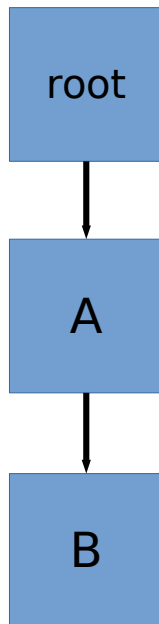


Rice gene functional data

Miguel Romero
Ómicas P5
Nov 6th, 2020

Gene functions

Hierarchy of
biological processes



By the *true-path* rule:

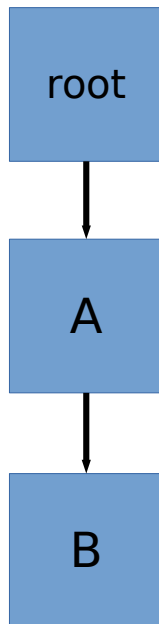
If a gene is associated to B then it must be associated to A.

If a gene is not associated to A then it must not be associated to B.

But, if a gene is associated to A, we do not know if it is associated to B.

Gene functions

Hierarchy of biological processes

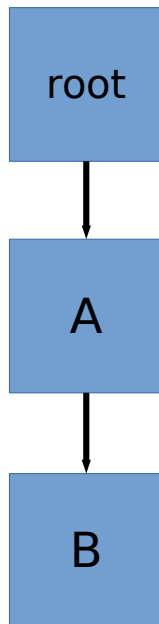


Gene-function association matrix

| Genes | A | B | ... |
|-------|-----|-----|-----|
| 1 | 0 | 0 | |
| 2 | 1 | 1 | |
| 3 | 1 | 1/0 | |
| 4 | 1/0 | 0 | |
| ... | | | |

Gene functions

Hierarchy of
biological processes

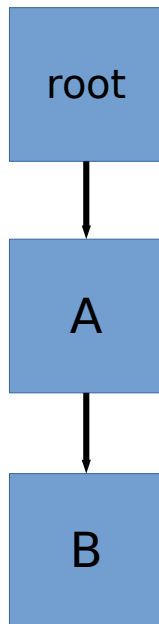


Gene-function
association matrix

| Genes | A | B | ... |
|-------|---|---|-----|
| 1 | 0 | 0 | |
| 2 | 1 | 1 | |
| 3 | 1 | 0 | |
| 4 | 1 | 1 | |
| 5 | 1 | 0 | |
| 6 | 1 | 1 | |
| 7 | 0 | 0 | |
| ... | | | |

Gene functions

Hierarchy of biological processes

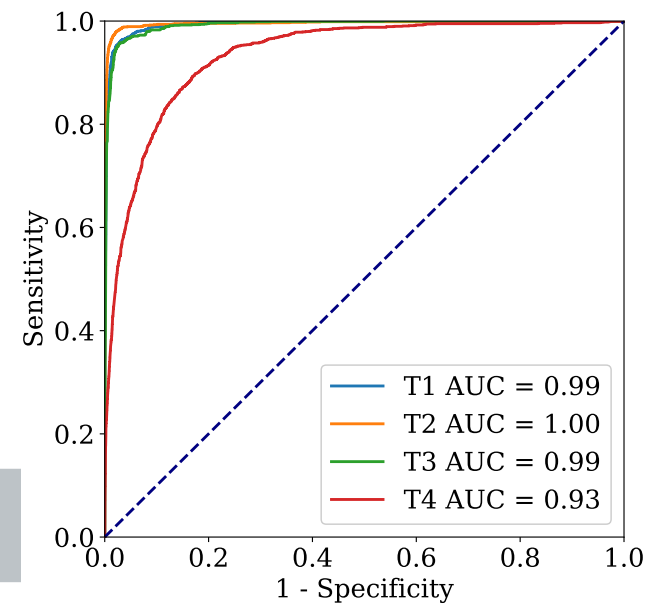
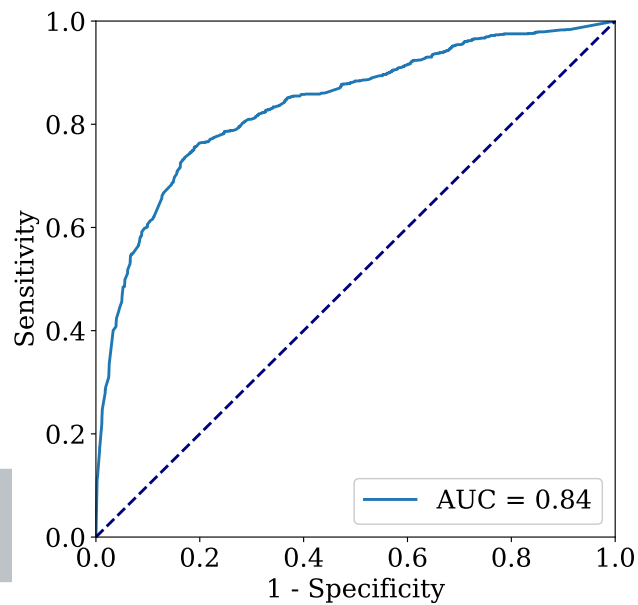
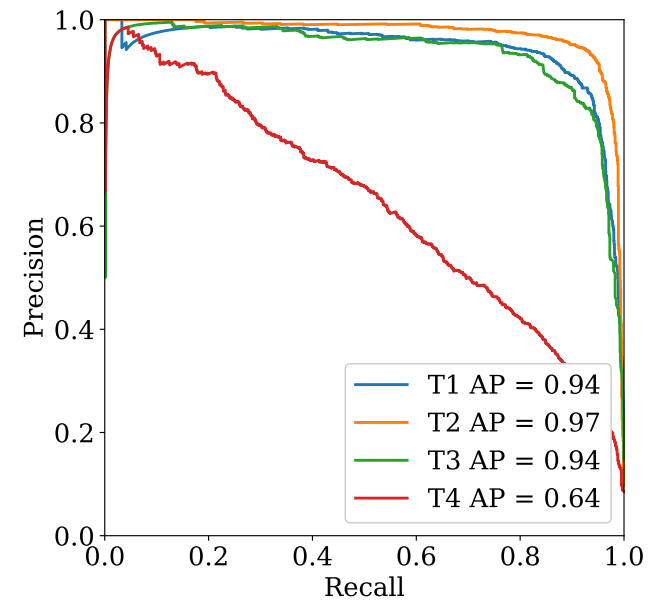
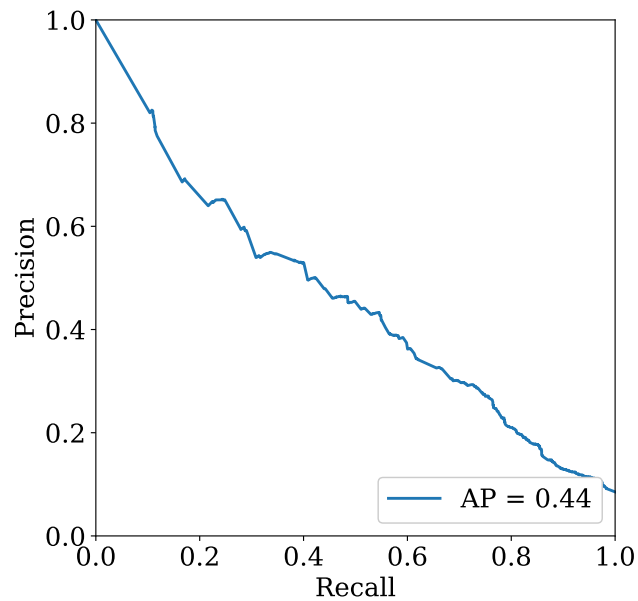


Gene-function association matrix

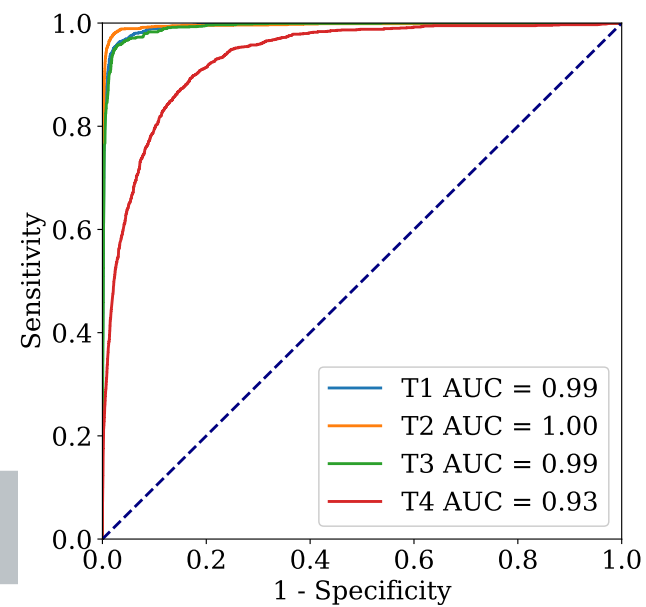
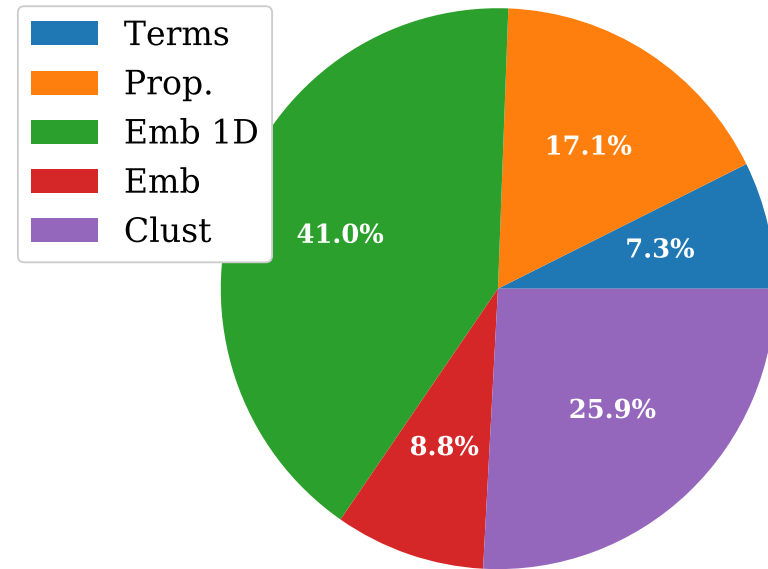
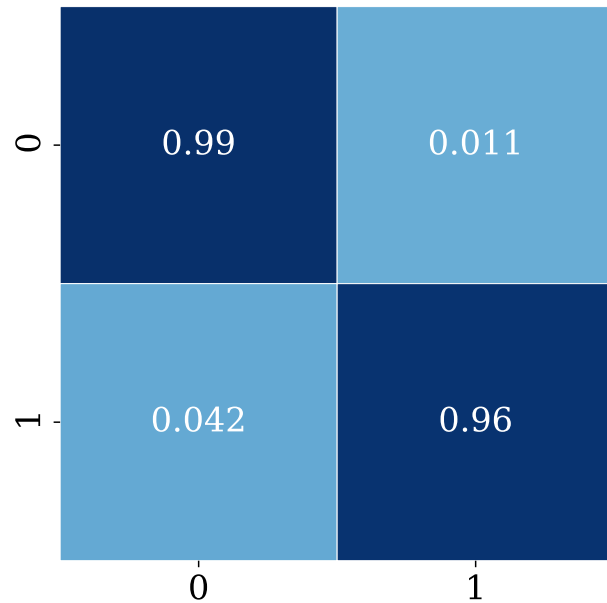
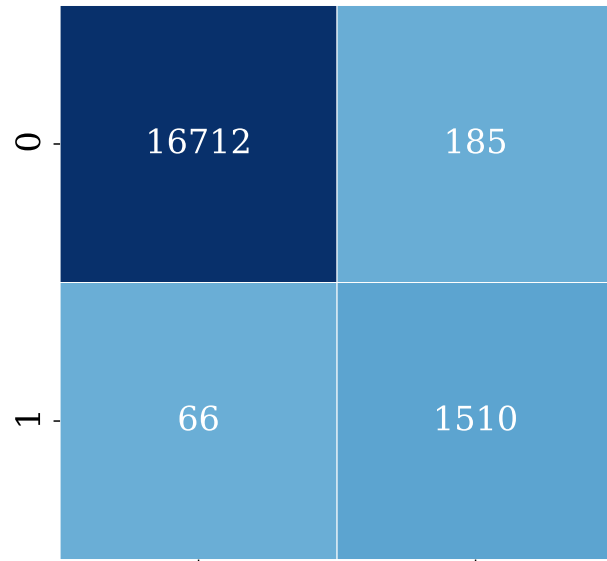
| Genes | A | B |
|-------|---|---|
| 1 | 0 | 0 |
| 2 | 1 | 1 |
| 3 | 1 | 0 |
| 4 | 1 | 1 |
| 5 | 1 | 0 |
| 6 | 1 | 1 |
| 7 | 0 | 0 |
| ... | | |

