An Iterative Self-Learning Framework for Medical Domain Generalization

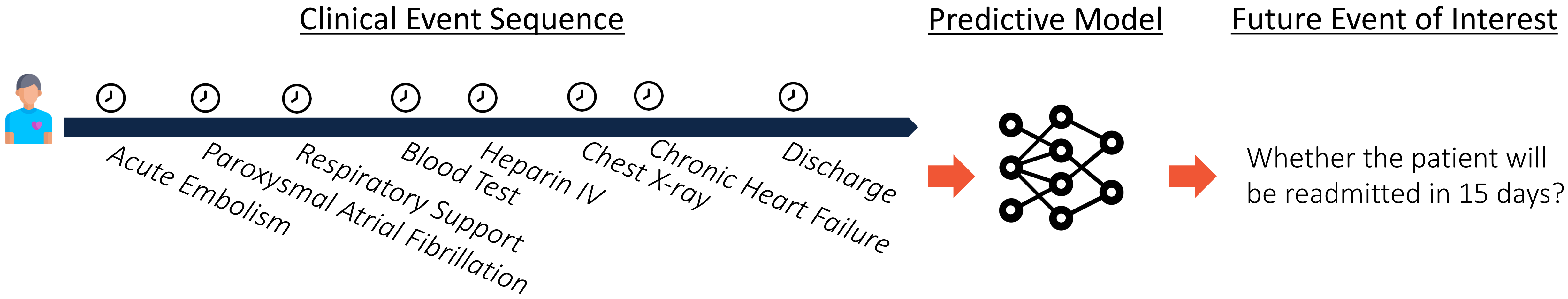
Zhenbang Wu[1], Huaxiu Yao[2], David M Liebovitz[3], Jimeng Sun[1]

