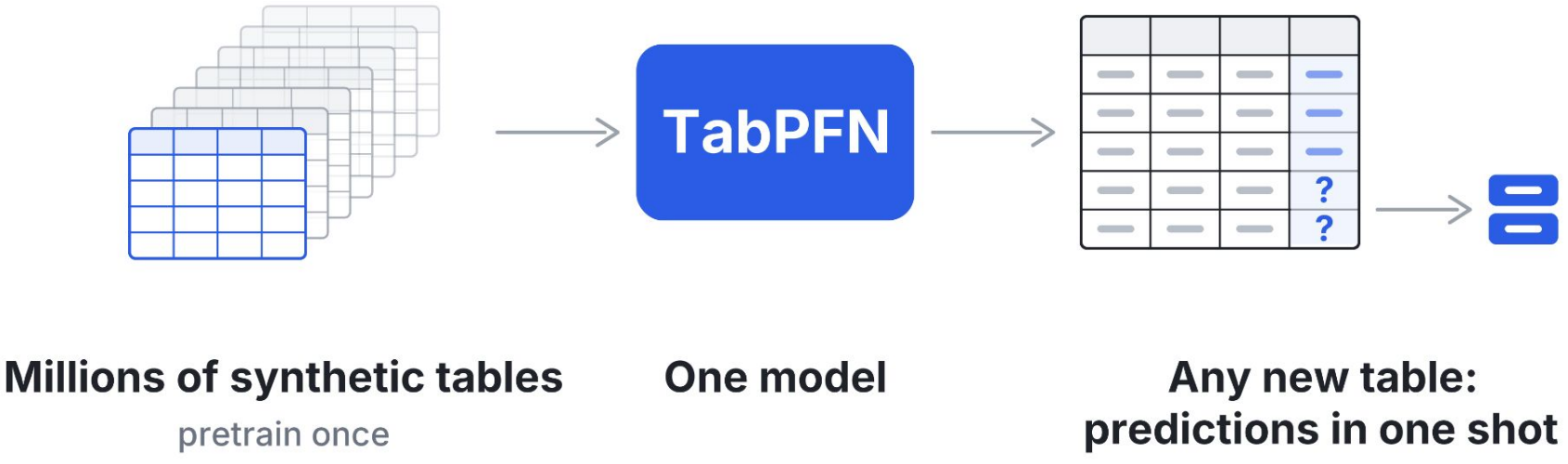
MulTaBench

BENCHMARKING MULTIMODAL TABULAR LEARNING

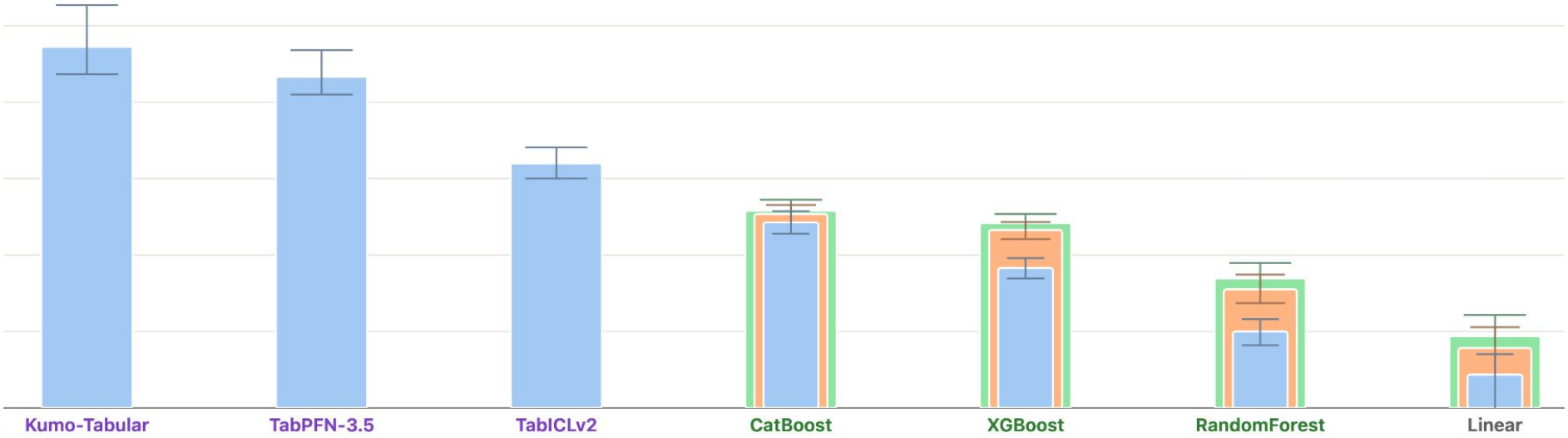
Benchmarking Multimodal Tabular Learning with Text and Image



