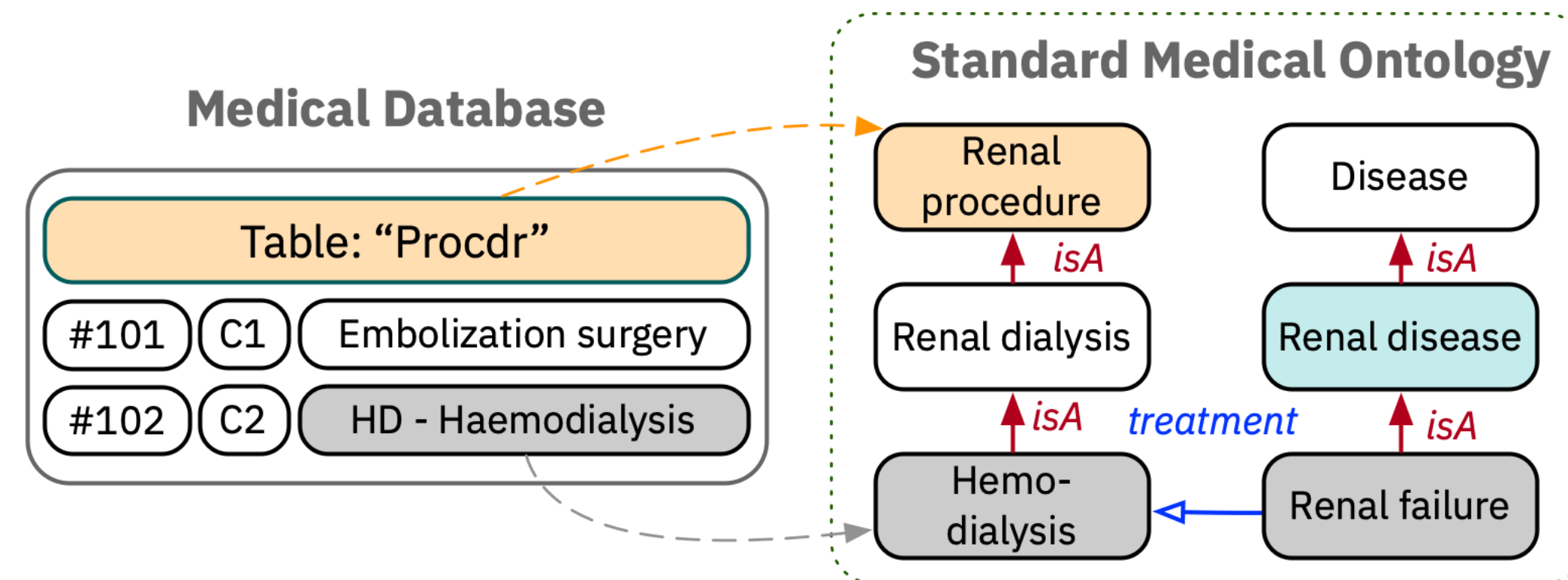


DATA-TO-ONTOLOGY MATCHING IN MEDICAL DOMAIN

Increasingly large-scale medical databases, in need of automatic AI-assisted analysis.
 Core task: Mapping database schema/tables to standard ontologies (for standardization)
 Existing methods focus on ontology matching, assuming ontologies are available for matching
 Effective data-to-ontology matching techniques