[1] University of Illinois Urbana-Champaign

[2] University of North Carolina at Chapel Hill

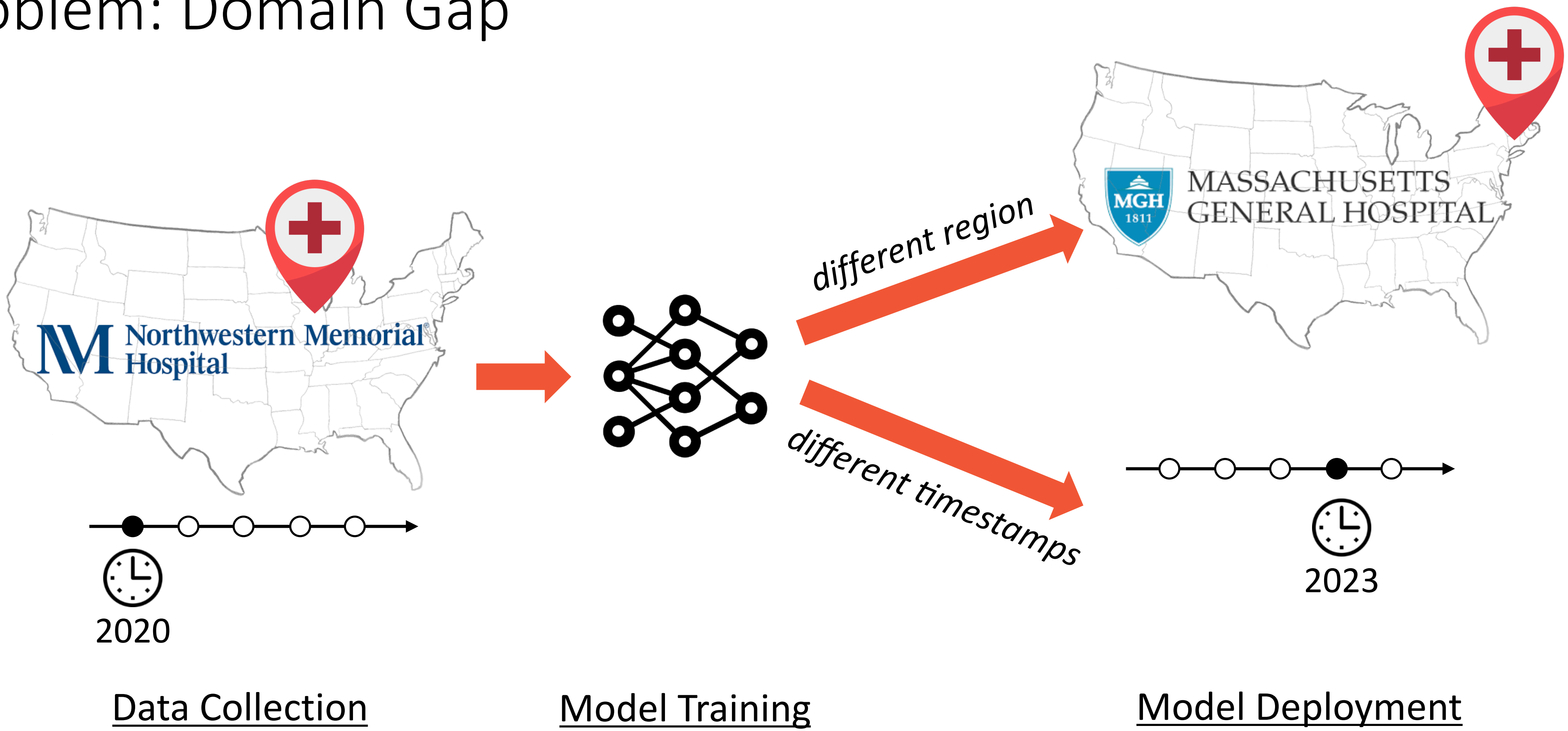
[3] Northwestern University

Background: Clinical Predictive Modeling



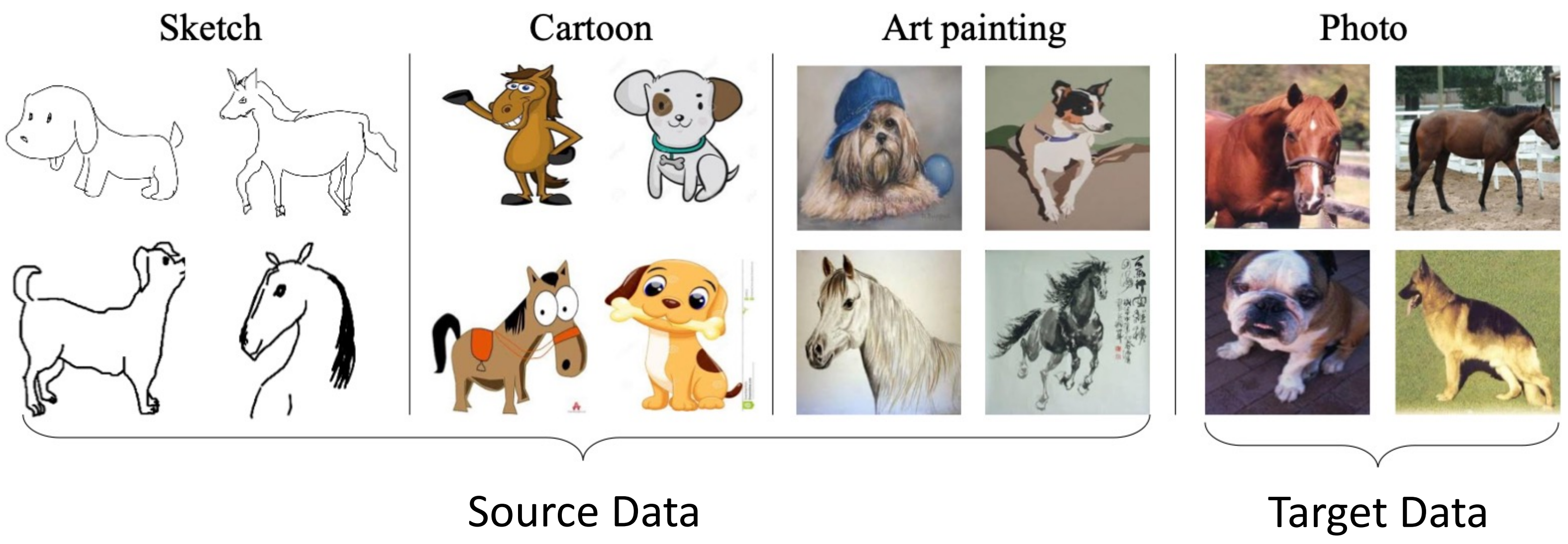
- Input: Existing event sequence
- Predict: Occurrence of future event of interest
  - E.g., hospital readmission, mortality, diagnosis of heart failure