TabPFN: A New Paradigm for Tabular Data



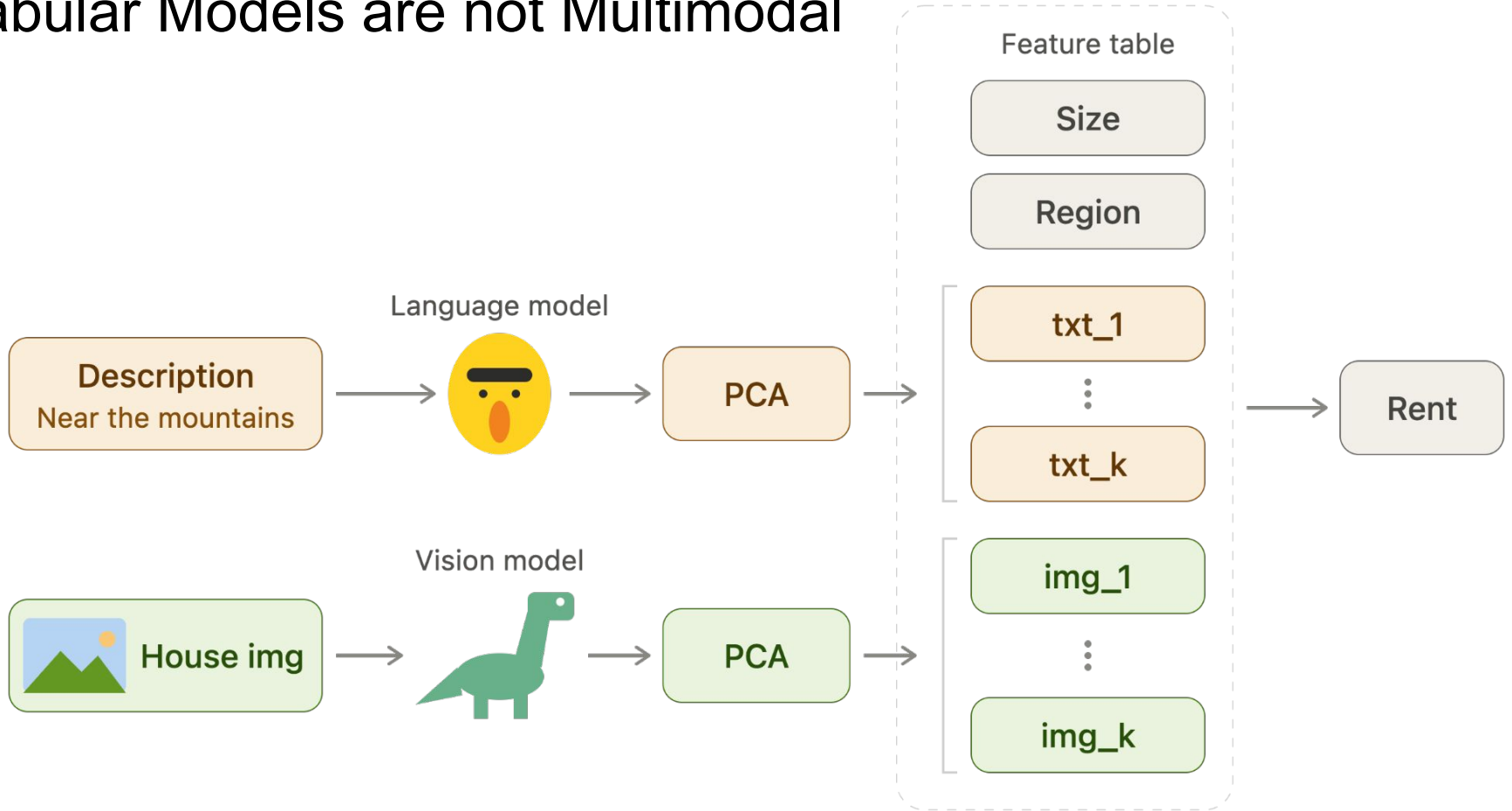
Tabular Foundation Models rule TabArena



Size	Region	House Description	House Img
45	North	Near the mountains	
62	South	Pearl of the desert	
38	Center	Near great restaurants	
55	South	One hour from beach	

