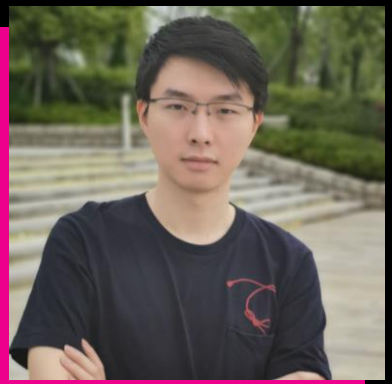
MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA



Physician-Centered EHR Data Utilization: A Pilot Study

Chengkai Wu

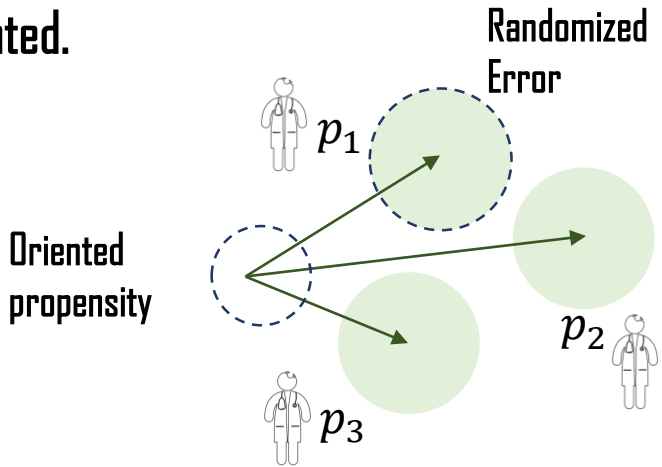
Postdoc

Zhejiang Laboratory

AIDH IMIA

1 Introduction

- EHR data records and reflects physicians' decisions.
- Physicians have varied knowledge and decision propensity.
- Such personal propensity are oriented and hardly eliminated.
- Direct mixed training cannot best utilize EHR data.

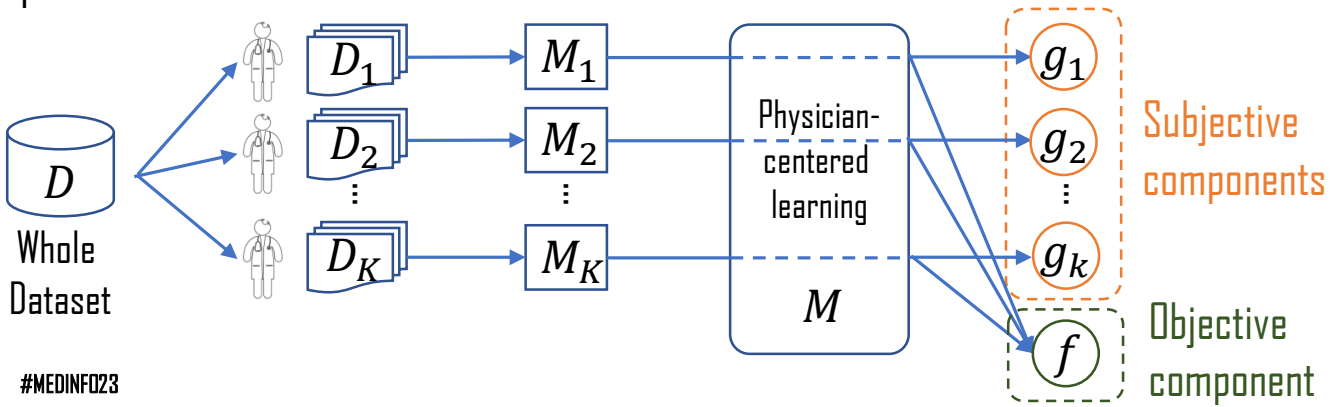


1 Introduction

- Patient-centered EHR data utilization.
  - A concept adopted from patient-centered health.
  - Suitable for applying machine learning techniques.
  - Ignores the differences between EHR decision providers (e.g., physicians).
- Physician-centered EHR data utilization.
  - Patients' personal conditions are still important. (no conflict)
  - Additional physician-centered view to reduce inconsistency/improve outcomes.