Problem: Domain Gap



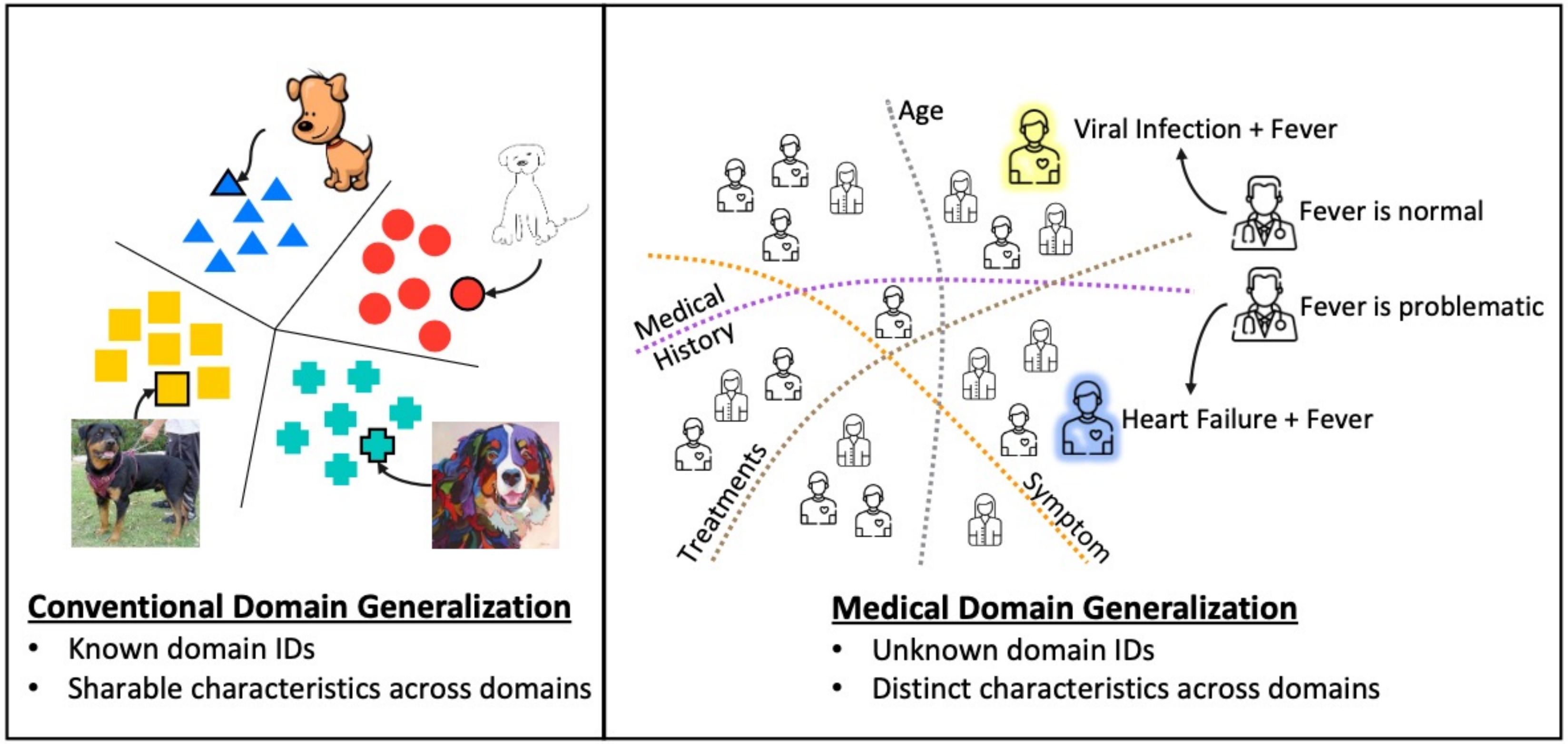
Domain Generalization (DG)

- Goal
  - Develop a model on the source data that can effectively handles potential domain shifts when applied to the target data



Jindong Wang et al. Generalizing to Unseen Domains: A Survey on Domain Generalization. IEEE Transactions on Knowledge and Data Engineering, 2022.