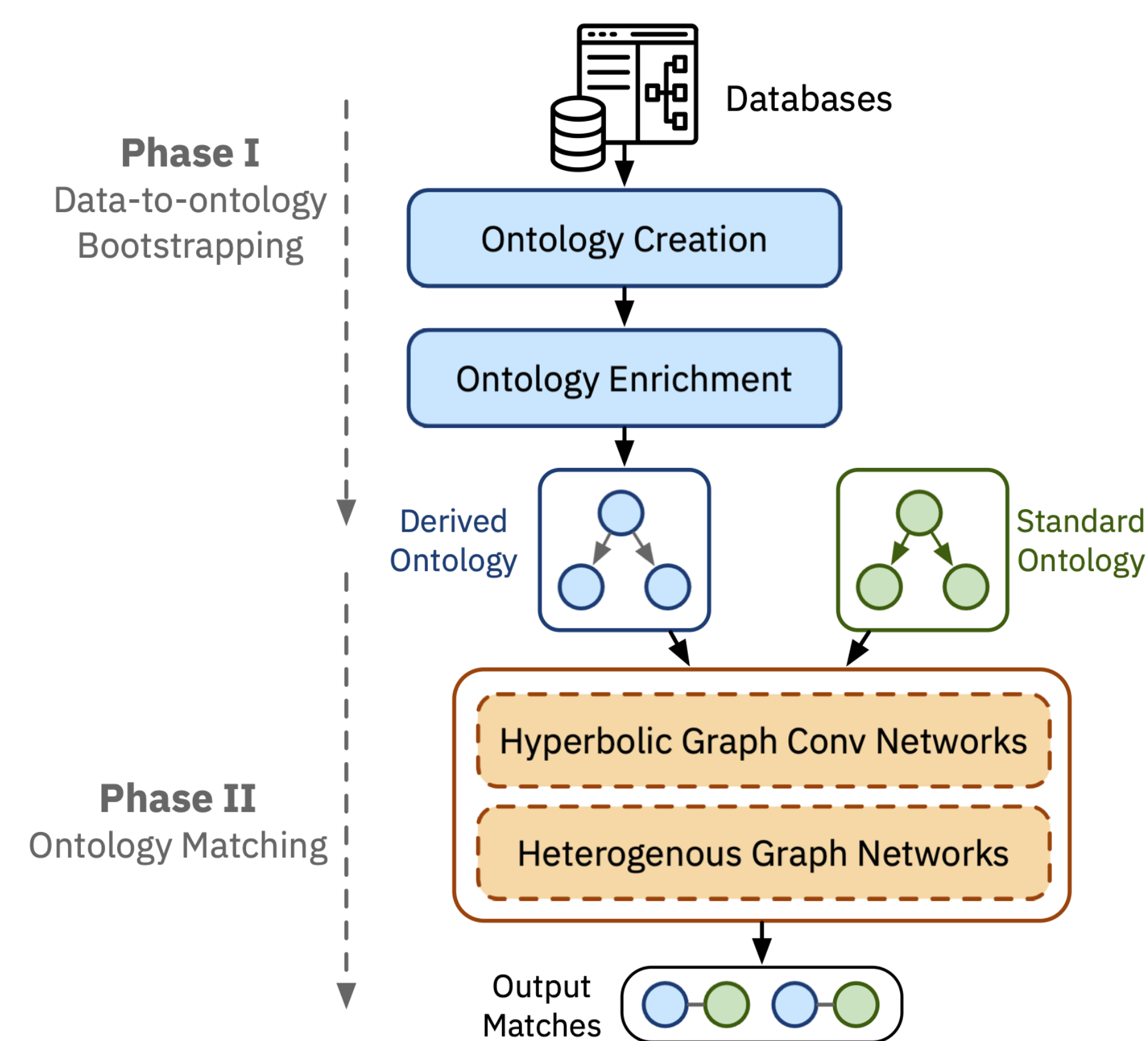
TRAINING

Matching: MLP, input as pairs of concept embeddings from O_1 and O_2
Training Loss: Matching loss + Graph decoders
 $L = L^M + \frac{1}{2} L_{O_1}^{HYP} + \frac{1}{2} L_{O_2}^{HYP} + \frac{1}{2} L_{O_1}^{HET} + \frac{1}{2} L_{O_2}^{HET}$

