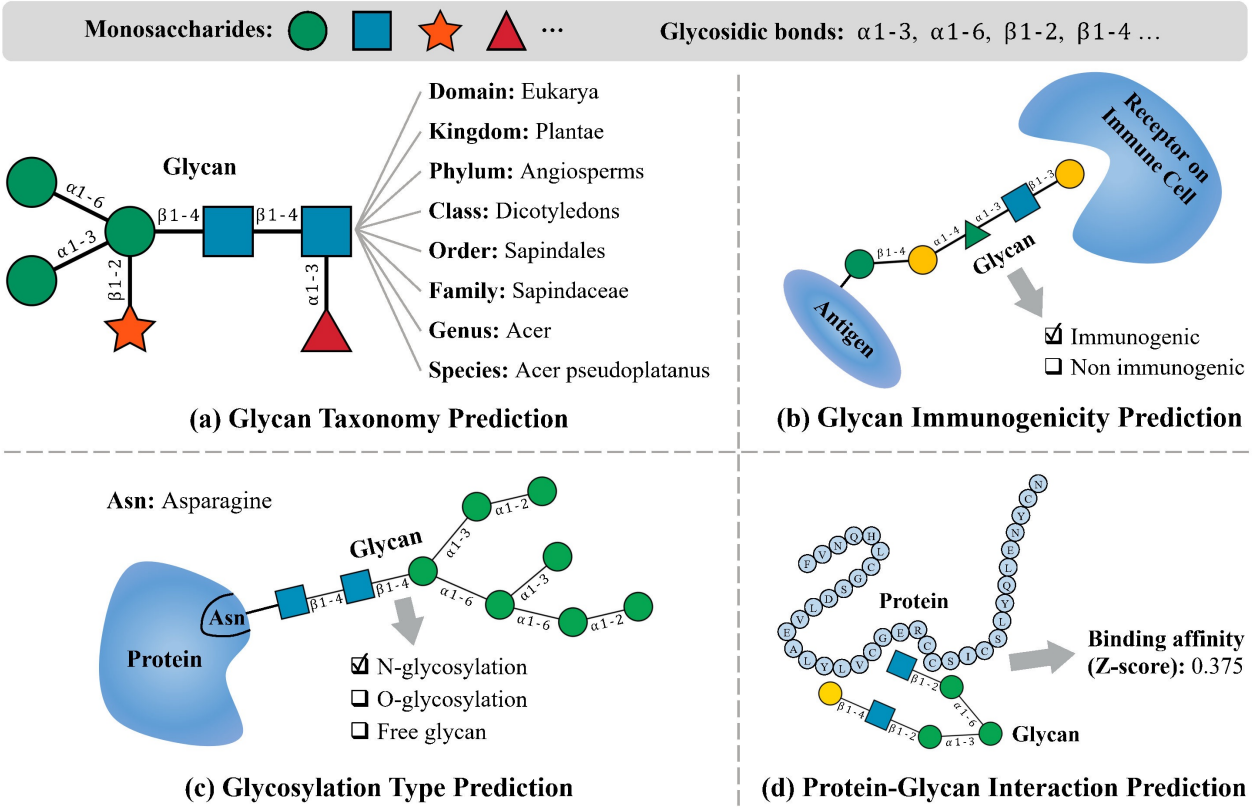
Rice

Sugars

Orange
(fructose)

Milk
(lactose)

