Pareto-Optimal Diagnostic Policy Learning in Clinical Applications via Semi-Model-Based Deep Reinforcement Learning

Zheng Yu[1] Yikuan Li[2] Joseph Kim[1]
Kaixuan Huang[1] Yuan Luo[2] Mengdi Wang[1]

[1]Princeton University

[2]Northwestern University

Apr, 2023

Sequential Decision Making in Medical Diagnostics

Sequential Decision Making in Medical Diagnostics

Trade-off between diagnose accuracy and testing costs

Sequential Decision Making in Medical Diagnostics

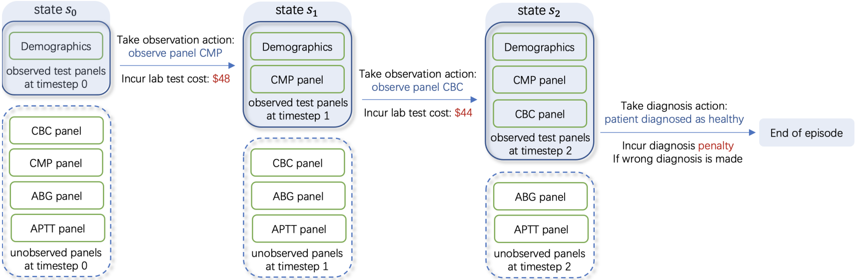


Figure: Sequential decision making model for medical diagnostics process

Cost-F_1 Pareto Front

Definition (Cost-F_1 Pareto Front of Multi-Objective Policy Optimization)

The Pareto front Π^* for cost-sensitive dynamic diagnosis is the set of policies such that

$$\Pi^* = \cup_{B>0} \operatorname*{argmax}_{\pi} \{F_1(\pi) \text{ subject to } \mathrm{Cost}(\pi) \le B\}$$

Here we consider the F_1 score metric:

$$F_1(\pi) = \frac{\mathrm{TP}(\pi)}{\mathrm{TP}(\pi) + \frac{1}{2}(\mathrm{FP}(\pi) + \mathrm{FN}(\pi))} = \frac{2\mathrm{TP}(\pi)}{1 + \mathrm{TP}(\pi) - \mathrm{TN}(\pi)}$$

Finding Cost-F_1 Pareto Front via Reward Shaping

Theorem

The Cost-F_1 Pareto front is a subset of the collection of all reward-shaped solutions, given by

$$\Pi^* \subseteq \overline{\Pi} := \cup_{\lambda \geq 0, \rho \leq 0} \operatorname*{argmax}_{\pi} \{ TN(\pi) + \lambda \cdot TP(\pi) + \rho \cdot Cost(\pi) \} .$$

Note the unconstrained policy optimization problem:

$$\max_{\pi} \ \mathrm{TN}(\pi) + \lambda \cdot \mathrm{TP}(\pi) + \rho \cdot \mathrm{Cost}(\pi).$$

is a standard cumulative-sum MDP problem, with reshaped reward:

$$R(s,a) = \begin{cases} \rho \cdot c(a), & \text{if } a \in [D] \text{ (choosing task panels)} \\ \lambda \cdot \mathbf{1}\{y = \mathrm{P}\}, & \text{if } a = \mathrm{P} \text{ (true positive diagnosis)} \\ \mathbf{1}\{y = \mathrm{N}\}, & \text{if } a = \mathrm{N} \text{ (true negative diagnosis)} \end{cases} .$$

Semi-Model-Based Deep Diagnosis Policy Optimization

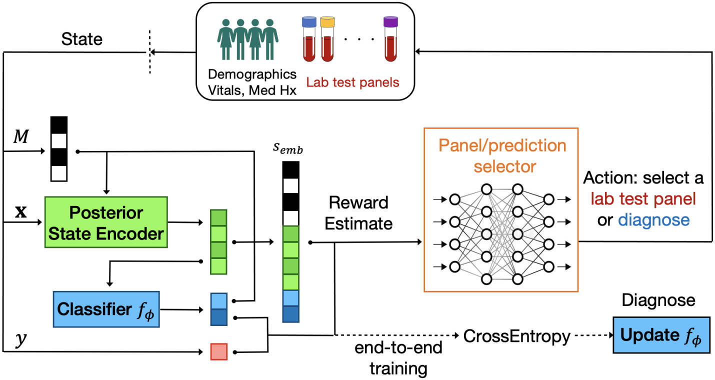
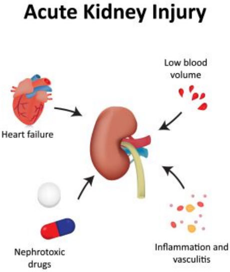


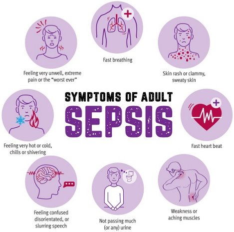
Figure: Dynamic diagnostic policy learning via semi-model-based proximal policy optimization. The full policy π comprises of three modules: posterior state encoder, classifier, and panel/prediction selector.