DATASET & BASELINES

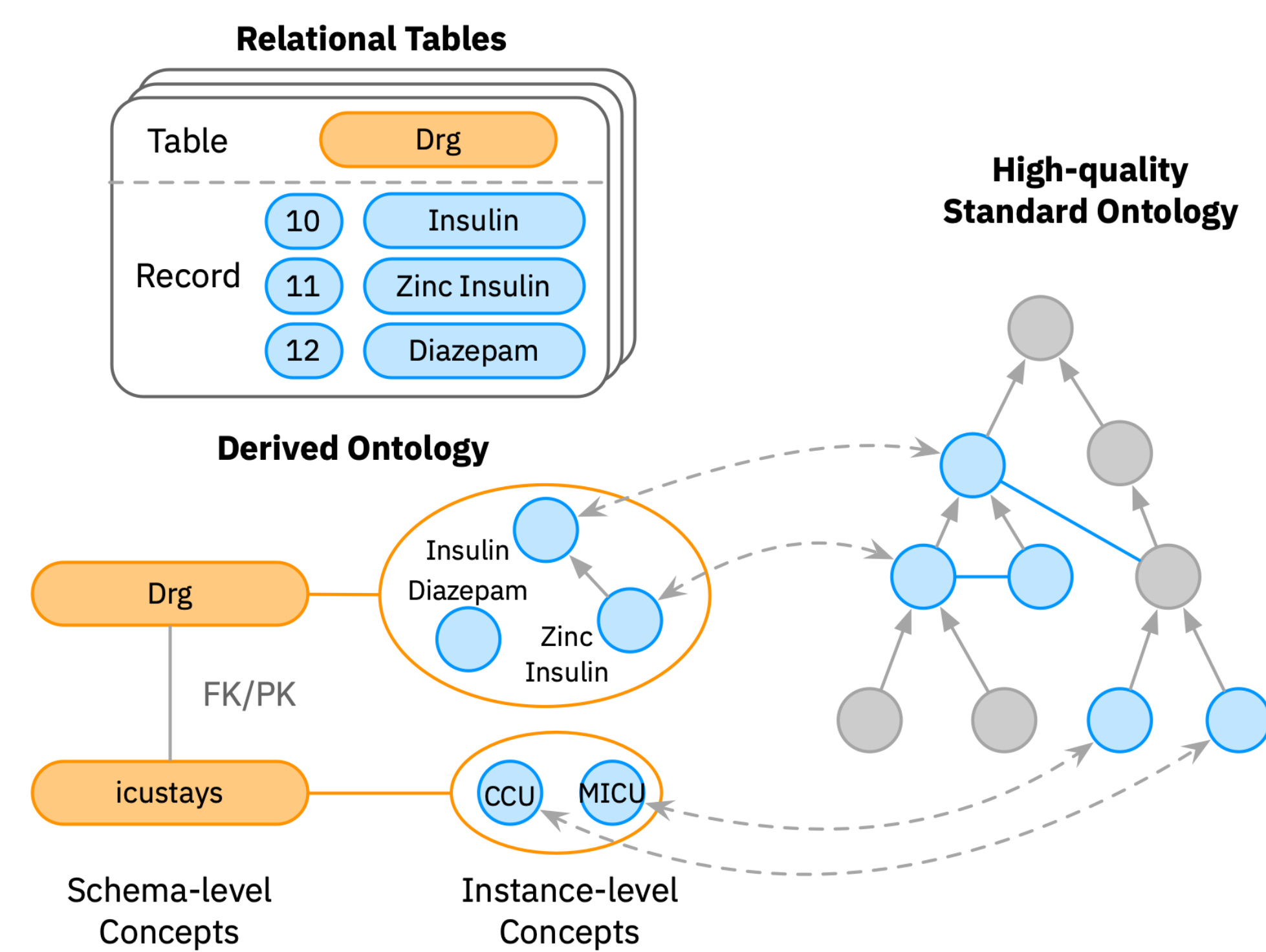
Medical Databases: MIMIC-III, IBM Micromedex (MDX)
Medical Ontologies: FMA, NCI, SNOMED-CT
Baselines: AML, LogMap, RDGCN, etc.