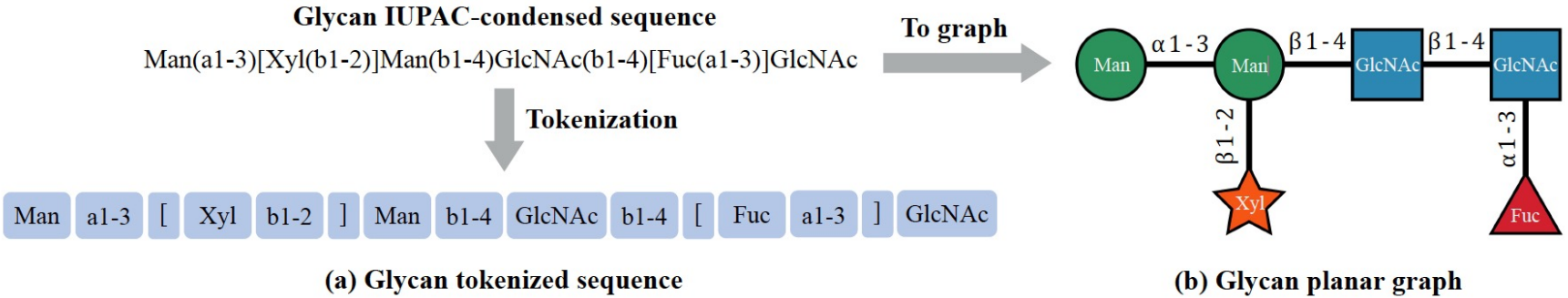
Glycan understanding with machine learning



GlycanML: Building a Comprehensive glycan machine learning benchmark

Task	Task type	#Mono. per glycan	#Sample	#Train/Validation/Test	Metric
Taxonomy prediction of *Domain*	Classification	$6.39_{(3.51)}$	13,209	11,010/1,280/919	Accuracy (%)
Taxonomy prediction of *Kingdom*	Classification	$6.39_{(3.51)}$	13,209	11,010/1,280/919	Accuracy (%)
Taxonomy prediction of *Phylum*	Classification	$6.39_{(3.51)}$	13,209	11,010/1,280/919	Accuracy (%)
Taxonomy prediction of *Class*	Classification	$6.39_{(3.51)}$	13,209	11,010/1,280/919	Accuracy (%)
Taxonomy prediction of *Order*	Classification	$6.39_{(3.51)}$	13,209	11,010/1,280/919	Accuracy (%)
Taxonomy prediction of *Family*	Classification	$6.39_{(3.51)}$	13,209	11,010/1,280/919	Accuracy (%)
Taxonomy prediction of *Genus*	Classification	$6.39_{(3.51)}$	13,209	11,010/1,280/919	Accuracy (%)
Taxonomy prediction of *Species*	Classification	$6.39_{(3.51)}$	13,209	11,010/1,280/919	Accuracy (%)
Immunogenicity prediction	Binary classification	$7.30_{(3.78)}$	1,320	1,026/149/145	AUPRC
Glycosylation type prediction	Classification	$9.04_{(3.96)}$	1,683	1,356/163/164	Accuracy (%)
Protein-Glycan interaction prediction	Regression	$6.56_{(4.54)}$	564,647	442,396/58,887/63,364	Spearman's ρ

GlycanML: Supporting Multiple glycan representation structures



(a) Glycan tokenized sequence

(b) Glycan planar graph

GlycanML: Maintaining a Leaderboard of latest models

Rank	Method	Mean Rank	Ranks: Domain → Interaction	Reference
1	RGCN	2.5	[1, 5, 1, 1, 1, 1, 2, 2, 2, 8 ,3]	paper
2	CNN	3.5	[7, 6, 2, 2, 2, 2, 3, 5, 3, 2, 4]	paper
3	CompGCN	3.9	[5, 1, 3, 3, 4, 3, 1, 1, 7, 10, 5]	paper
4	GIN	5.1	[2, 3, 4, 4, 10, 5, 6, 6, 6, 4, 6]	paper
5	MPNN	5.6	[6, 7, 5, 5, 3, 4, 4, 4, 10, 3, 10]	paper
6	ResNet	6.0	[8, 8, 7, 6, 5, 8, 8, 9, 4, 1, 2]	paper
7	LSTM	6.3	[9, 9, 6, 7, 6, 6, 9, 10, 1, 5, 1]	paper
8	GAT	6.6	[4, 2, 8, 9, 7, 7, 5, 3, 9, 9, 9]	paper
9	GCN	7.2	[3, 4, 10, 8, 8, 9, 7, 7, 8, 7, 8]	paper
10	Transformer	8.5	[10, 10, 9, 10, 9, 10, 10, 8, 5, 6, 7]	paper

Thank you!

Paper

Project Page

Code