Rent
No
Yes
No
No

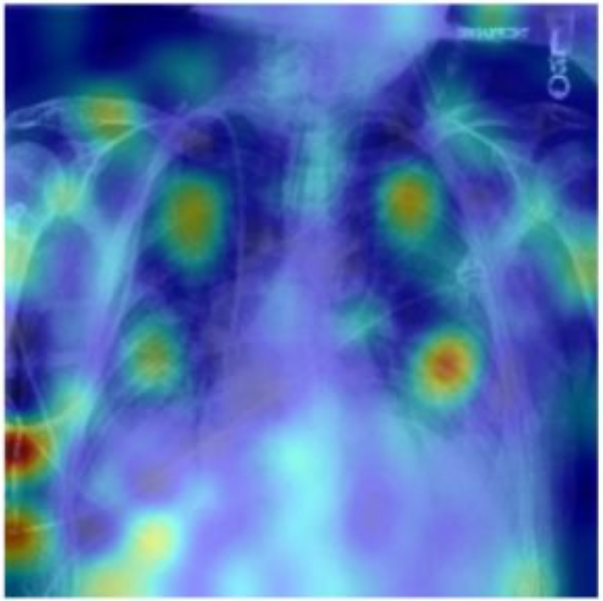
Tabular Models are not Multimodal



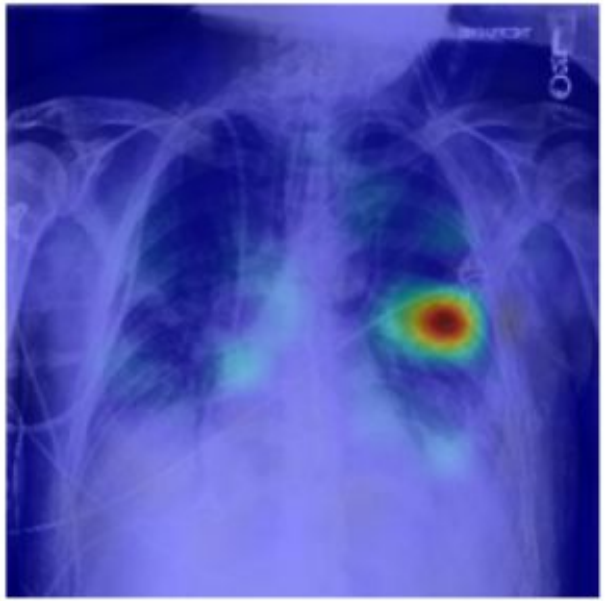
Image



Frozen



Target-Aware

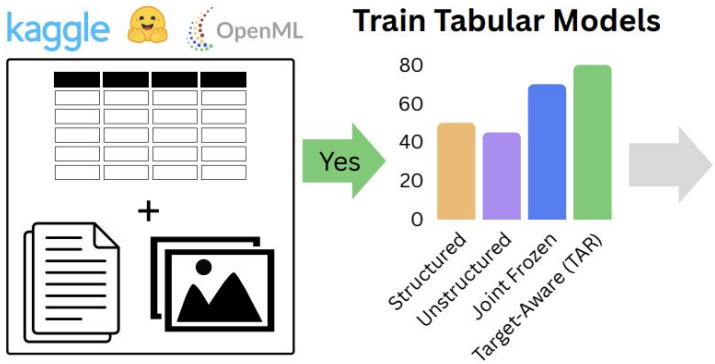


How to build a Multimodal Tabular Benchmark?