2 Methods

- Assuming decisions contains objective and subjective components.
  - Objective component: knowledge, common ones among physicians
  - Subjective components: physician's department, historical experiences, focuses on varied diseases, and personal cautiousness



2 Methods

- Adopting to a positive and unlabeled (PU) learning scheme.
- With input x:
  - A physician's personal decision: $f_k(x) = P(y_k = 1|x)$
  - The global objective decision: $f(x) = P(y = 1|x)$
  - A physician's propensity: $g_k(x) = P(y_k = 1|y = 1, x)$
  - PU scheme: $P(y = 1|y_k = 1, x) = 1$

2 Methods

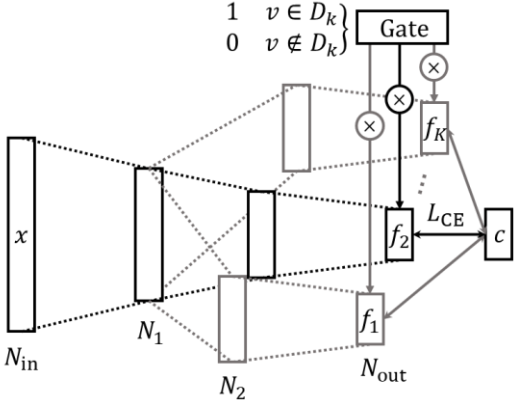
- Using the Bayes formula

$$P(y_k = 1|x) \cdot P(y = 1|y_k = 1, x) = P(y = 1|x) \cdot P(y_k = 1|y = 1, x)$$
$$\Leftrightarrow f_k(x) = f(x) \cdot g_k(x)$$

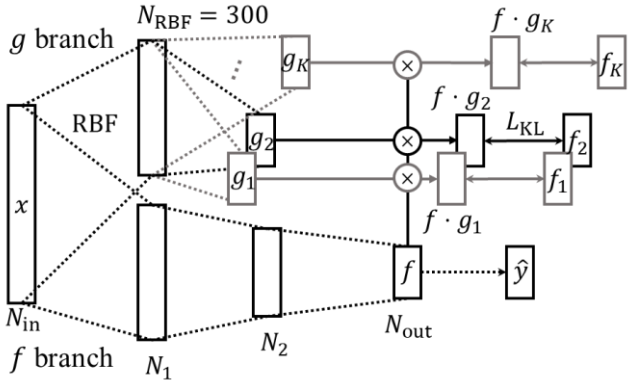
- Using multiple observed personal decisions f_k to estimate the objective global decision f and the multiple individual propensity g_k.

2 Methods

- Two-step Network design (phyC)



Network 1 for getting f_k



Network 2 for getting f and g_k

f branch: MLP network
g branch: RBF network
Loss:

$$L_{KL} = \sum_{k=1}^{K} \mathrm{KL}\left(f_k, f \cdot g_k\right)/K$$

3 Experimental Setups

- # Feature: 111 + 111, # label: 10 diagnosis for diseases
- Training set: 228,154 visits from 7 physicians
  - Each physician has a maximum diagnosis rate for at least one disease
- Testing set A: 30k visits from other physicians for each disease
  - physicians has a maximum diagnosis rate for this disease
- Testing set B: all 611,540 visits from 9 physicians
- Testing metrics: area under the ROC curve AUC_{ROC}
- Comparing methods: PhyC (ours), MLP, and RBF.

4 Results

Table 1. Evaluation results between MLP, RBF and PhyC using AUC_{ROC}.