A and B are **correlated**

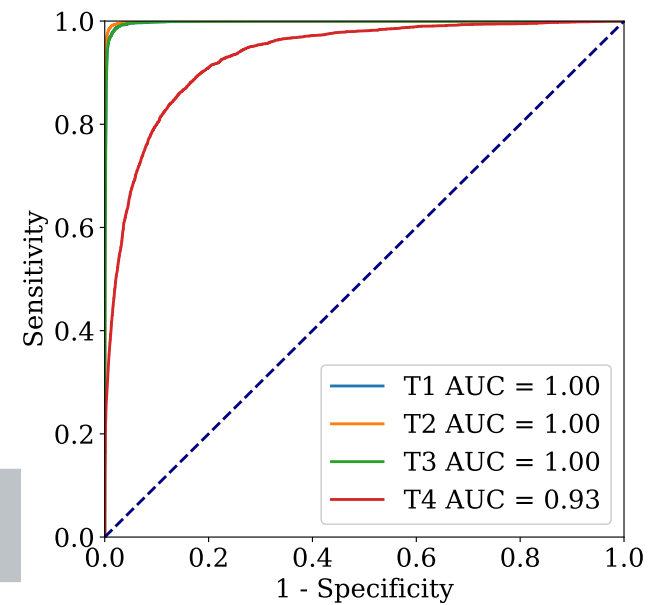
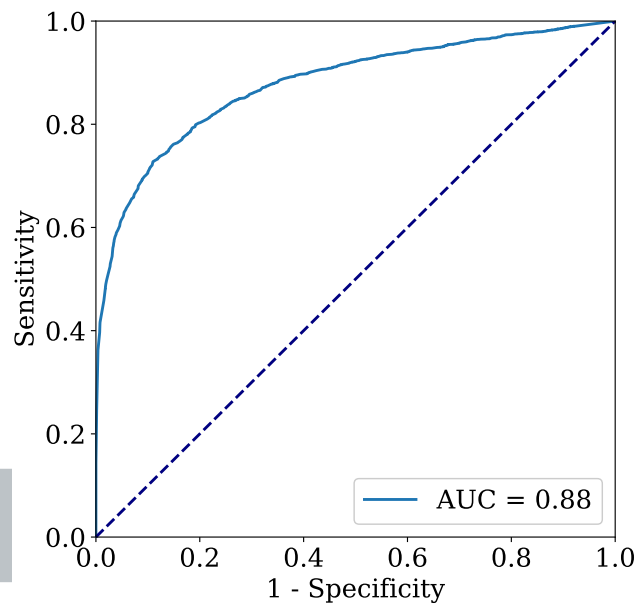
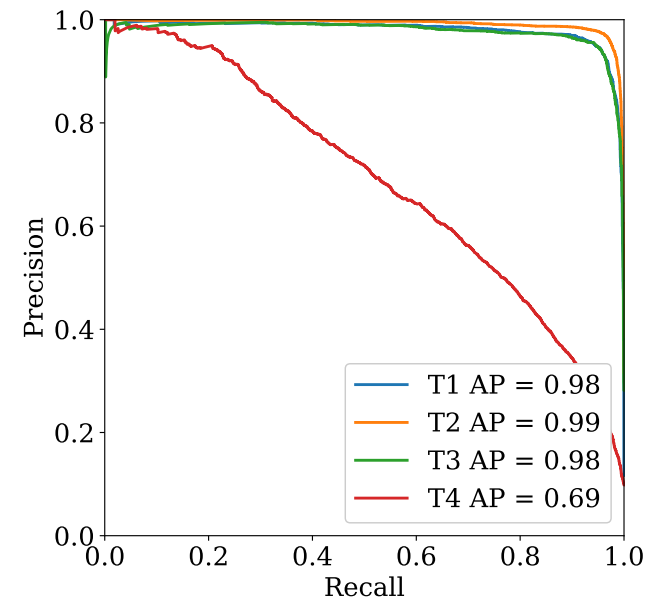
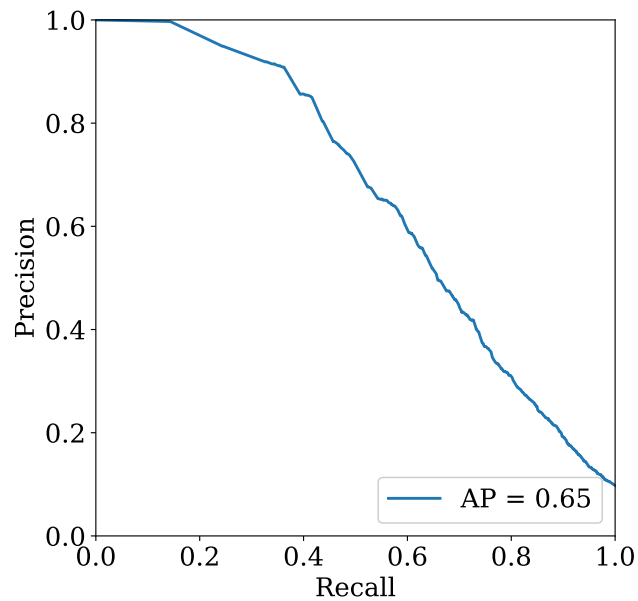
HBN vs. ML: GO:0044085



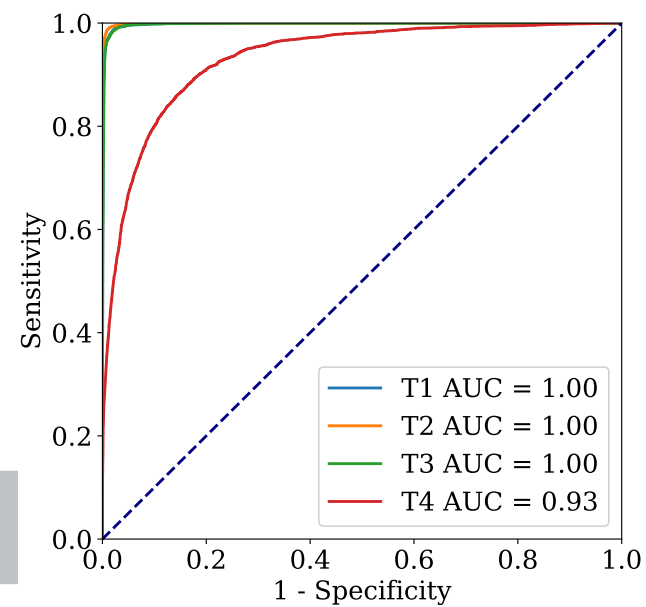
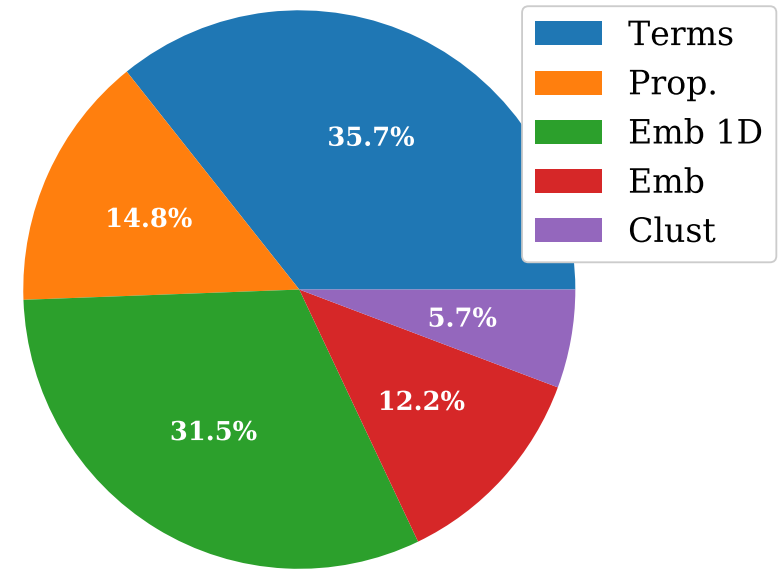
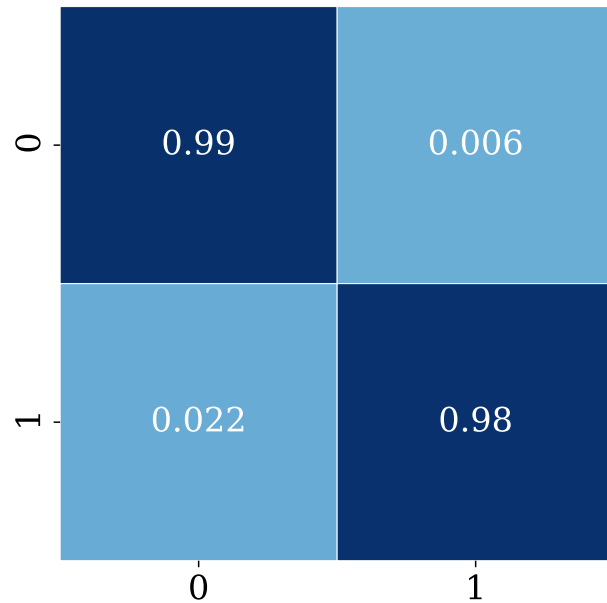
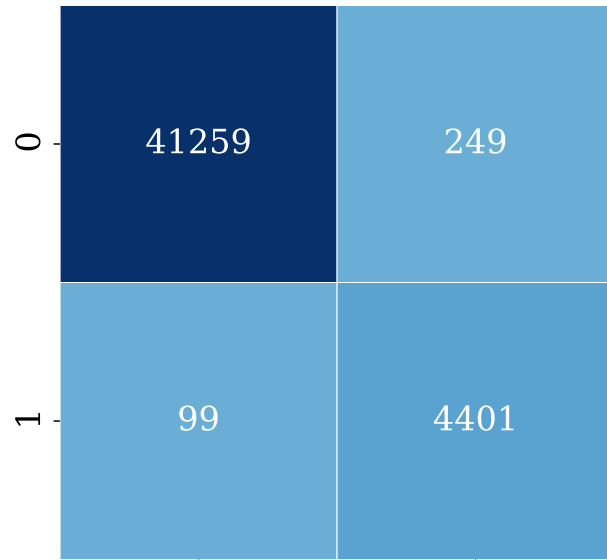
HBN vs. ML: GO:0044085



HBN vs. ML: GO:0000003



HBN vs. ML: GO:0000003

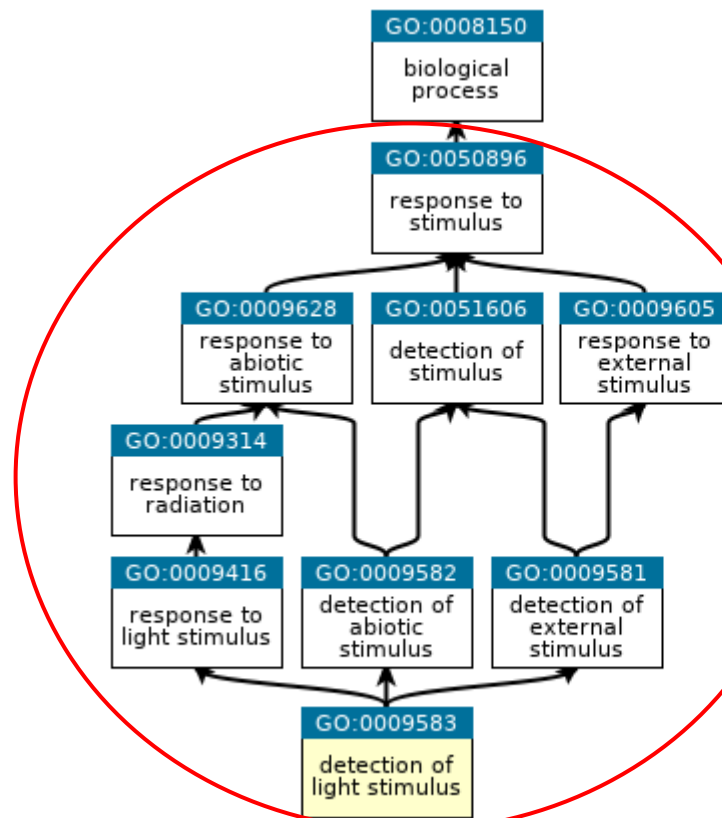


Gene functions

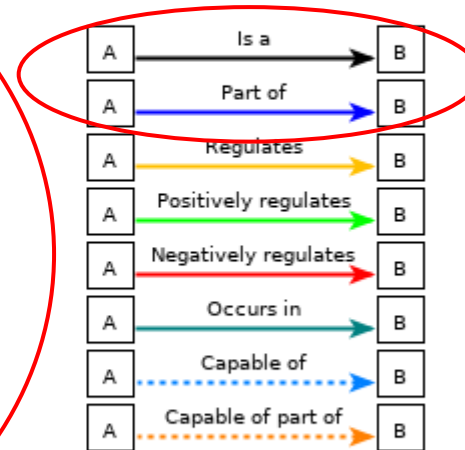
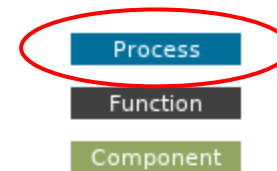
GO:0009583 detection of light stimulus