MEDTO: SYSTEM ARCHITECTURE

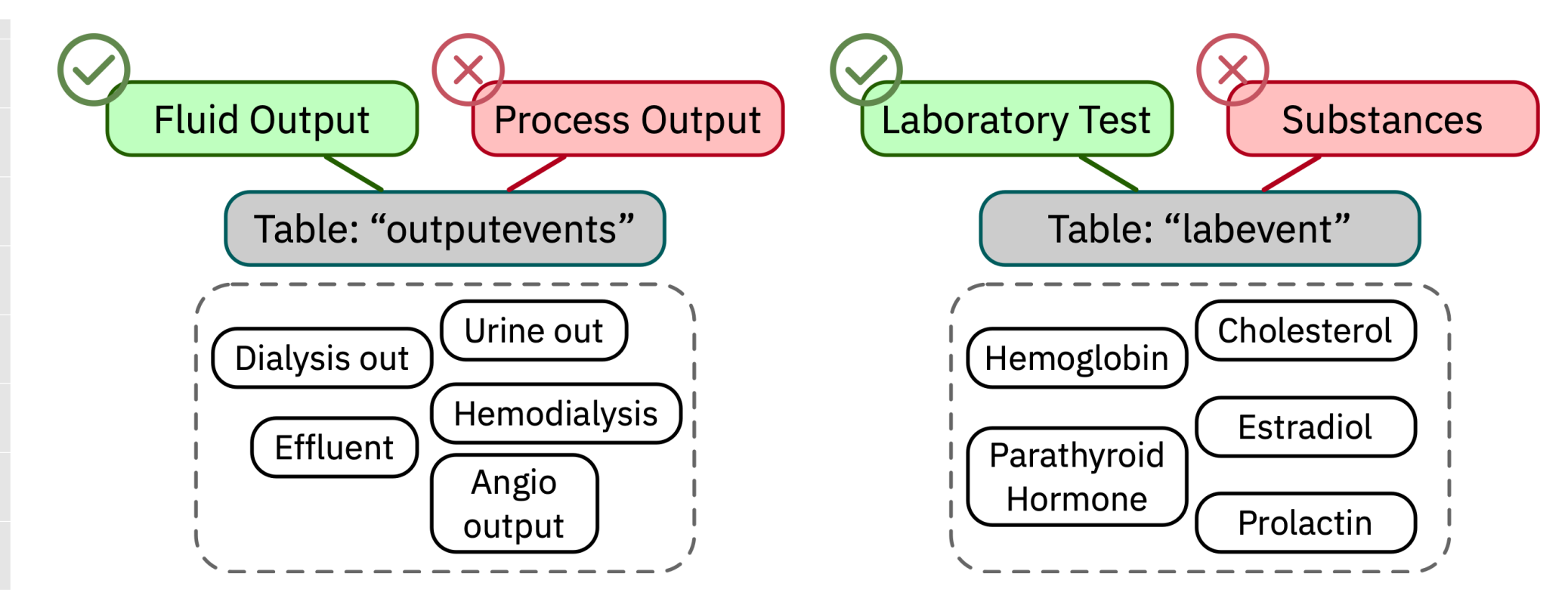
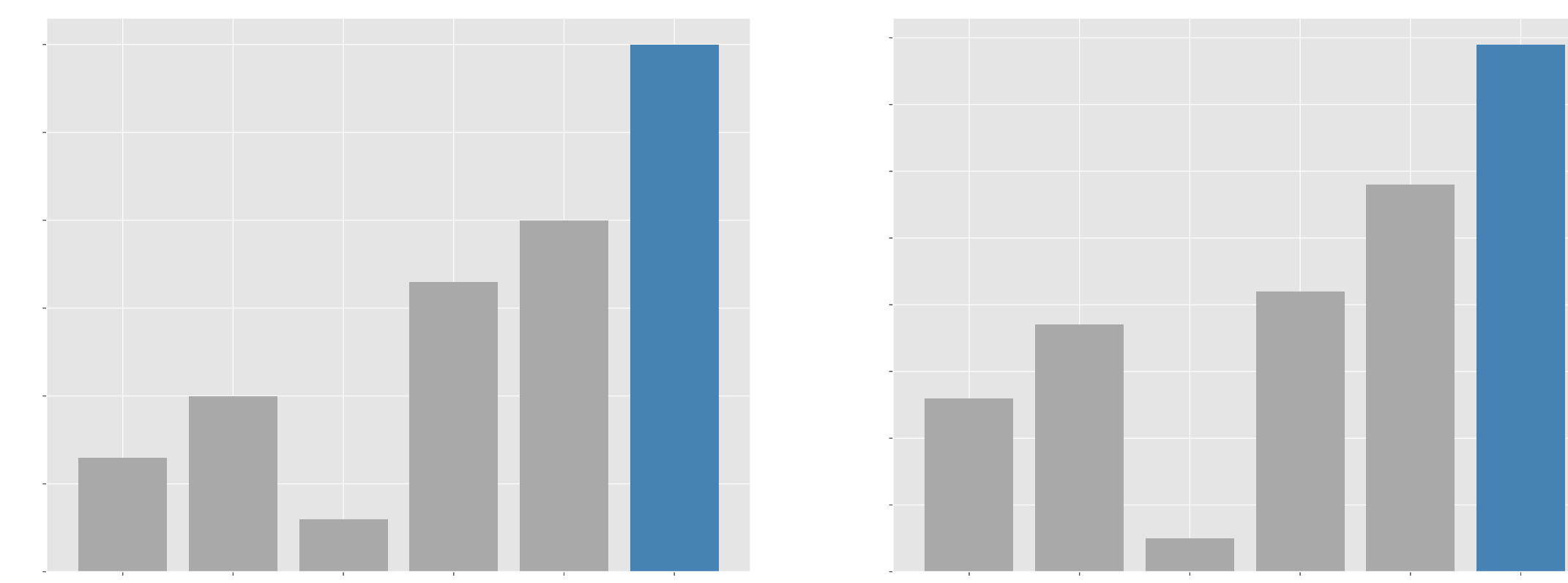


PHASE 1: BOOTSTRAPPING

A two-step process from database tables: (1) Ontology creation; (2) Ontology Enrichment.