Disease	Set A			Set B		
	MLP	RBF	PhyC	MLP	RBF	PhyC
ARR	0.576±0.004	0.554±0.002	**0.583**±0.007	**0.562**±0.004	0.546±0.002	0.561±0.006
AF	0.687±0.006	0.669±0.002	**0.697**±0.004	0.725±0.006	0.713±0.003	**0.735**±0.004
CAS	**0.673**±0.002	0.664±0.001	0.672±0.002	**0.681**±0.001	0.674±0.001	0.680±0.002
DM	0.692±0.006	0.686±0.002	**0.708**±0.002	0.698±0.005	0.691±0.001	**0.713**±0.002
HTN	0.590±0.004	0.588±0.002	**0.593**±0.003	0.599±0.003	0.596±0.002	**0.603**±0.002
CM	0.703±0.007	0.678±0.003	**0.712**±0.004	0.673±0.003	0.657±0.001	**0.674**±0.002
CHD	0.787±0.003	0.791±0.002	**0.801**±0.001	0.781±0.002	0.783±0.002	**0.794**±0.001
HLP	0.620±0.002	0.612±0.001	**0.629**±0.001	0.630±0.002	0.621±0.001	**0.639**±0.001
HUA	0.641±0.007	0.659±0.002	**0.670**±0.009	0.640±0.007	0.649±0.002	**0.668**±0.004
GERD	0.630±0.016	0.637±0.006	**0.669**±0.011	0.633±0.016	0.635±0.005	**0.673**±0.010
Average	0.660	0.654	**0.674**	0.662	0.656	**0.673**

Results of 10 repetitive tests

4 Results

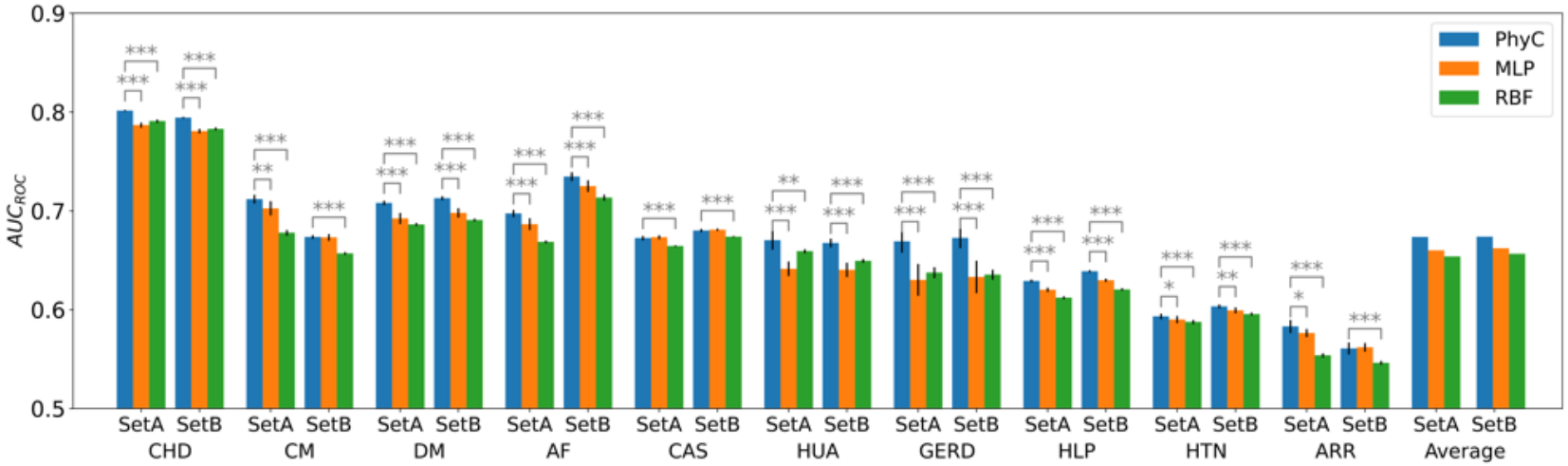


Figure 2. Comparative graph on AUC_{ROC} scores

* for p-value<0.05, ** for p-value<0.01, and *** for p-value<0.001

5 Discussion

- A) significant improvements show learning from physicians can help to build better artificial intelligent models.
- B) Both improvements on Set A and B show the improvements being led by the learning of the objective component.
- C) This study is a pilot study for physician-centered EHR data utilization to explore a framework and suitable methods.
- D) We hope this physician-centered methodologic concept can be paid attention to by the community.

5 Discussion

Thank you