Ancestor Chart

Function
and **all its**
ancestors



QuickGO - <https://www.ebi.ac.uk/QuickGO>

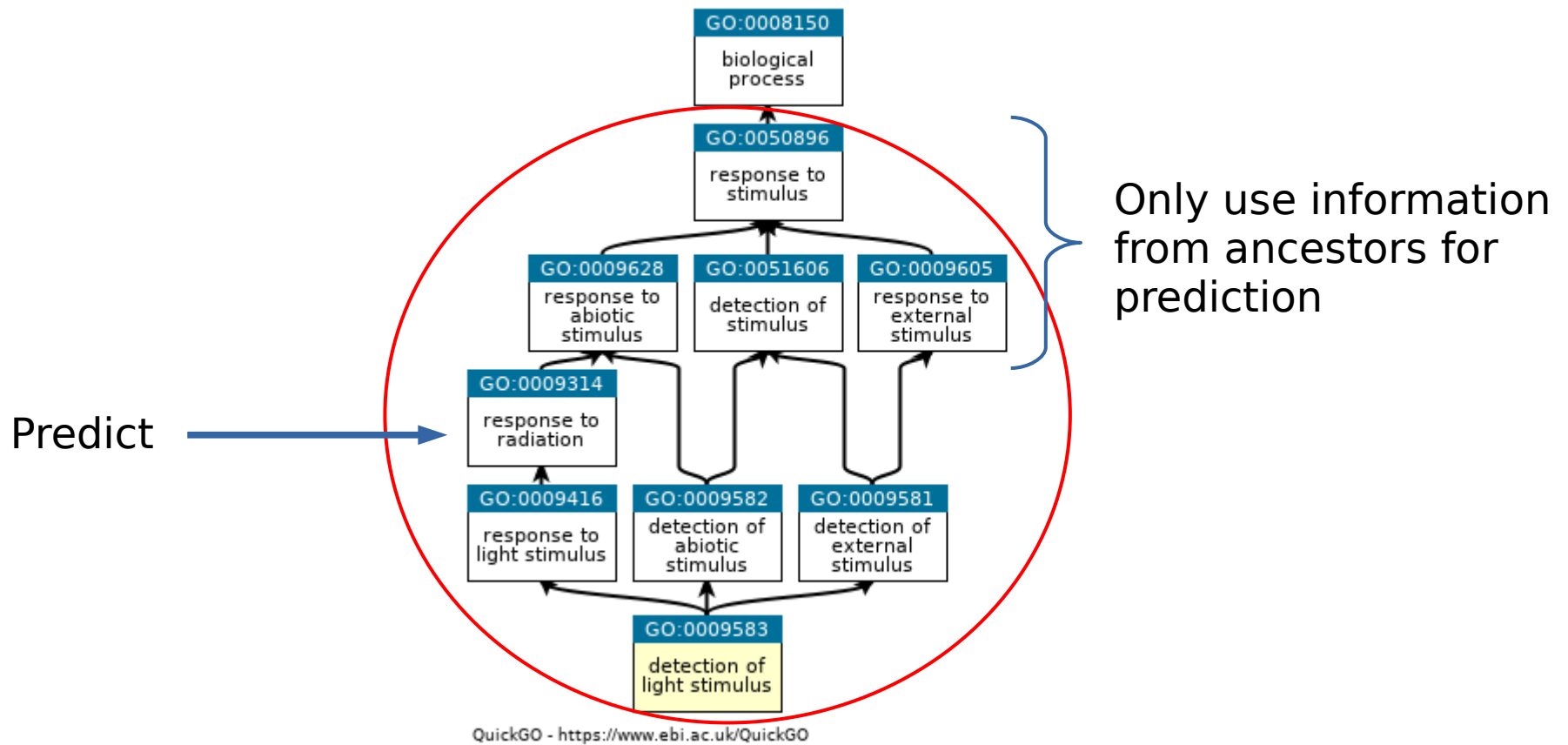


Biological
processes

Gene functions

GO:0009583 detection of light stimulus

Ancestor Chart



Prediction models

- **HBN:** probabilistic model.
- **Hierarchical** (Hie): uses all ancestors of function.
- **Ancestor** (Anc): uses only the parent of the function.
- **Structural properties** (Str): does not include information from ancestors.

Prediction models

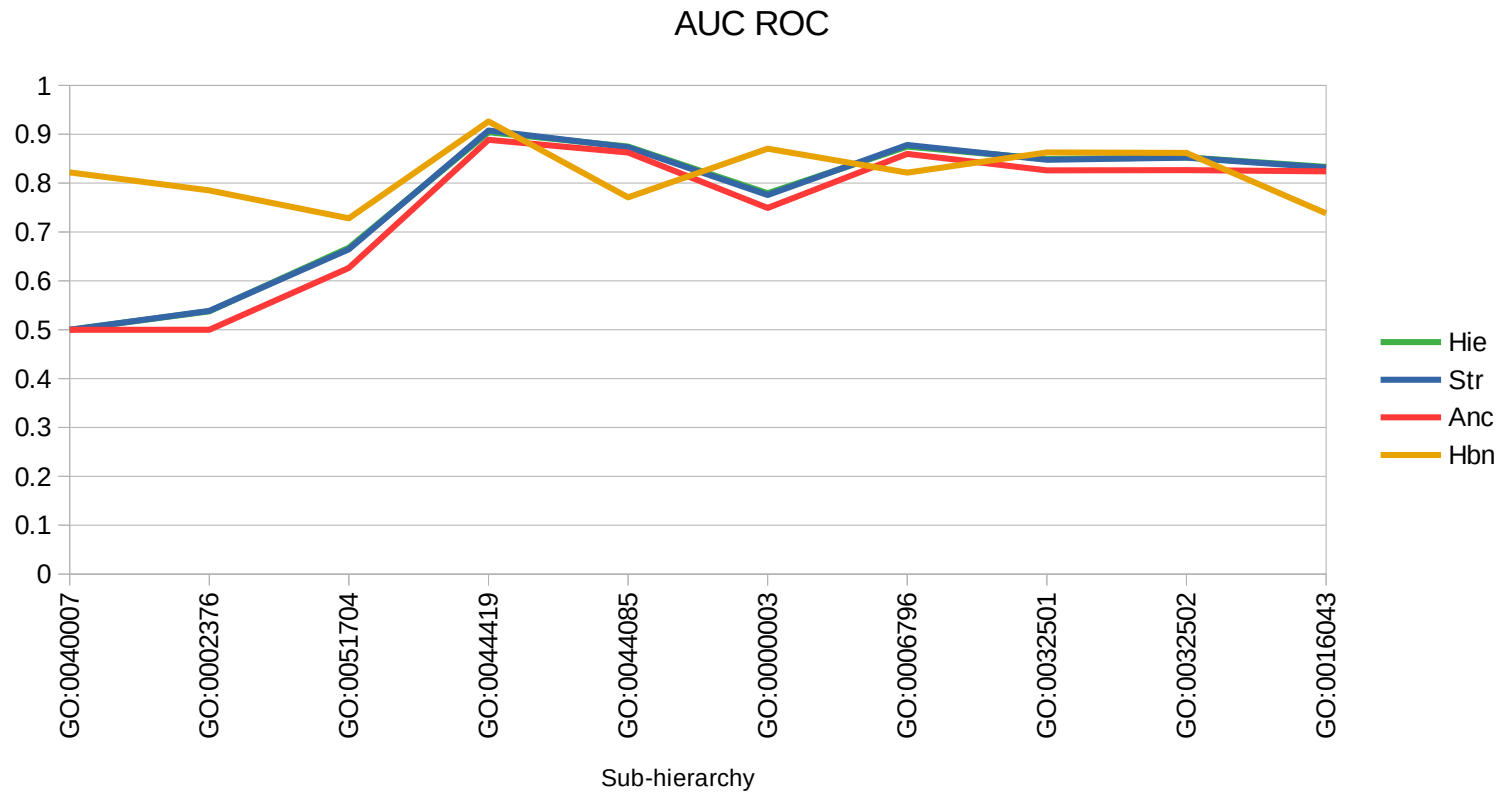
Hierarchical, Ancestor and Structural properties:

- Are **different** datasets used for the ML approach.
- Use the features related to **structural** properties of the GCN.
- Use **BFS** to predict functions.

Sub-hierarchies

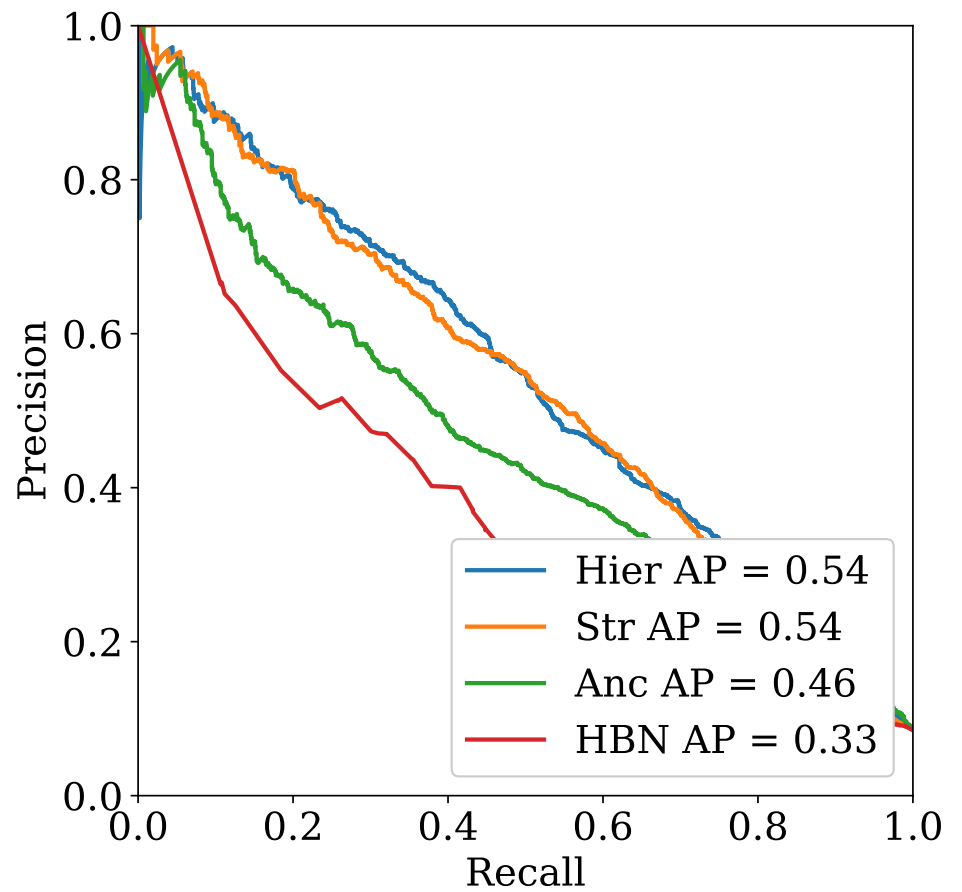
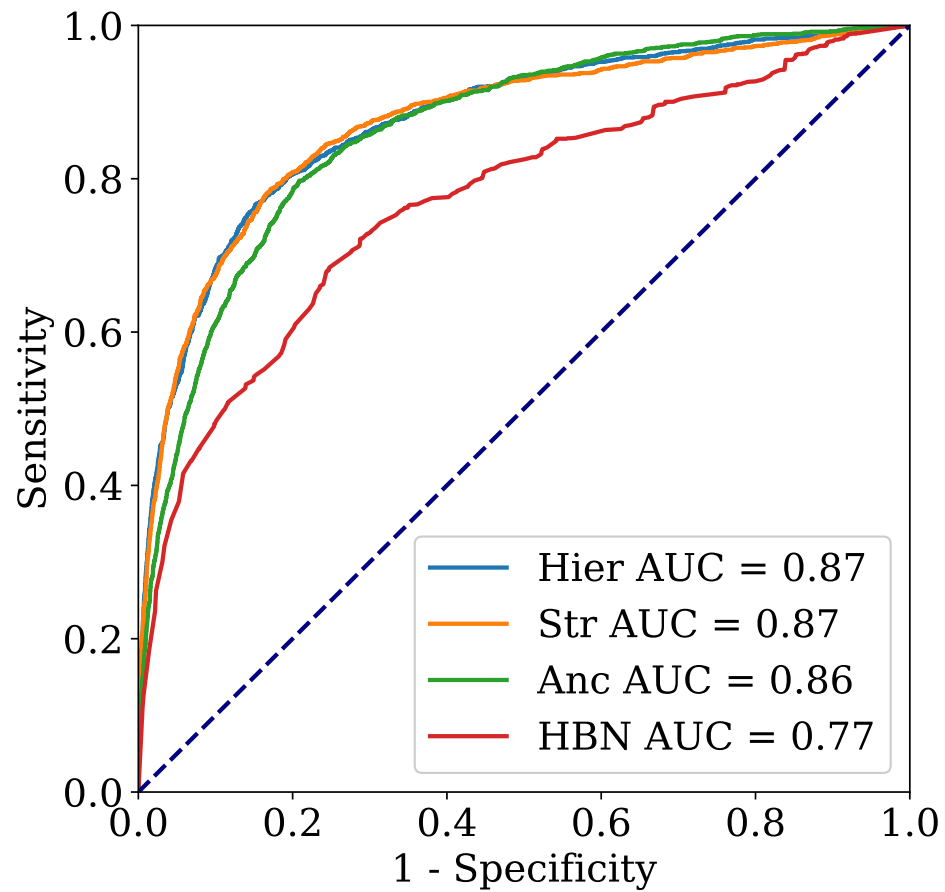
| Root | Terms | Pred | Genes | Desc |
|------------|-------|------|-------|---|
| GO:0040007 | 22 | 10 | 108 | growth |
| GO:0002376 | 36 | 11 | 131 | immune system process |
| GO:0051704 | 40 | 22 | 144 | multi-organism process |
| GO:0044419 | 78 | 37 | 777 | interspecies interaction between organisms |
| GO:0044085 | 121 | 50 | 377 | cellular component biogenesis |
| GO:0000003 | 146 | 72 | 648 | reproduction |
| GO:0006796 | 209 | 118 | 1270 | phosphate-containing compound metabolic process |
| GO:0032501 | 250 | 118 | 1043 | multicellular organismal process |
| GO:0032502 | 290 | 149 | 1063 | developmental process |
| GO:0016043 | 298 | 140 | 661 | cellular component organization |