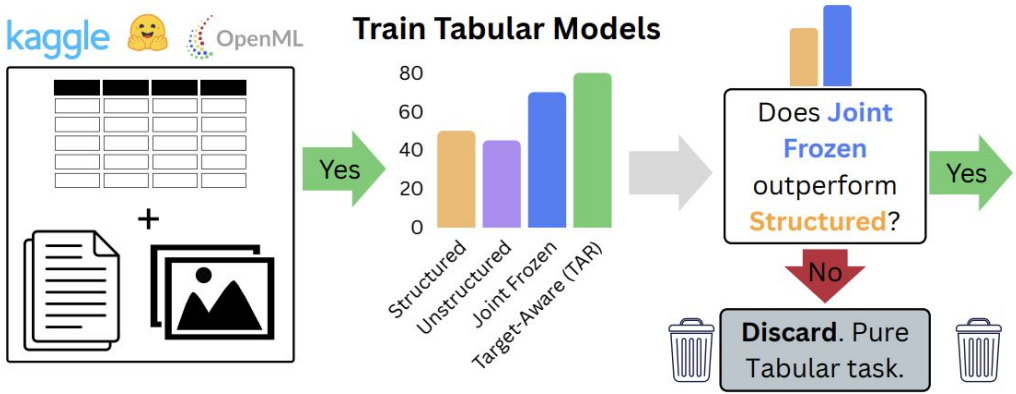
🤗 Joint Signal: both modalities should improve predictive performance.

🎯 Task-awareness: adapting the encoder should boost performance.

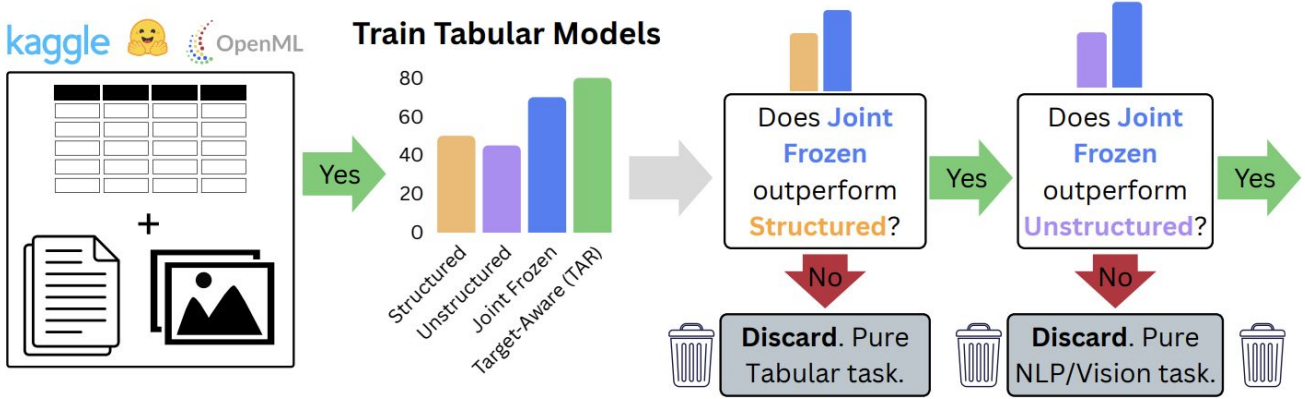
Our Curation Algorithm: The 4 Conditions