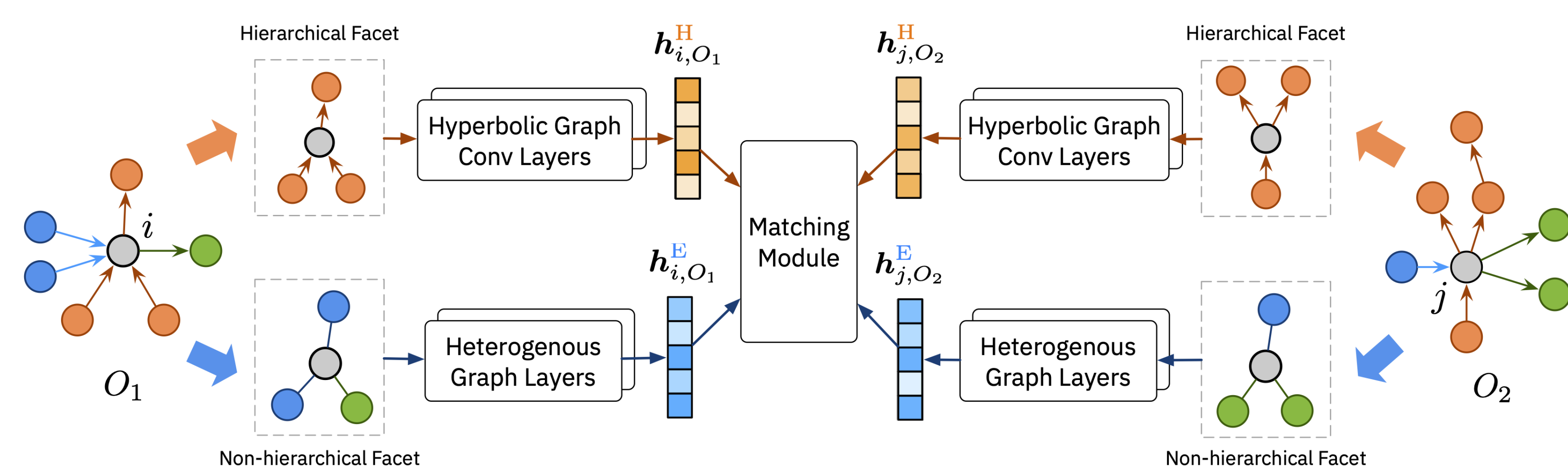
EXPERIMENTS AND CASE STUDY ON MEDICAL DATABASE



EXPERIMENTS ON ONTOLOGY MATCHING

Model Groups	Datasets Metrics	FMA-NCI		FMA-SNOMED		NCI-SNOMED	
		F1	MRR	F1	MRR	F1	MRR
Rule-Based	AML	0.920	-	0.806	-	0.810	-
	LogMap	0.905	-	0.819	-	0.805	-
GNN-based Entity Alignment	MTransE	0.633	0.416	0.490	0.372	0.304	0.349
	GCN-Align	0.798	0.561	0.746	0.526	0.760	0.467
	RDGCN	0.849	0.761	0.786	0.683	0.816	0.679
Ours	MEDTO	0.908	0.783	0.813	0.690	0.849	0.704

PHASE 2: ONTOLOGY MATCHING (GRAPH ENCODERS)