First trial



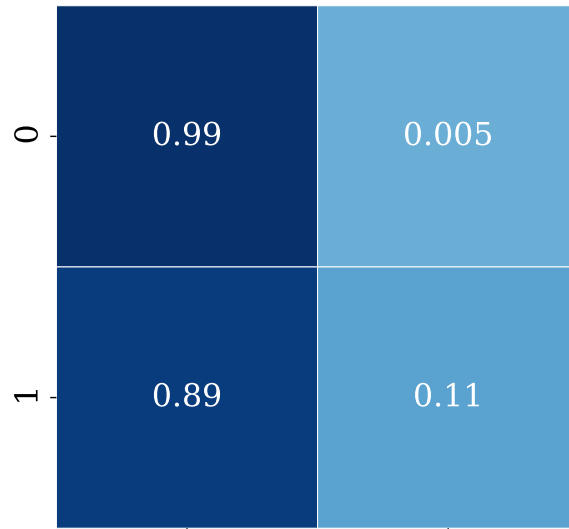
The first trial is straightforward as defined.

HBN vs. ML: GO:0044085

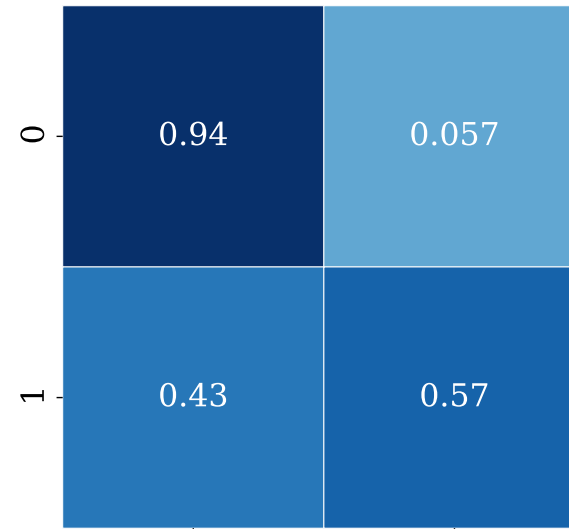


HBN vs. ML: GO:0044085

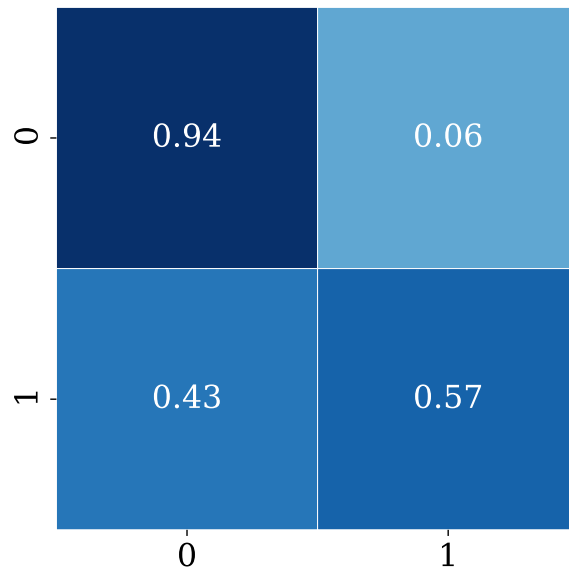
HBN



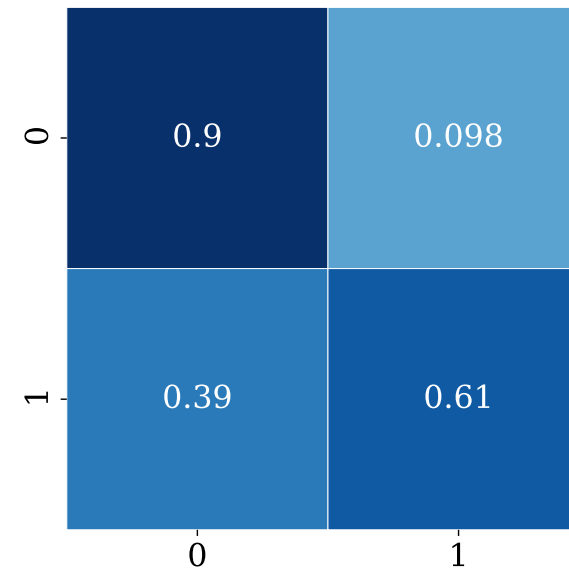
Str



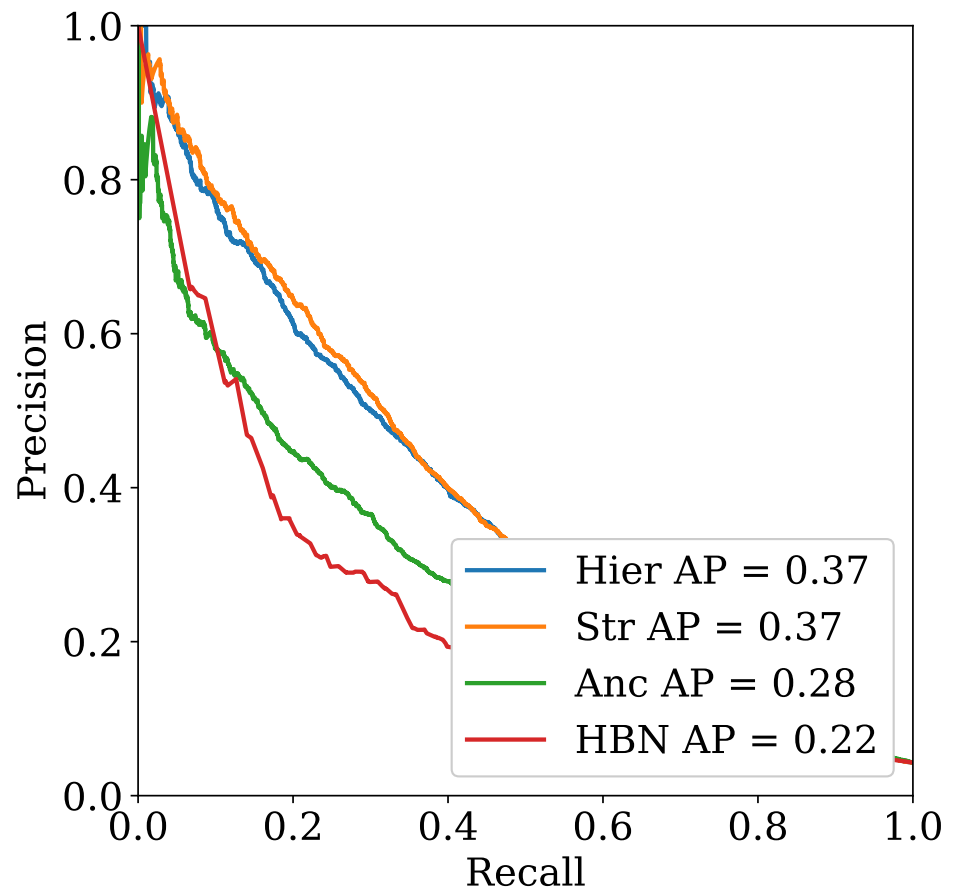
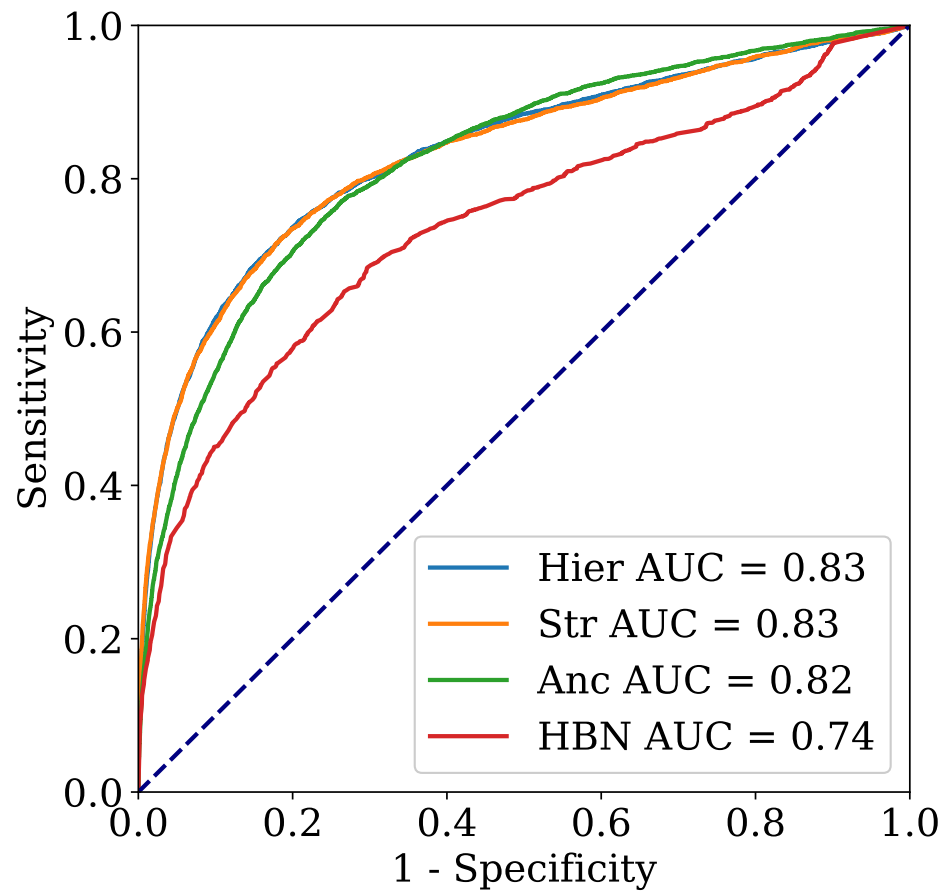
Hie