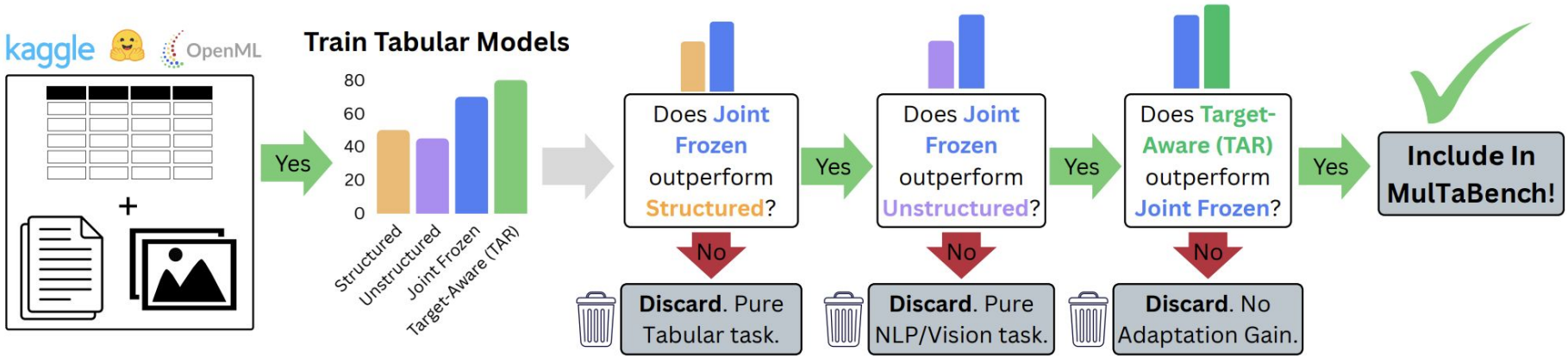
Our Curation Algorithm: Joint Signal



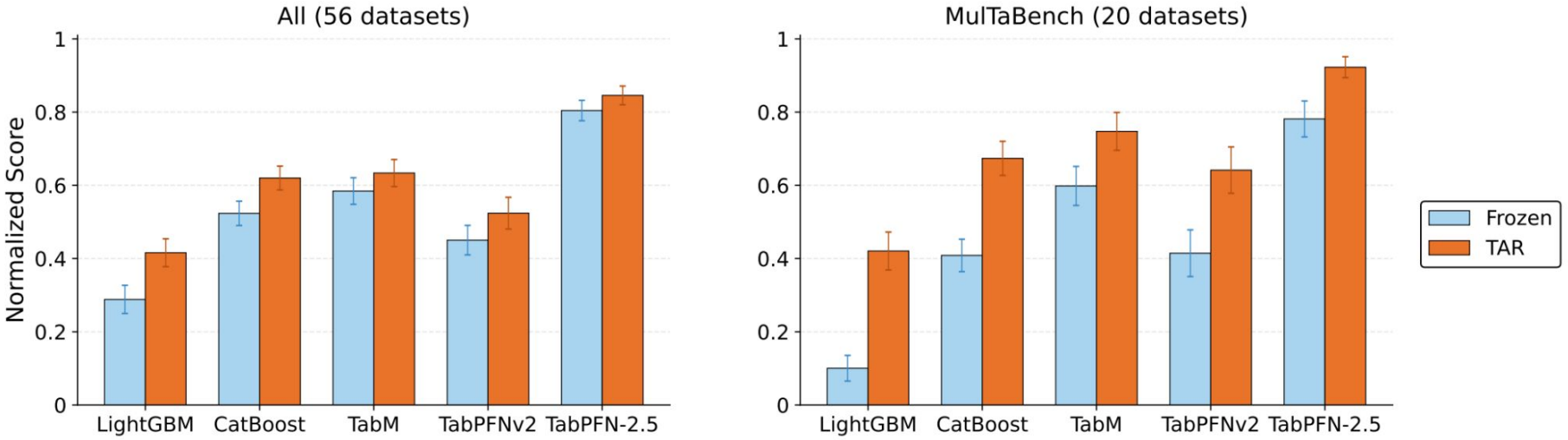
Our Curation Algorithm: Joint Signal

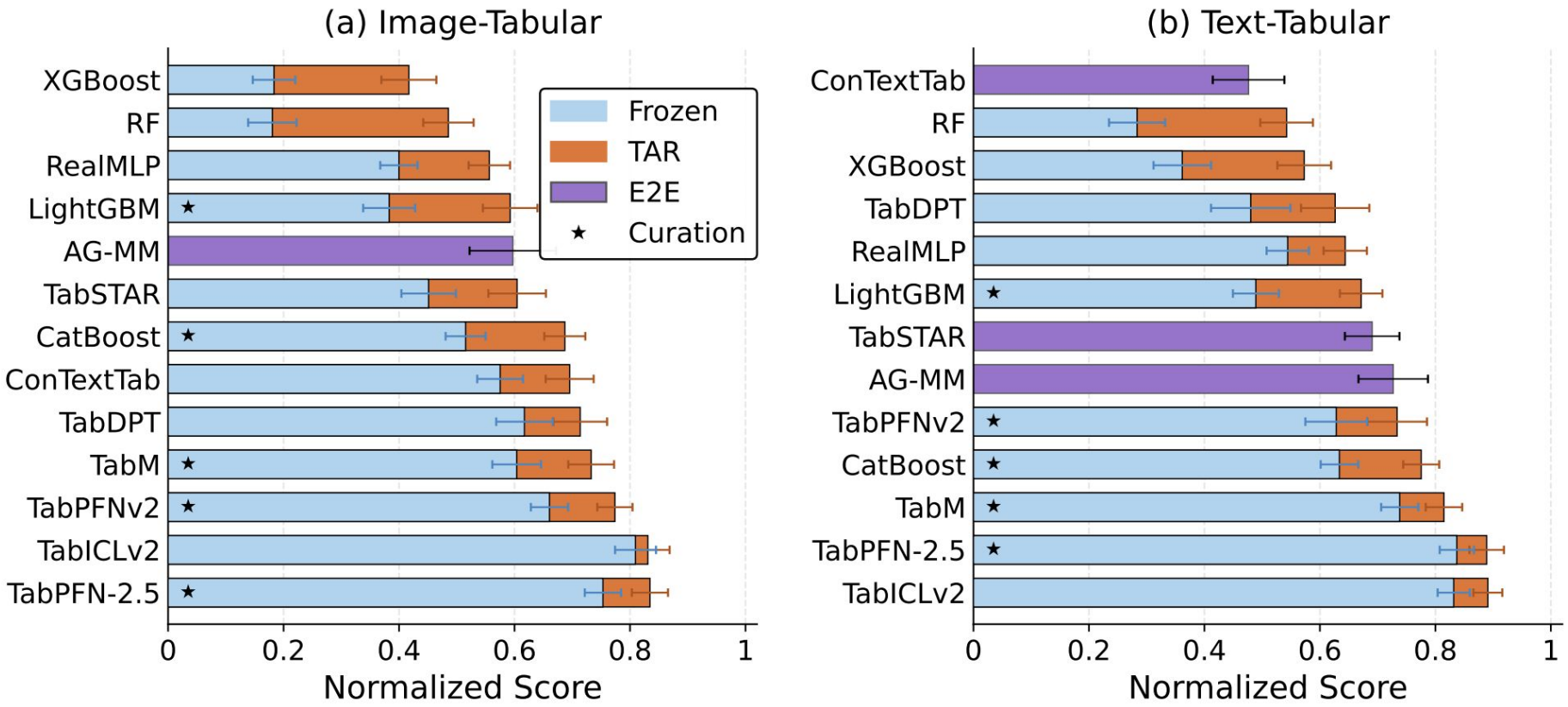


Our Curation Algorithm: Task-Awareness