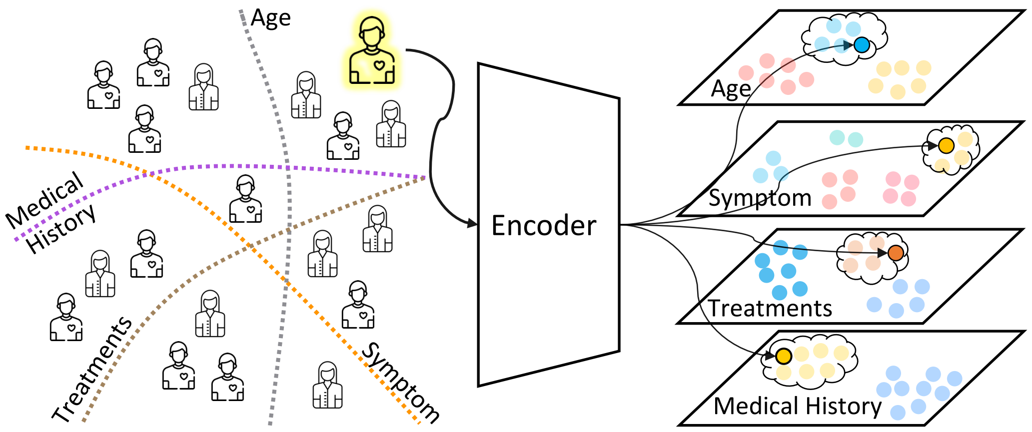
New Challenges in Medical DG



- Existing DG relies on domain IDs
  - Patients can be divided into numerous latent domains based on different features
- Existing DG attempts to train a single model
  - Patients from different domains possess distinct characteristics and require different treatment approaches

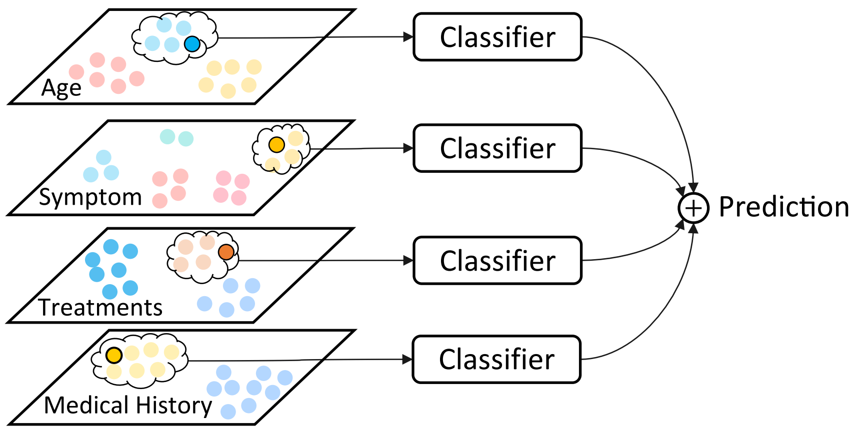
Insight 1: Decoupled Domain Discovery

- Problem:
  - Patients can be categorized into numerous latent domains based on different features
  - The categorization can be difficult to obtain and vary across different tasks
- Idea:
  - Decouple clinical features
  - Discover the domains for each type of feature