Anc

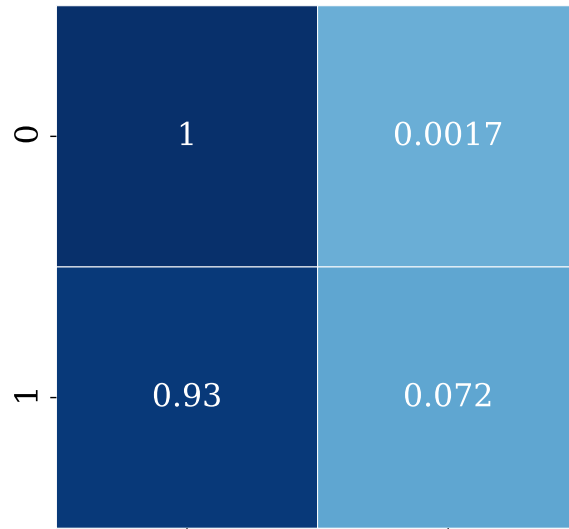


HBN vs. ML: GO:0016043

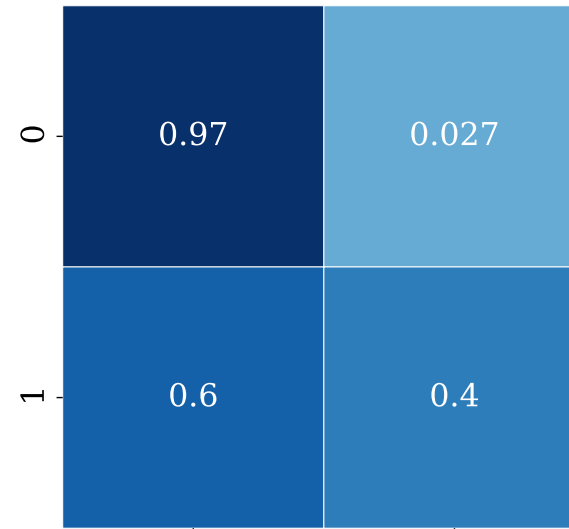


HBN vs. ML: GO:0016043

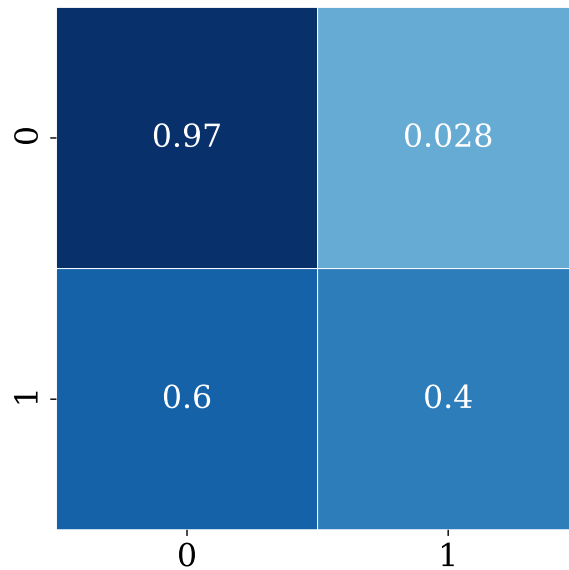
HBN



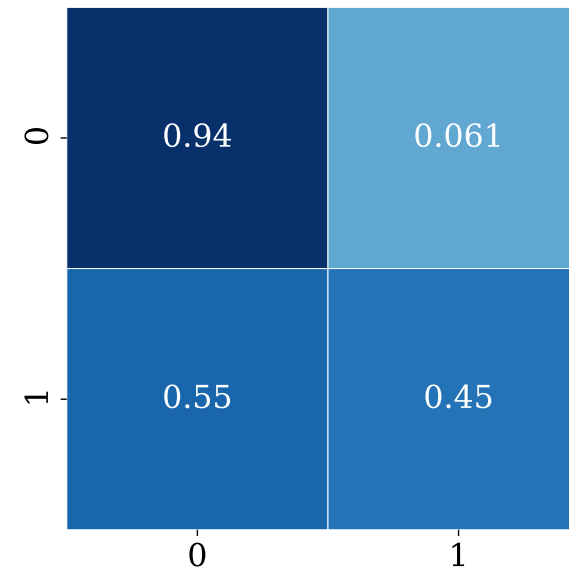
Str



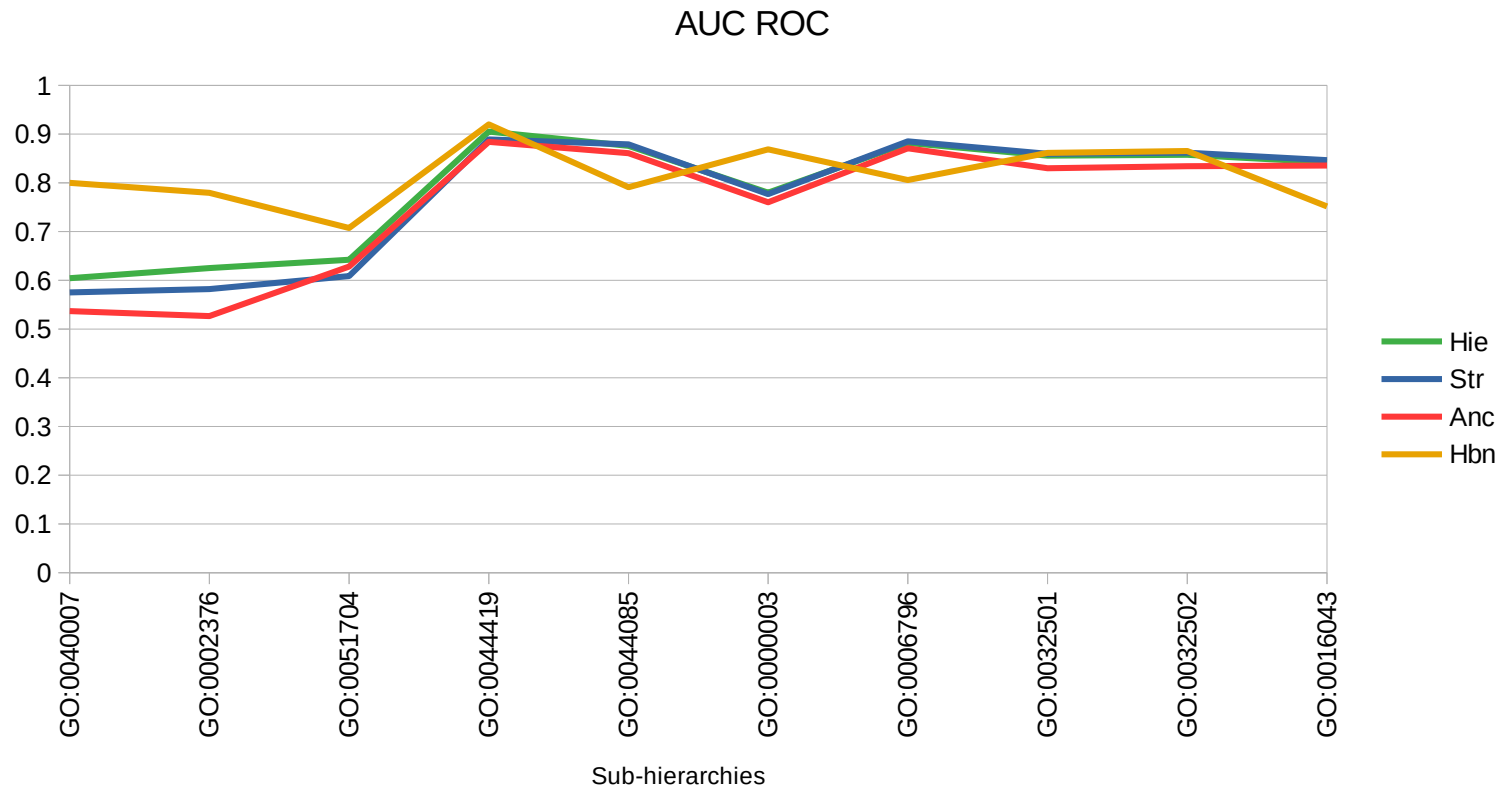
Hie



Anc



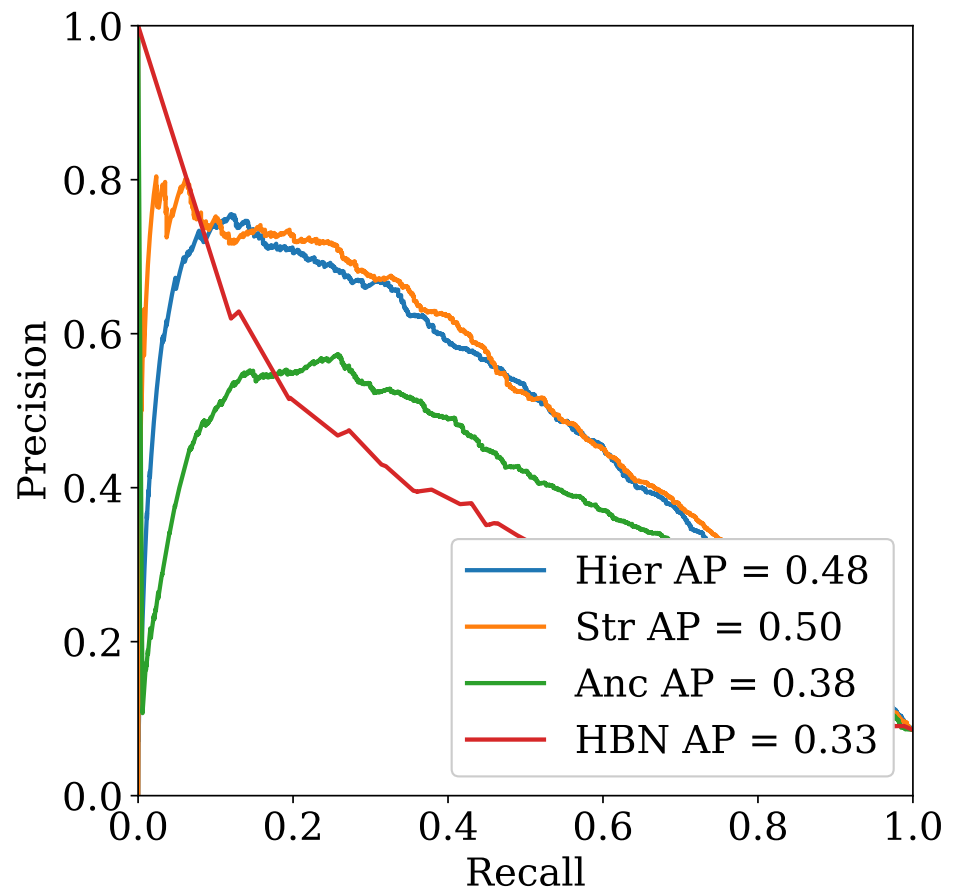
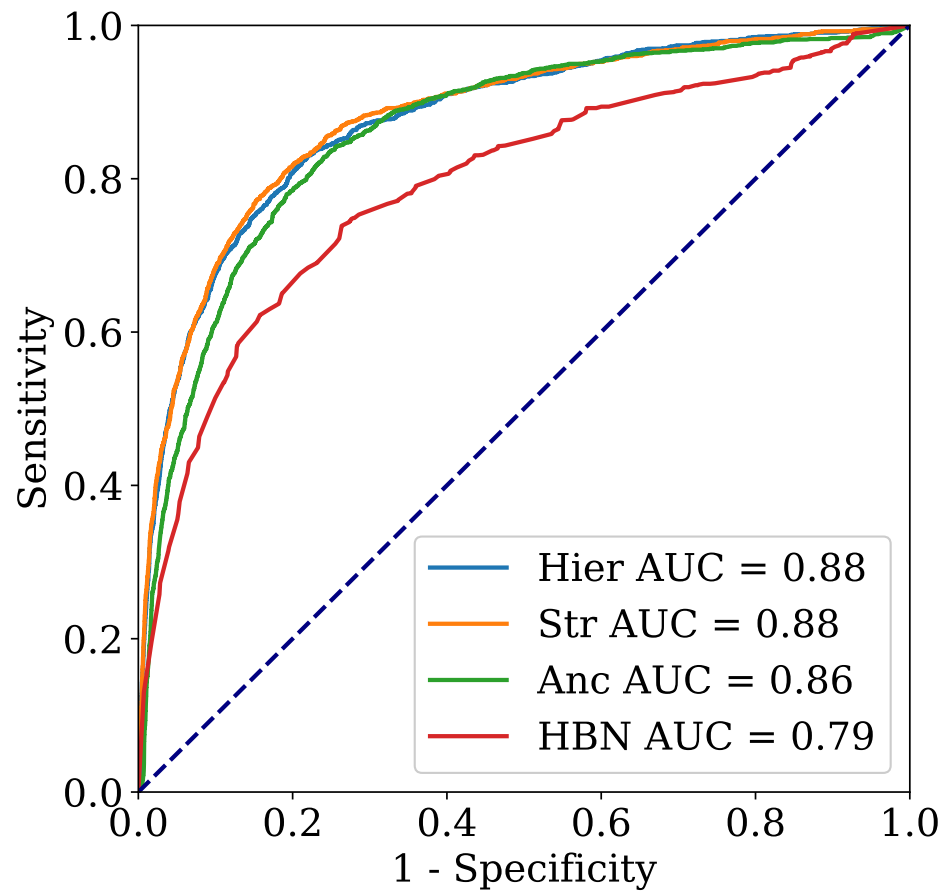
Second trial



The second trial include new features that captures the information of the neighborhood of a gene in terms of the function and its parent.

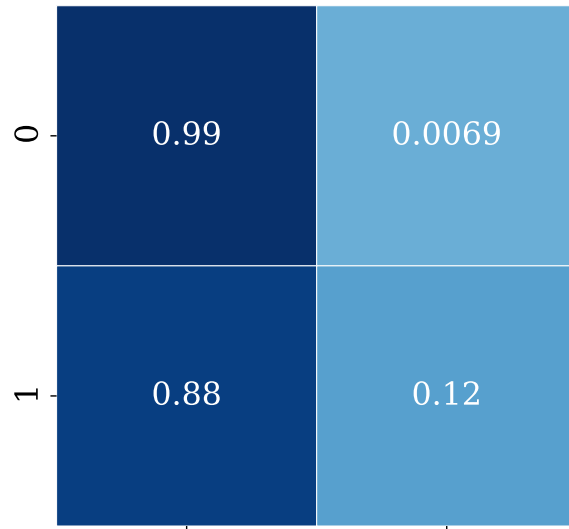
Only the **Anc** dataset include the new features.