Empirical Results on Three Clinical Tasks

(a) Ferritin abnormality detection



(b) Acute kidney injury prediction



(c) Sepsis mortality prediction

Empirical Results on Three Clinical Tasks

Table: Comparison with full observation, fixed selection, random selection and dynamic selection baselines under no budget constraints. Our approach achieves up to 85% reduction in testing costs.

Models	Ferritin			AKI			Sepsis			Test Selection
Metrics	F_1	*AUC*	*Cost*	F_1	*AUC*	*Cost*	F_1	*AUC*	*Cost*	Strategy
LR	0.539	0.935	$290	0.452	0.797	$591	0.506	0.825	$591	Full
RF	0.605	0.938	$290	0.439	0.764	$591	0.456	0.801	$591	Full
XGBoost	0.617	0.938	$290	0.404	0.785	$591	0.431	0.828	$591	Full
LightGBM	**0.627**	**0.941**	**$290**	0.474	0.790	$591	0.500	0.844	$591	Full
3-layer DNN	0.616	0.938	$290	0.494	0.802	$591	0.517	0.845	$591	Full
LR (2 panels)	0.401	0.859	$92	0.473	0.797	$92	0.488	0.811	$92	Fixed
RF (2 panels)	0.504	0.887	$92	0.425	0.768	$92	0.478	0.828	$92	Fixed
XGBoost (2 panels)	0.519	0.895	$92	0.410	0.781	$92	0.459	0.877	$92	Fixed
LightGBM (2 panels)	0.571	0.901	$92	0.491	0.792	$92	0.502	0.864	$92	Fixed
FS	0.585	0.927	$74	0.434	0.787	$98	0.500	0.837	$90	Fixed
RS	0.437	0.845	$145	0.424	0.748	$295	0.473	0.789	$295	Random
CWCF	0.554	0.718	$256	0.283	0.510	$326	0.112	0.503	$301	Dynamic
SM-DDPO$_{\text{pretrained}}$	0.607	0.925	$80	**<u>0.519</u>**	**<u>0.789</u>**	**<u>$90</u>**	**<u>0.567</u>**	**<u>0.836</u>**	**<u>$85</u>**	Dynamic
SM-DDPO$_{\text{end2end}}$	<u>0.624</u>	<u>0.928</u>	<u>$62</u>	0.495	0.795	$97	0.562	0.845	$90	Dynamic

Empirical Results on Three Clinical Tasks

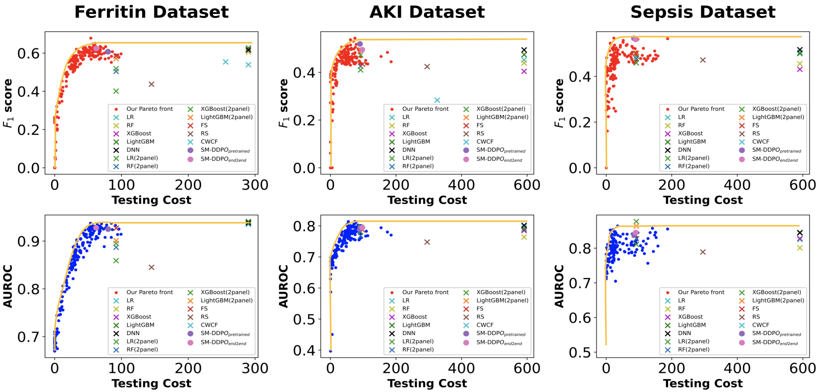


Figure: Cost-F_1 Pareto Front for maximizing F_1-score on Ferritin, AKI and Sepsis Datasets

Future work

1. Extension to time-series diagnostic tasks
2. Consideration of temporal costs and constraints
3. Inclusion of various types of diagnostic data via multimodal learning