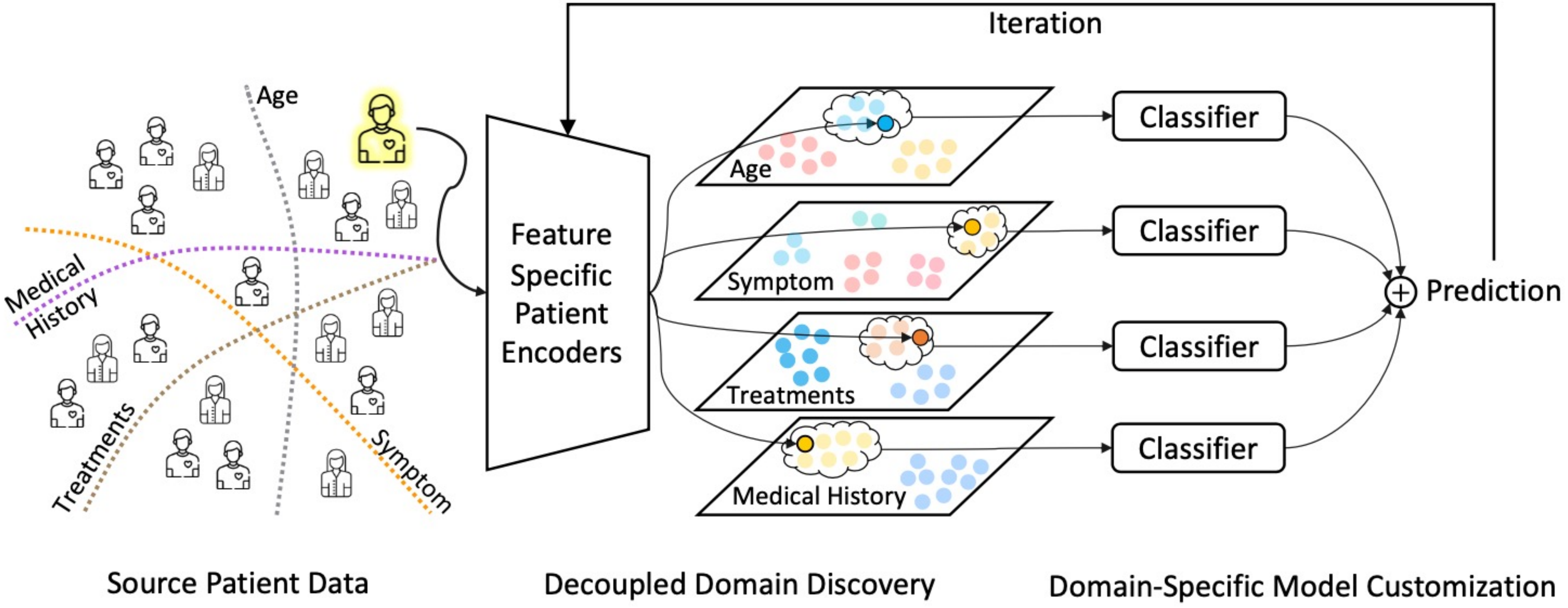


Insight 2: Domain-Specific Model customization

- Problem:
  - Unique characteristics of patients in different domains
- Idea:
  - Train customized classifiers for each domain



Method: A Self-Learning Framework for Domain Generalization



SLDG Is Robust Against Spatial and Temporal Domain Shifts

Method	eICU				MIMIC-IV			
	Readmission		Mortality		Readmission		Mortality	
	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC
Oracle	0.219 (0.01)	0.677 (0.01)	0.271 (0.01)	0.839 (0.01)	0.282 (0.01)	0.693 (0.00)	0.428 (0.00)	0.898 (0.01)
Base	0.104 (0.02)	0.510 (0.01)	0.230 (0.01)	0.803 (0.01)	0.237 (0.01)	0.665 (0.01)	0.374 (0.01)	0.861 (0.00)
DANN	0.135 (0.01)	0.538 (0.01)	0.245 (0.01)	0.808 (0.01)	0.247 (0.01)	0.673 (0.01)	0.380 (0.02)	0.873 (0.02)
MLDG	0.104 (0.01)	0.525 (0.01)	0.224 (0.01)	0.797 (0.01)	0.205 (0.01)	0.637 (0.02)	0.360 (0.01)	0.857 (0.01)
ManyDG	0.150 (0.01)	0.549 (0.01)	0.259 (0.01)	0.814 (0.01)	0.249 (0.01)	0.676 (0.01)	0.388 (0.01)	0.880 (0.01)
IRM	0.136 (0.01)	0.538 (0.01)	0.252 (0.02)	0.811 (0.01)	0.242 (0.00)	0.668 (0.01)	0.387 (0.01)	0.876 (0.01)
MMLD	0.167 (0.01)	0.578 (0.00)	0.256 (0.01)	0.818 (0.01)	0.250 (0.02)	0.679 (0.01)	0.393 (0.01)	0.887 (0.01)
DRA	0.148 (0.01)	0.551 (0.01)	0.249 (0.01)	0.810 (0.01)	0.246 (0.01)	0.670 (0.01)	0.387 (0.01)	0.875 (0.01)
SLDG	**0.186 (0.01)***	**0.623 (0.01)***	**0.268 (0.01)***	**0.824 (0.01)***	**0.274 (0.01)***	**0.690 (0.01)***	**0.416 (0.00)***	**0.899 (0.01)***

Conclusion

Zhenbang Wu
CS Ph.D. Student @ UIUC
zw12@illinois.edu

- Goal
  - Develop a clinical predictive model on the source data that effectively handles potential domain shifts when applied to the target data
- Challenges
  - Unknown domain IDs
  - Distinct characteristics across domains
- Method
  - SLDG: a self-learning framework for domain generalization
  - Iteratively discovers decoupled domains and trains customized classifiers for each discovered domain
- Result
  - Achieves up to 11% improvement in the AUPRC score over the best baseline