HBN vs. ML: GO:0044085

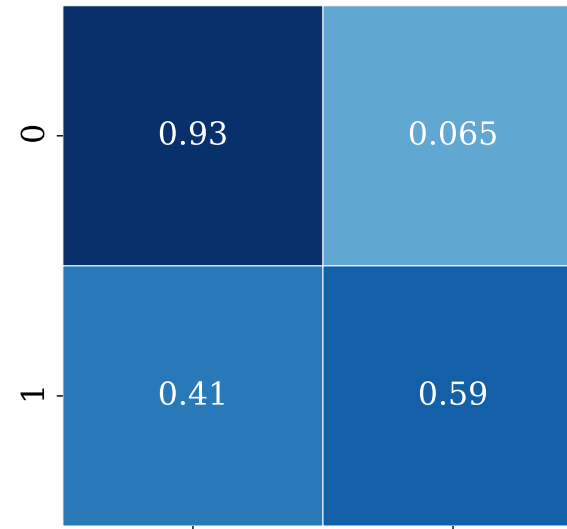


HBN vs. ML: GO:0044085

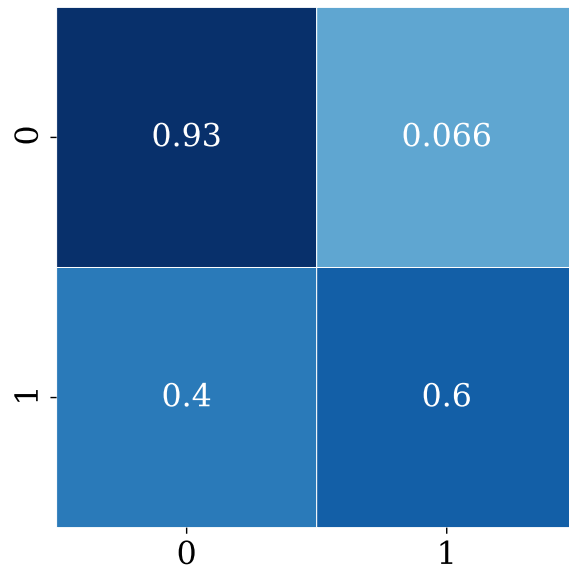
HBN



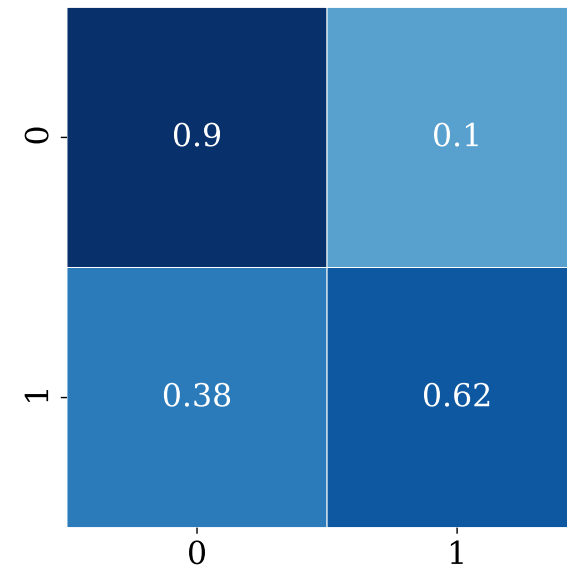
Str



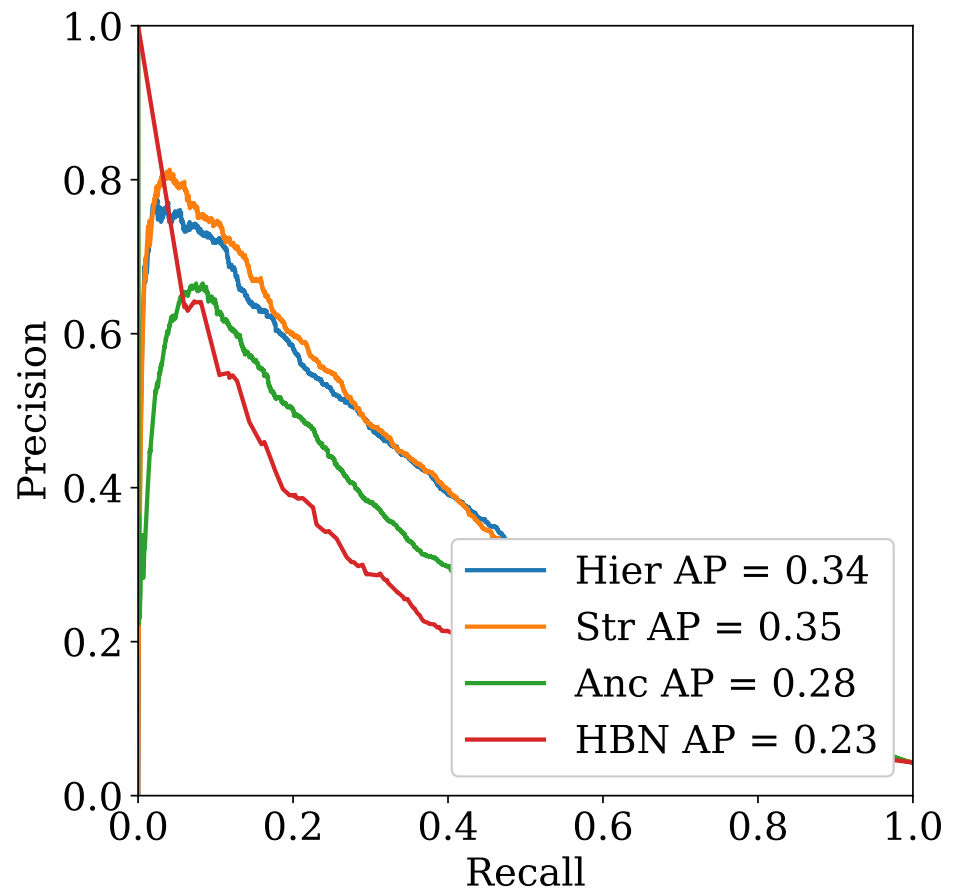
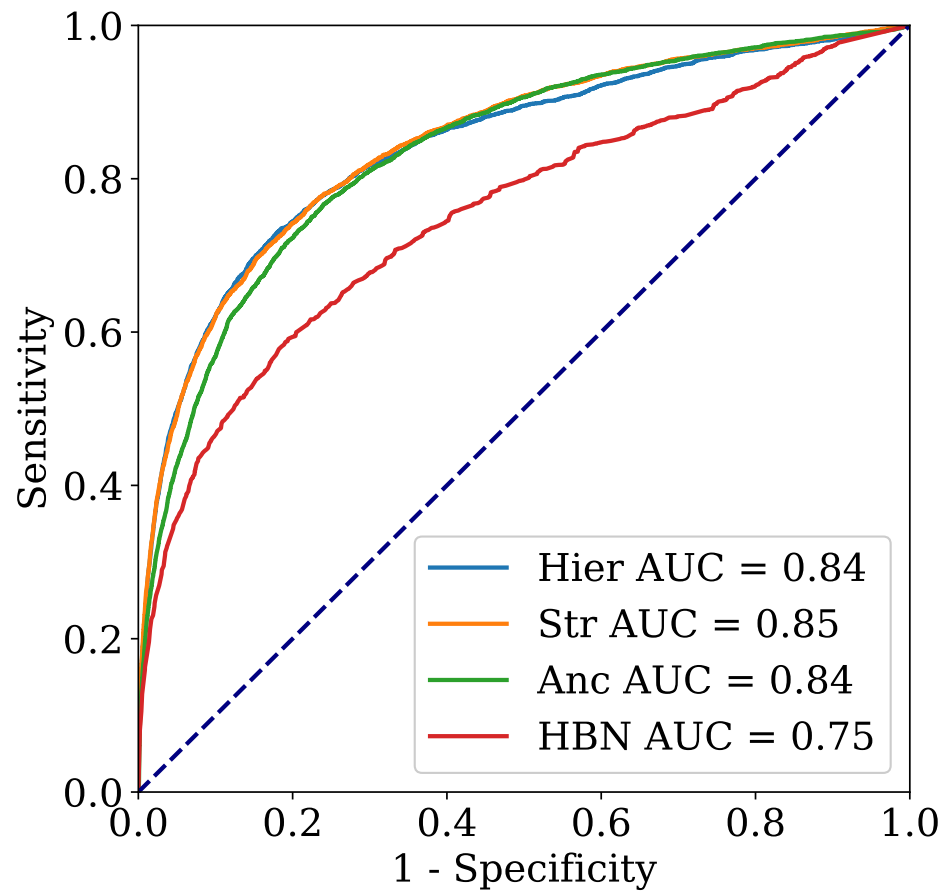
Hie