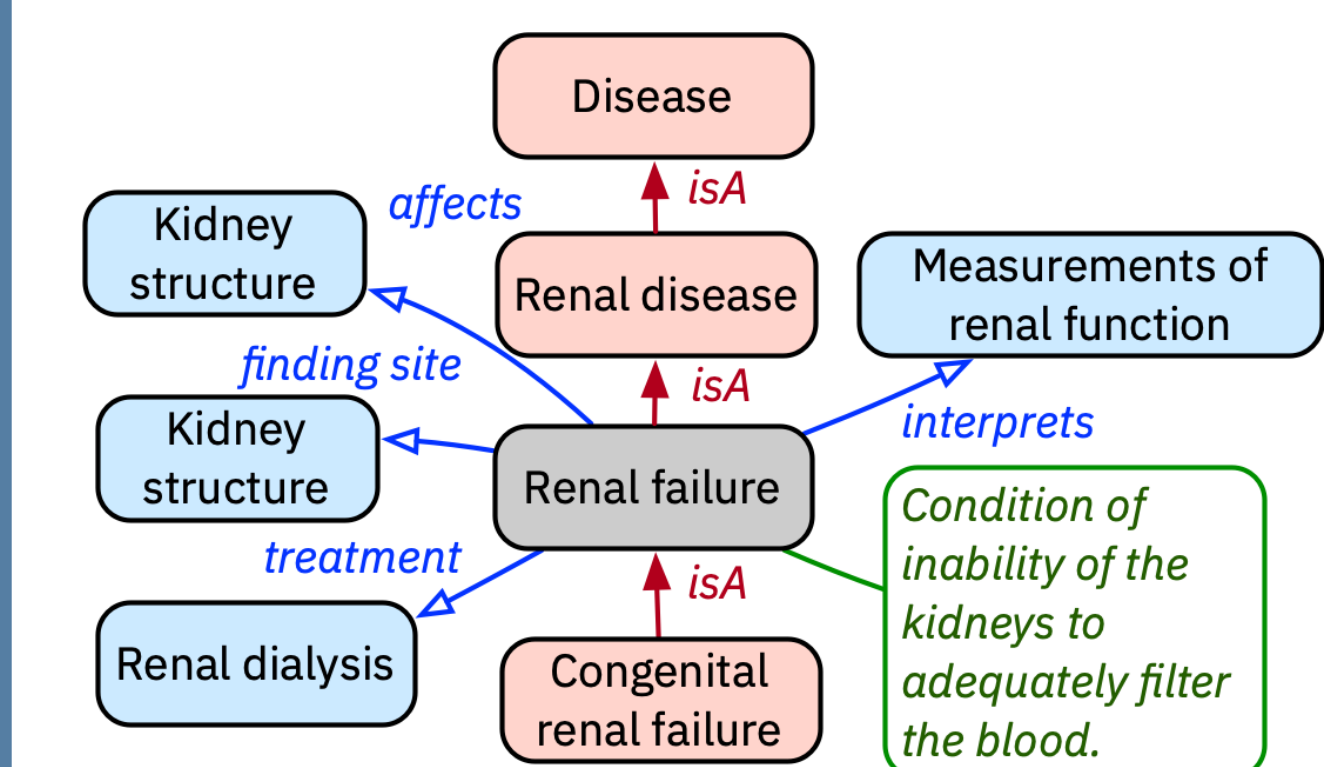


Hyperbolic Graph Module (HYP): Better capture concept hierarchies or taxonomies in the medical ontologies in hyperbolic space (HGCM).