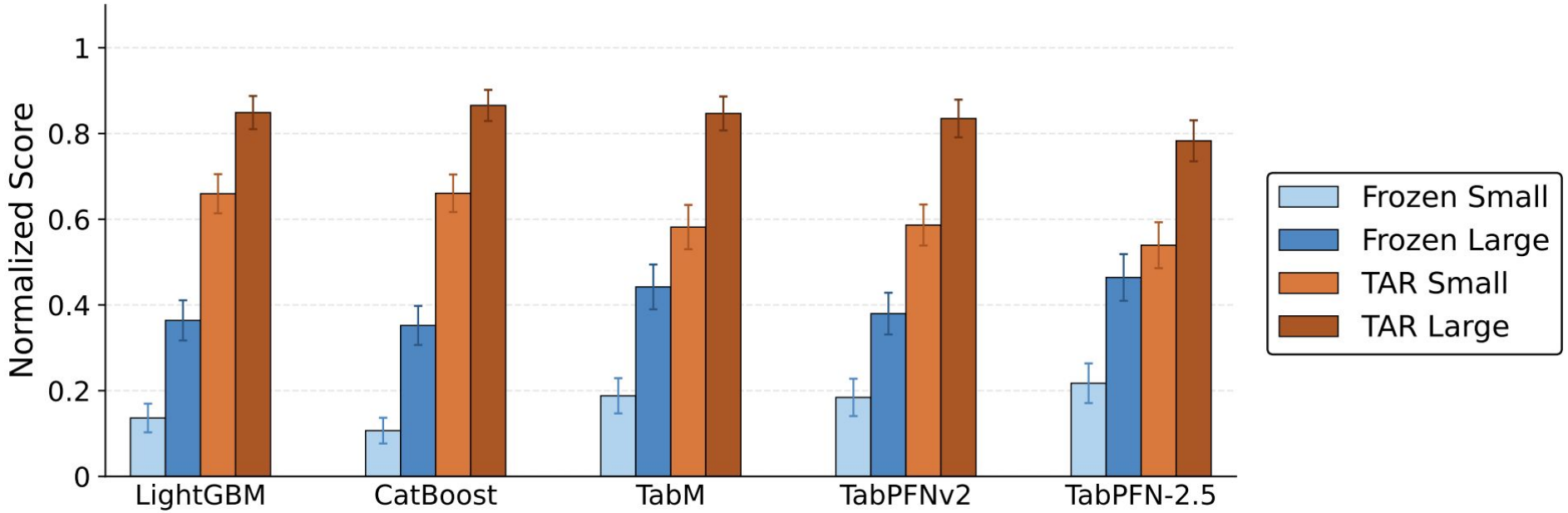


Task-Awareness is a property of the dataset




(a) Image-Tabular
XGBoost
RF
RealMLP
LightGBM
AG-MM
TabSTAR
CatBoost
ConTextTab
TabDPT
TabM
TabPFNv2
TabICLv2
TabPFN-2.5
Frozen
TAR
E2E
Curation
0
0.2
0.4
0.6
0.8
1
Normalized Score
(b) Text-Tabular
ConTextTab
RF
XGBoost
TabDPT
RealMLP
LightGBM
TabSTAR
AG-MM
TabPFNv2
CatBoost
TabM
TabPFN-2.5
TabICLv2
0
0.2
0.4
0.6
0.8
1
Normalized Score

Robustness Analysis: Larger Encoders



Towards Multimodal Tabular Foundation Models