Anc

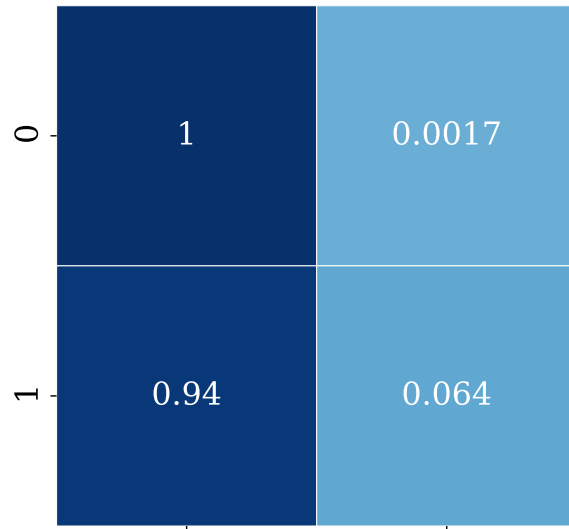


HBN vs. ML: GO:0016043

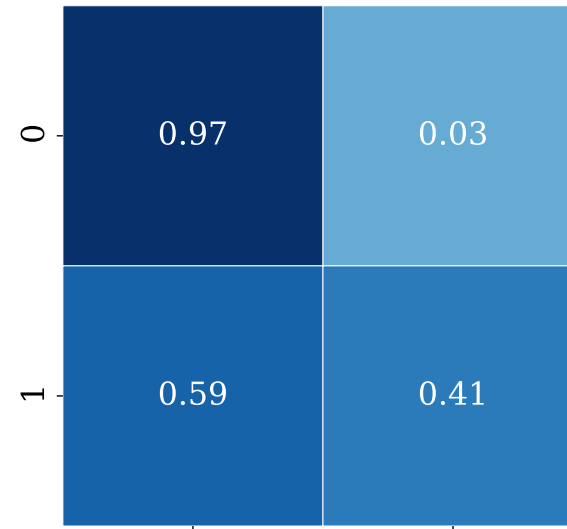


HBN vs. ML: GO:0016043

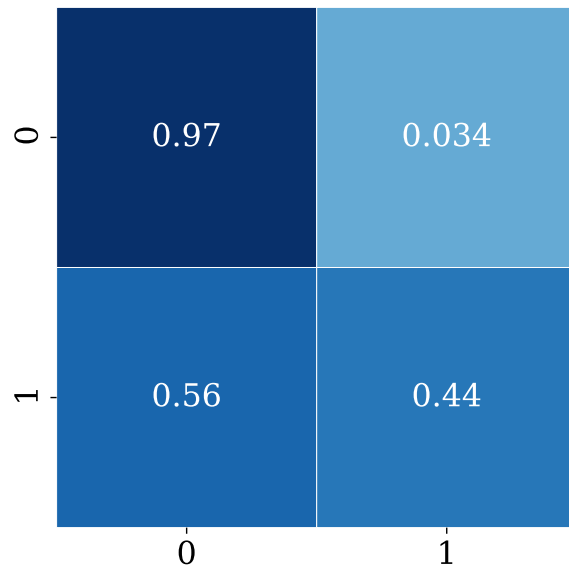
HBN



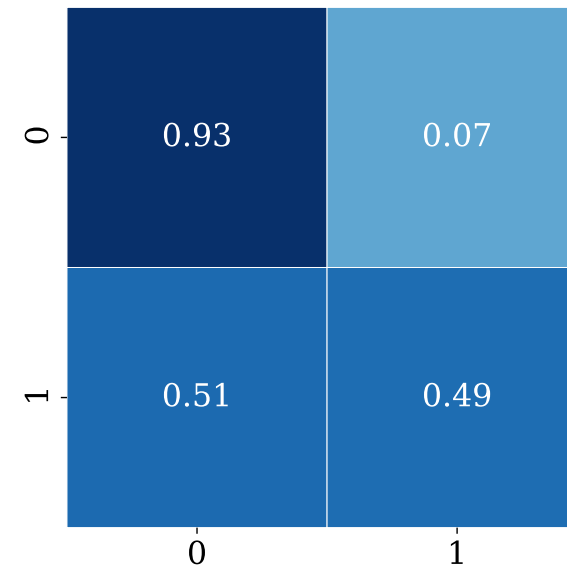
Str



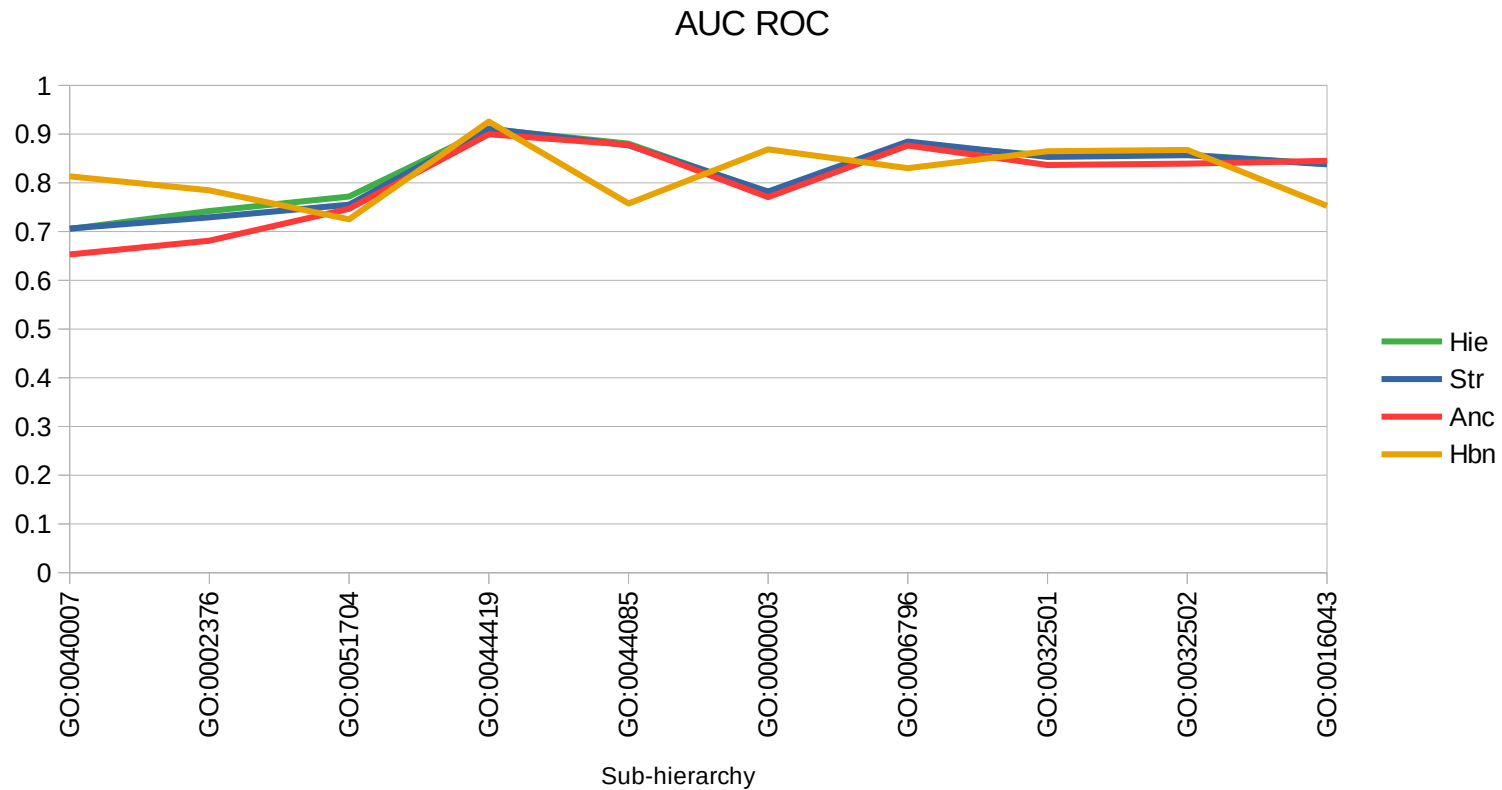
Hie



Anc

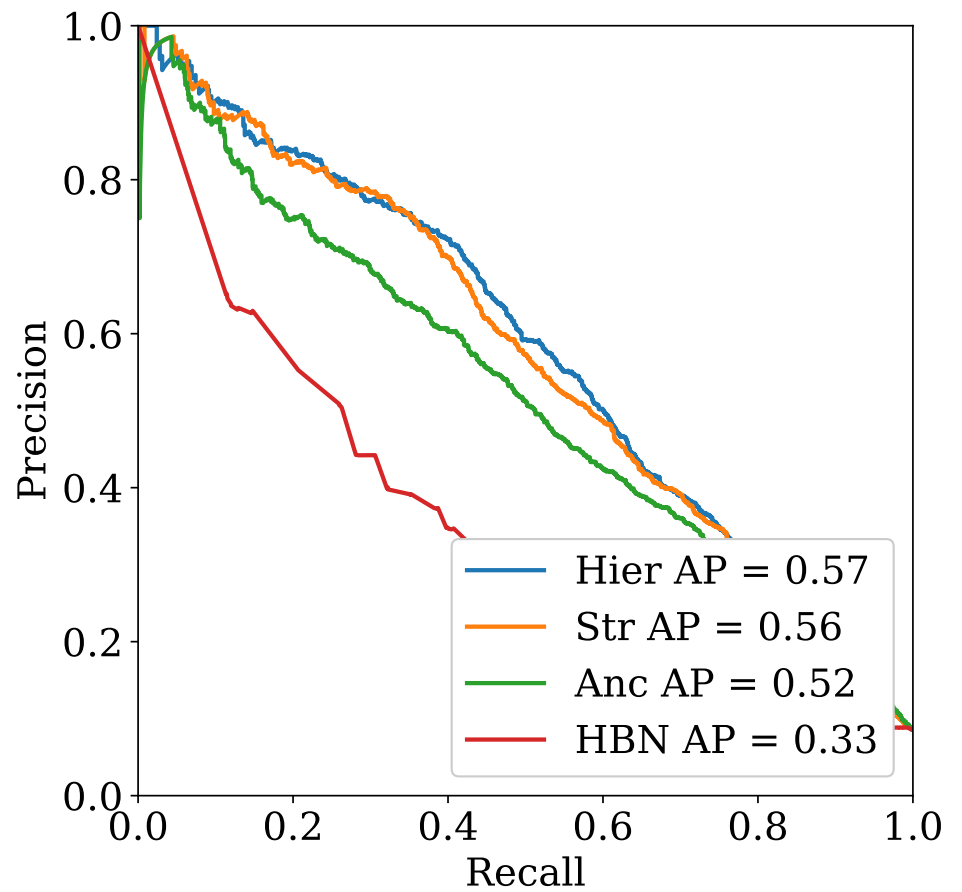
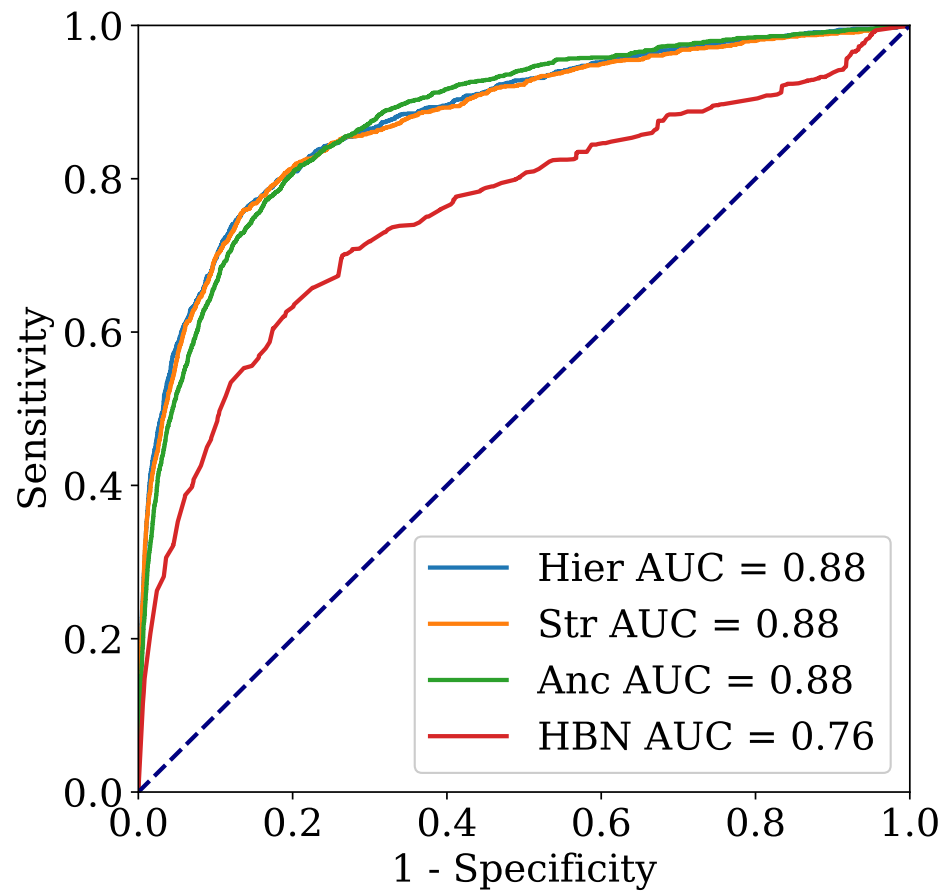


Third trial



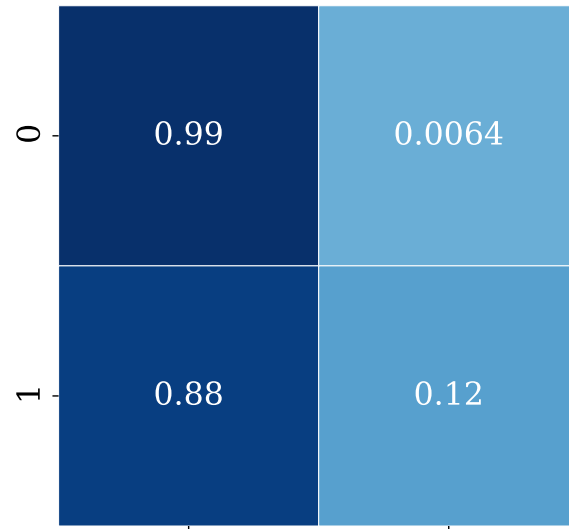
The third trial include the probabilistic features in all datasets.