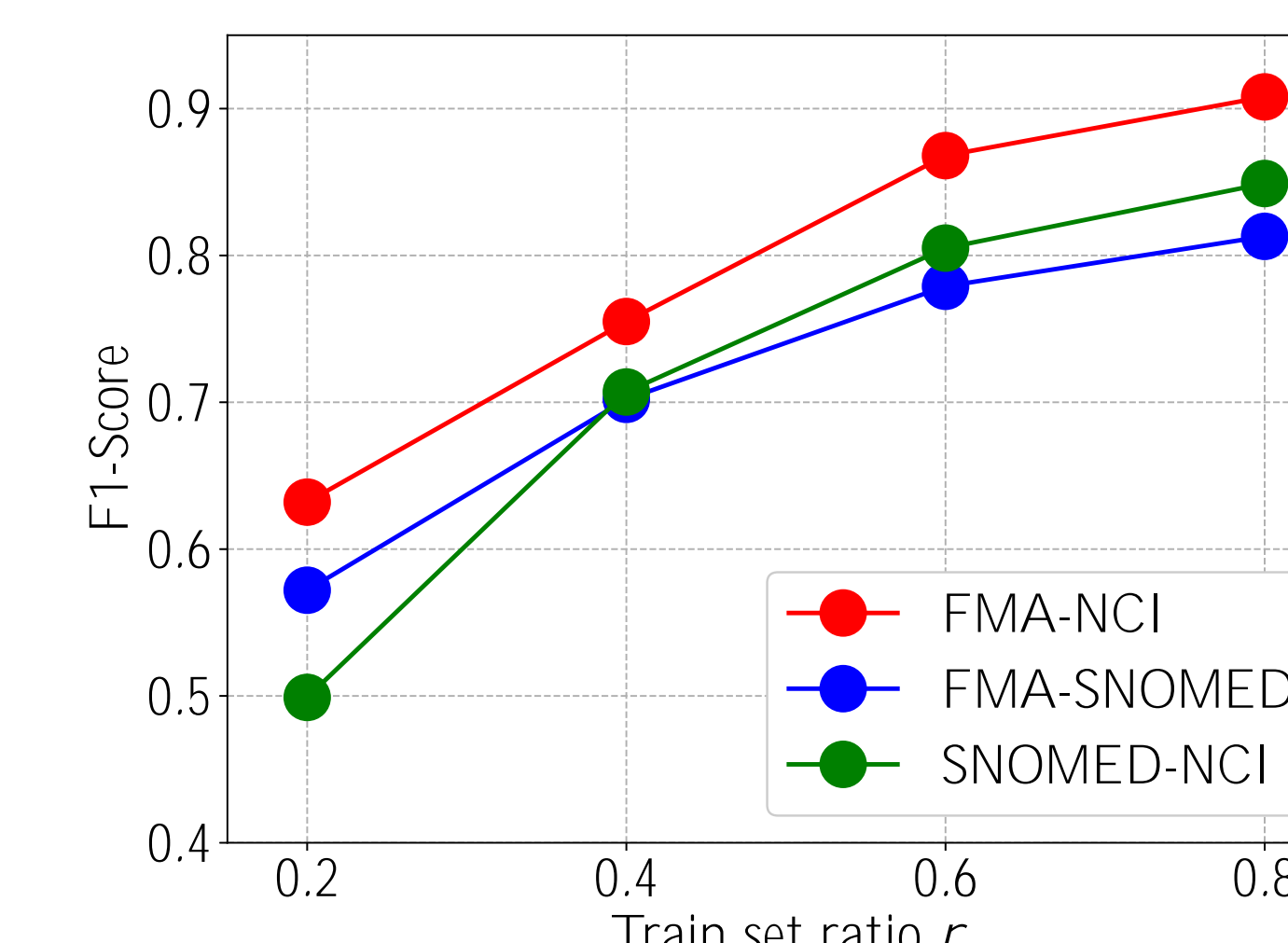
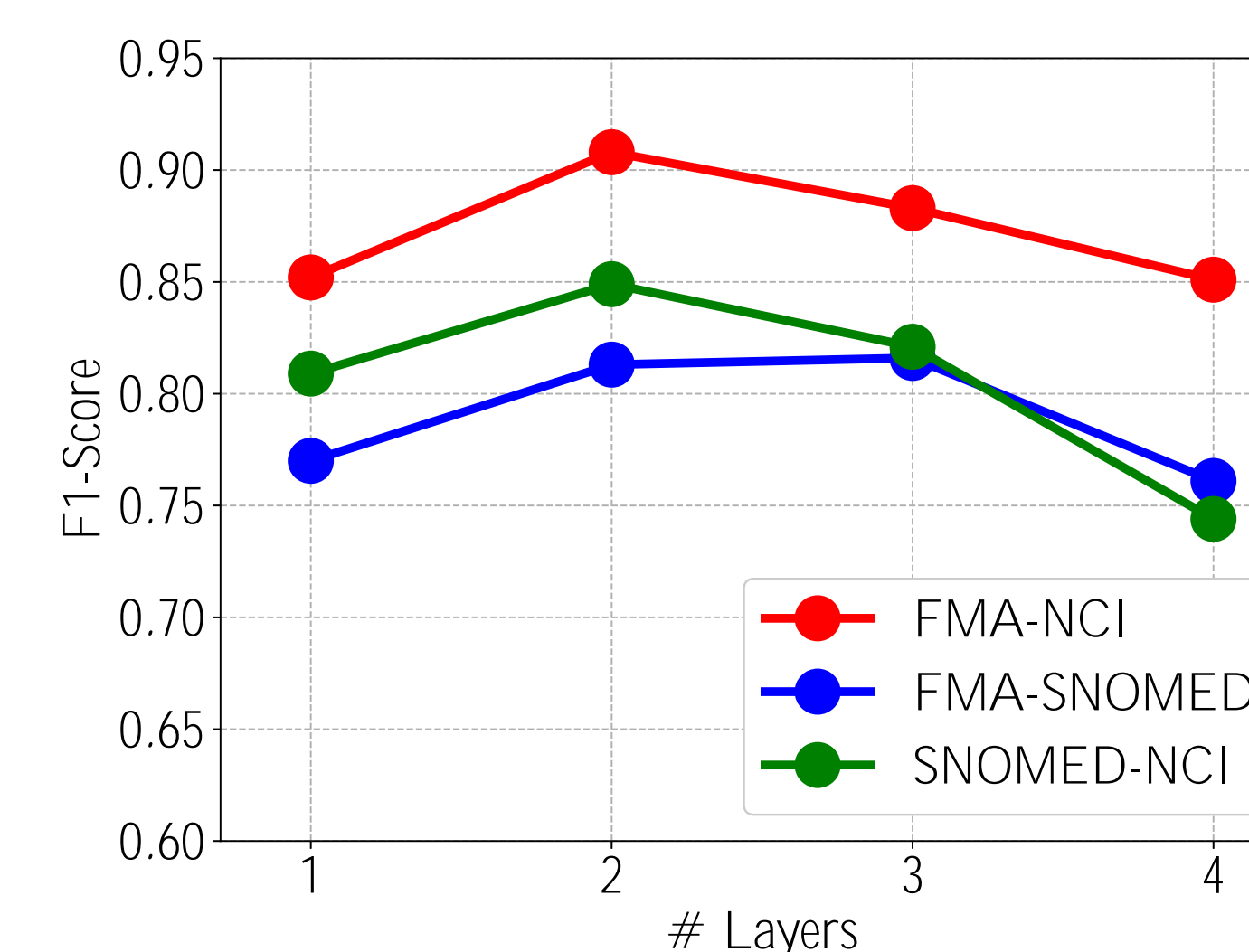
$$h_i^{i,H} = W^i K_i^{-1} h_i^{i,H} \quad K_i^{-1} b_i^i; h_i^{i,H} = K_i^{-1} K_i^{-1} \text{AGG}^{K_i^{-1}} h_i^{i,H}$$