HBN vs. ML: GO:0044085

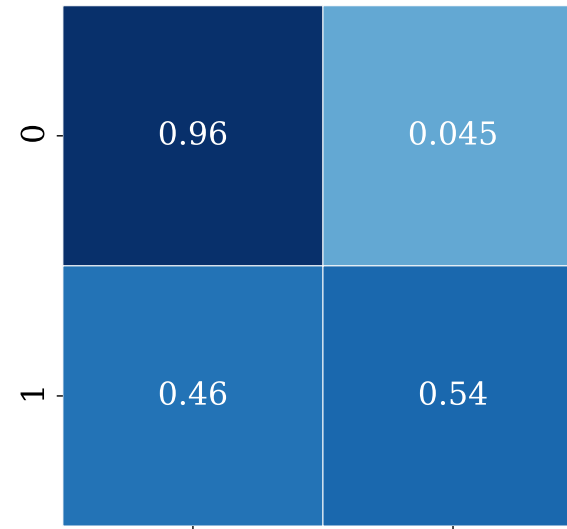


HBN vs. ML: GO:0044085

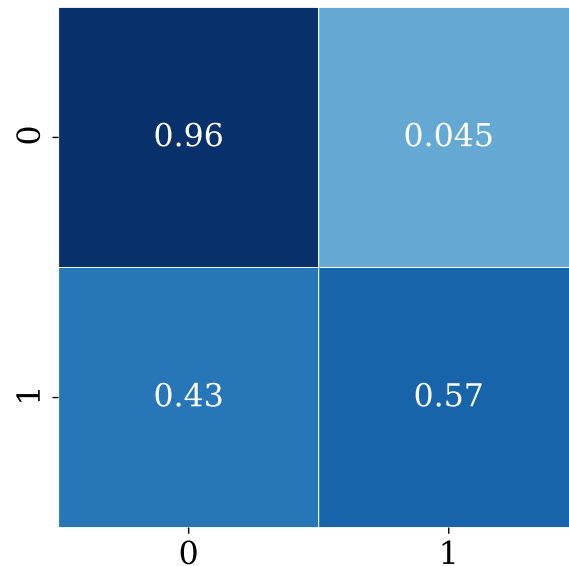
HBN



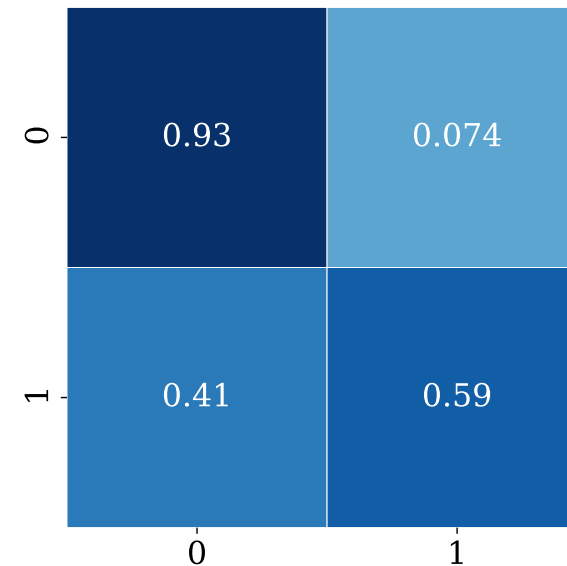
Str



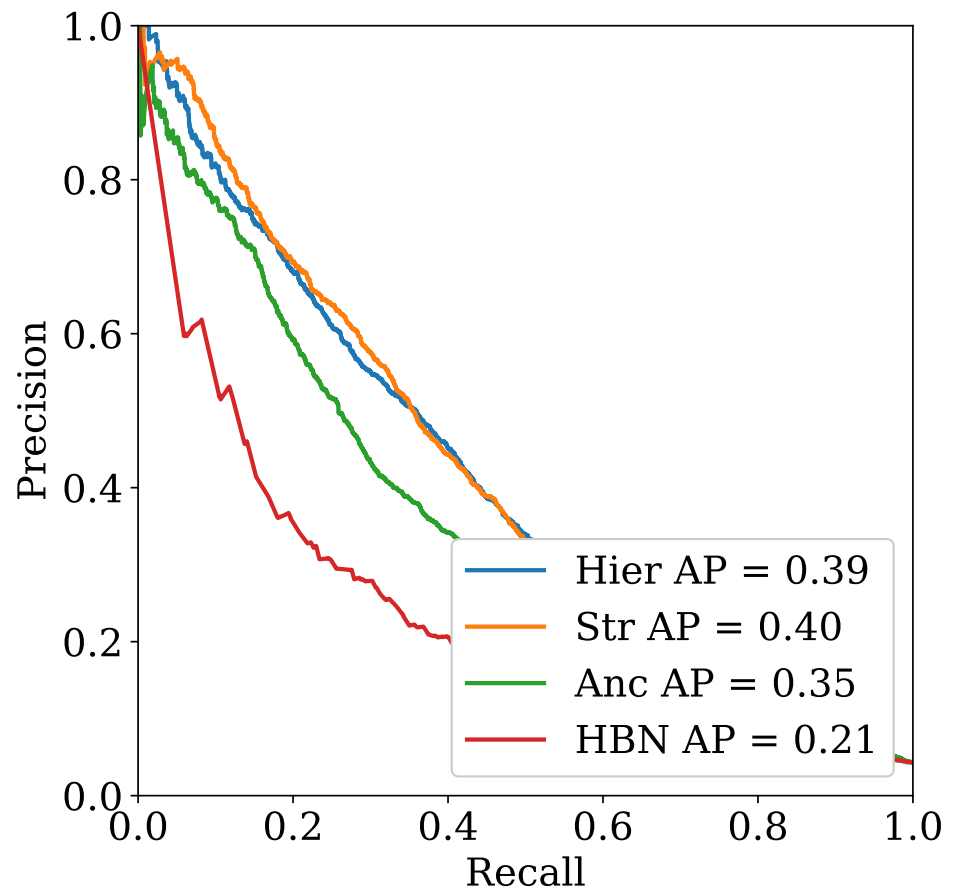
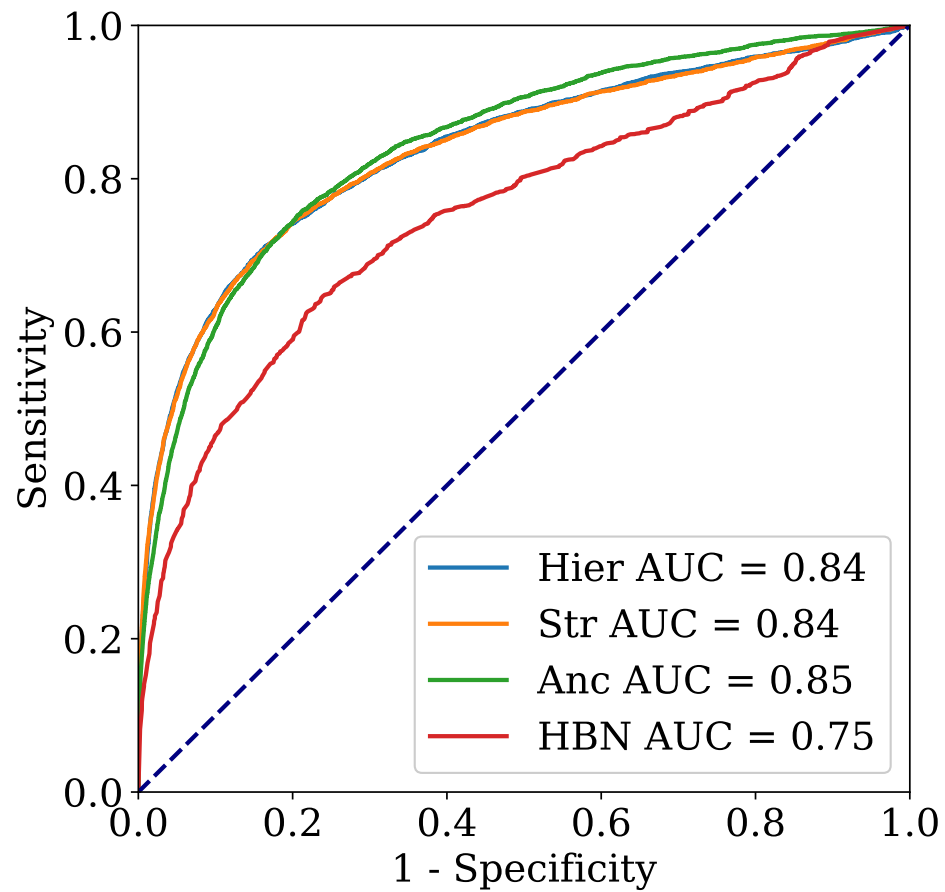
Hie



Anc

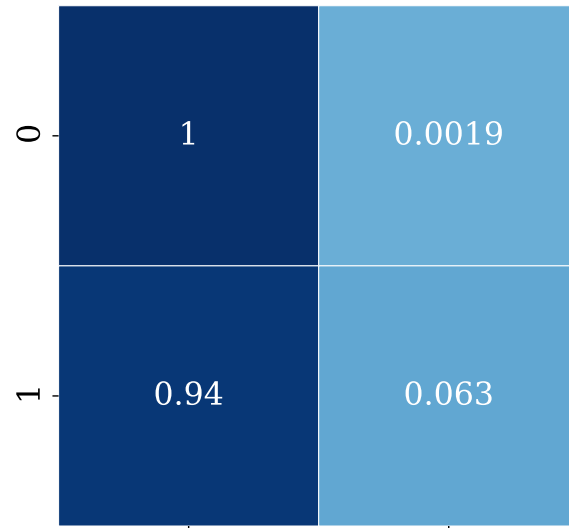


HBN vs. ML: GO:0016043

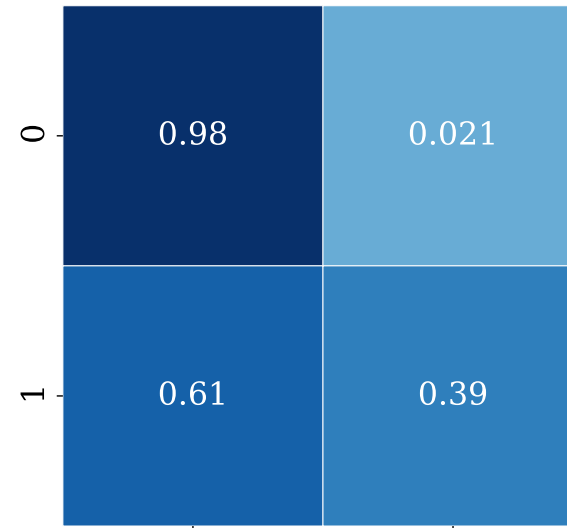


HBN vs. ML: GO:0016043

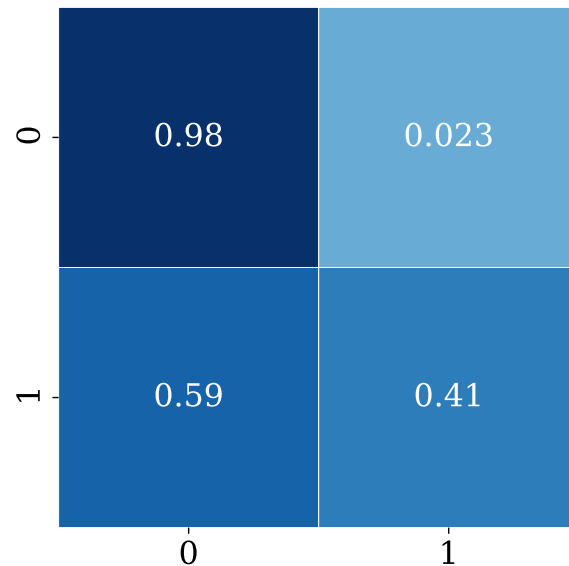
HBN



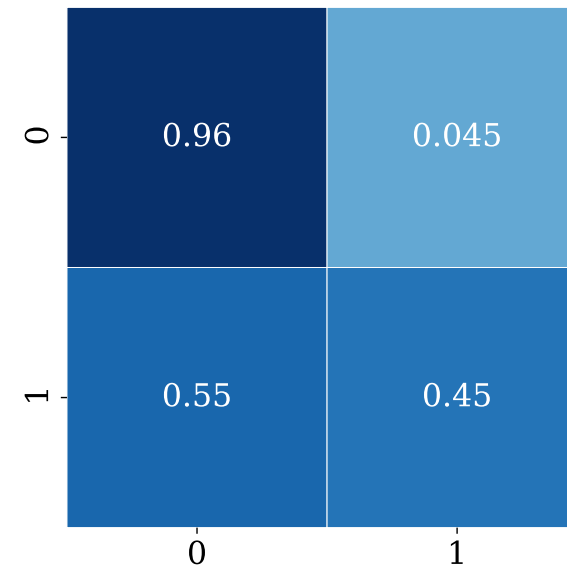
Str



Hie

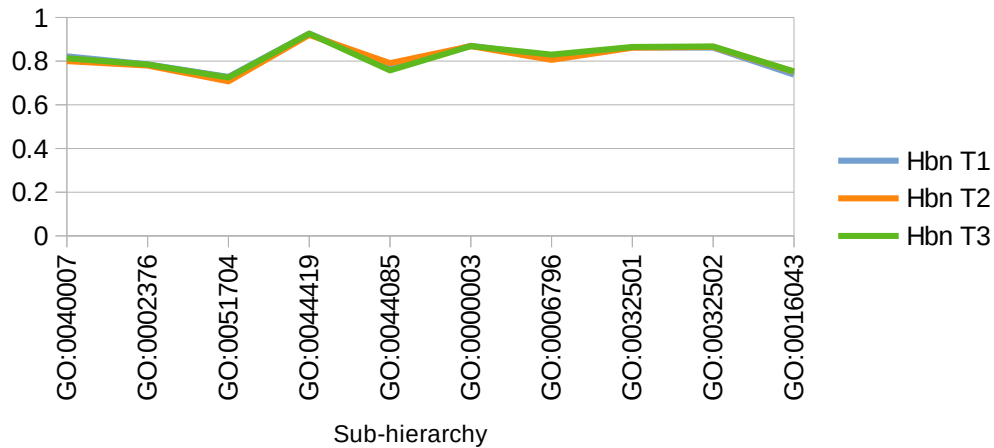


Anc

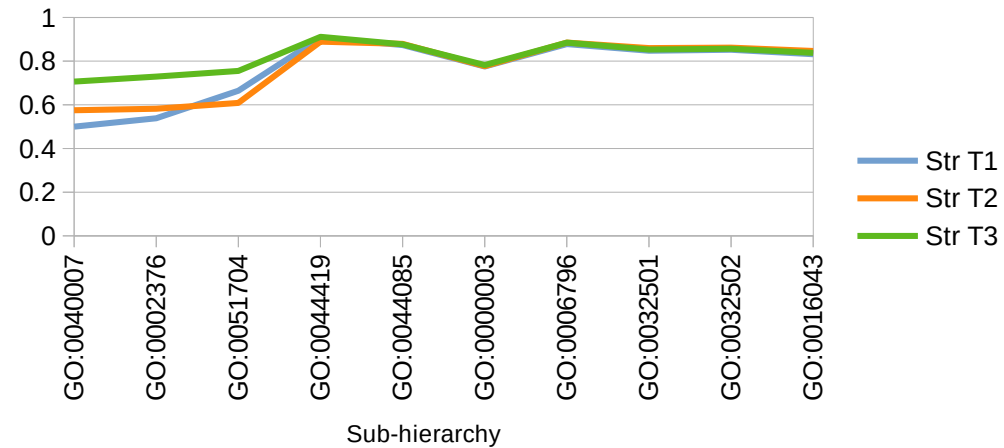


Comparison between trials

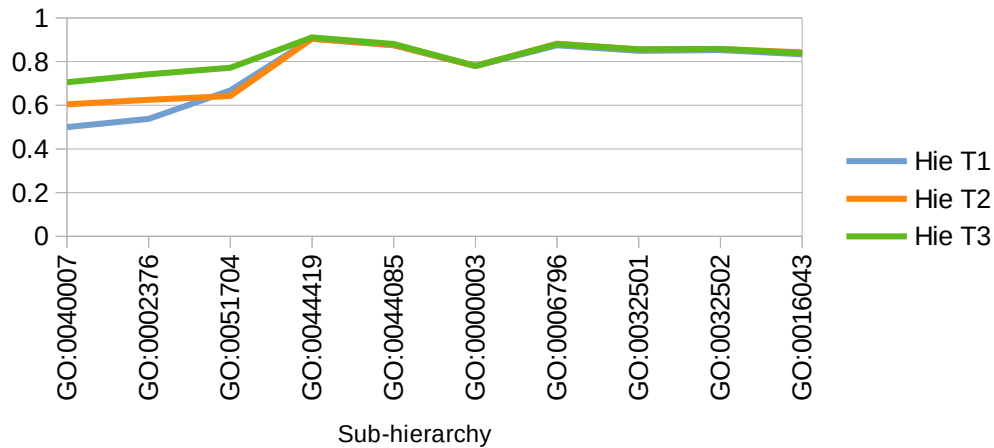
HBN



Structural



Hierarchical



Ancestral