Heterogeneous Graph Module (HET): Model non-hierarchical relational facts with other concepts of multiple types in the ontologies

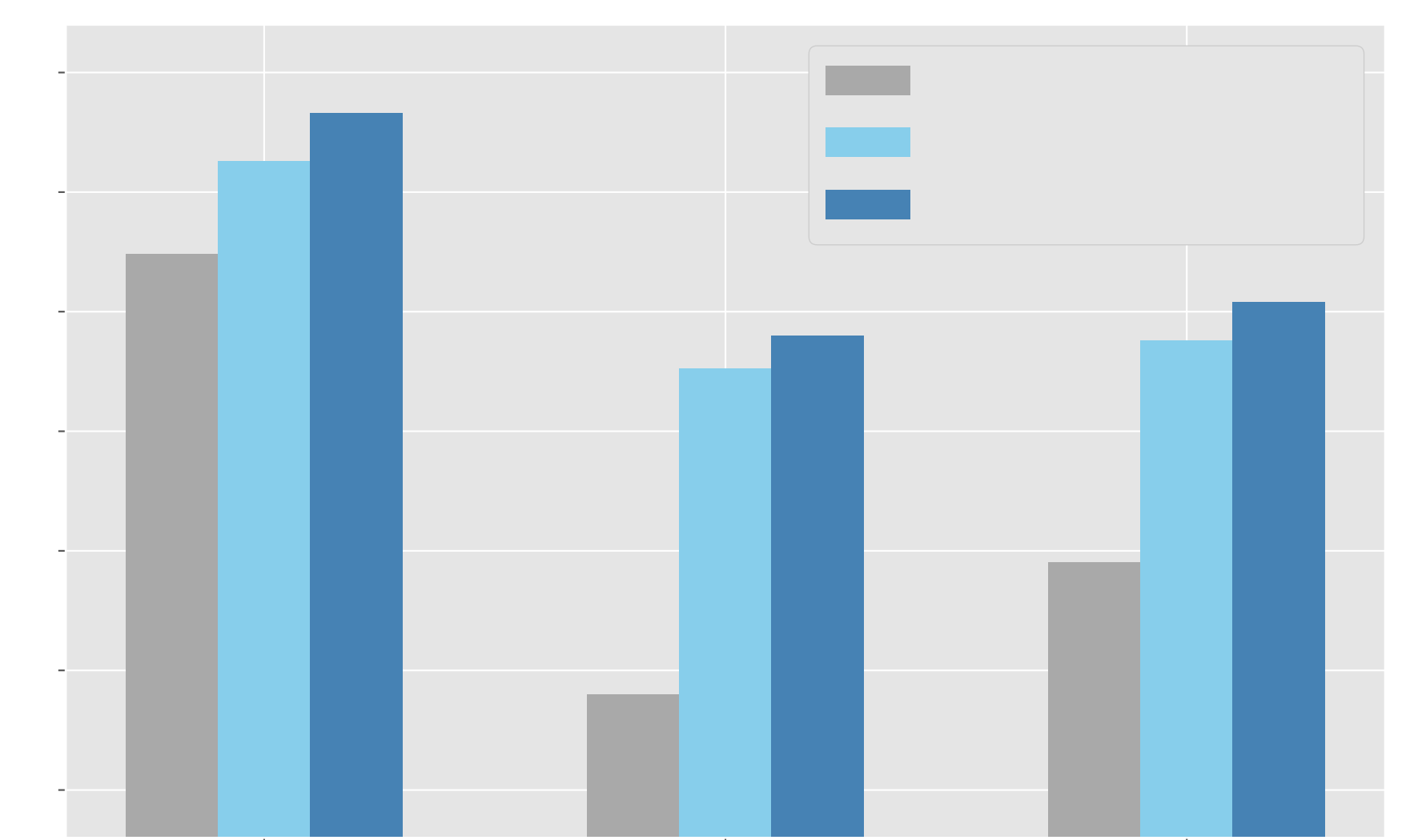
$$h_i^{i,E} = (W_0^i [h_i^{i,E} j j g_i^{i,E}]) + \times \times \frac{1}{r_2 R j 2 N_i^r} W_r^i [h_j^{i,E} j j g_j^{i,E}]$$



HYPERPARAMETERS



MEDTO VARIANTS



ACKNOWLEDGEMENT

This work was done during the internship at IBM Research - Almaden. Special thanks to the members at IBM Research (Berthold Reinwald, Xiao Qin, Nasrullah Sheikh, Lingfei Wu, Yunyao Li, etc).