*MEDINFO 2023 — The Future Is Accessible*
*J. Bichel-Findlay et al. (Eds.)*

*doi:10.3233/SHTI231061*

# Physician-Centered EHR Data Utilization: A Pilot Study

Chengkai WU[a,#] , Tianshu ZHOU[a,#], Yu TIAN[b], Huiyao SUN[a], Zhong LIU[c,1] and Jingsong LI[a,b,1]

[a] *Research Center for Healthcare Data Science, Zhejiang Laboratory, Hangzhou, China*
[b] *Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China*
[c] *Health Management Center, the First Affiliated Hospital, Medical School of Zhejiang University, No.79 Qingchun Rd., Hangzhou, China*

**Abstract.** The utilization of vast amounts of EHR data is crucial to the studies in medical informatics. Physicians are medical participants who directly record clinical data into EHR with their personal expertise, making their roles essential in follow-up data utilization, which current studies have yet to recognize. This paper proposes a physician-centered perspective for EHR data utilization and emphasizes the feasibility and potentiality of digging into physicians' latent decision patterns in EHR. To support our proposal, we design a physician-centered CDS approach named PhyC and test it on a real-world EHR dataset. Experiments show that PhyC performs significantly better in the auxiliary diagnosis of multiple diseases than globally learned models. Discussions on experimental results suggest that physician-centered data utilization can help to derive more objective CDS models, while more means for utilization need further exploration.

**Keywords.** Physician-centered utilization, EHR data utilization, decision pattern

## 1. Introduction

Electronic health records (EHR) are essential data sources for clinical decision support (CDS) research [1]. EHR data obtains the advantages of easy accessibility, vast amounts, and high population coverage, making it suitable for early-stage, large-scale, and multi-target CDS, such as individual diagnosis in outpatient scenarios.

The relatively low quality of EHR data has been discussed [2], and improved views on EHR data have been proposed to utilize EHR more objectively, including semi-supervised (SS) [3] and positive and unlabeled (PU) [4] learning, which both discuss the existence of unlabeled records. However, the current studies still regard EHR data as globally mappable and oversimplify the CDS into a "record to model" mapping [5,6].

However, EHR data is not globally consistent from the data generation aspect. A record in EHR is only the digital reflection of the decisions one specific physician makes on a patient at a particular visit. Besides objective medical knowledge, clinical decisions largely depend on the physician's subjective factors [7], e.g., the physician's major,

[1] Corresponding Authors: Jingsong Li, e-mail: ljs@zju.edu.cn; Zhong Liu, e-mail: liuzhongzheyi@zju.edu.cn.
[#] These authors contributed equally to this work.

historical experiences, focuses on varied diseases, and personal cautiousness. Instead of global mapping, these factors are oriented and often lead to personal bias [8]. Unlike randomized errors, such bias can hardly be eliminated during mixed training.

On the other hand, as a medical provider, a physician will have one's own decision pattern being recorded in the EHR [9], which has more consistency among one's records than a global model. Each decision pattern can contain both objective and subjective components. Within the two components, the former can help to derive objective global CDS models, while the latter can help to model high-quality clinical experience.

Therefore, we propose that a more refined **physician-centered** perspective should be established to utilize vast amounts of EHR data, by digging into physicians' decision patterns. In this section, we highlight the crucial role of physicians in utilizing EHR data, which current studies have yet to recognize. In the following sections, we will introduce an example CDS method called PhyC, designed under the perspective of physician-centered utilization, as a pilot study to provide experimental support for our proposal.

## 2. Methods

### *2.1. Study Settings and Notifications*

This study concentrates the target medical scenario on a single-department, single-visit (cross-sectional), multi-target auxiliary diagnosis. We regard EHR data as a PU dataset, in which recorded positive labels are considered accurate. For a dataset $D = \bigcup_{k=1}^{K} D_k$ that contains visit records from a total of $K$ physicians $p_1, \dots p_K$, we use $v = (x, c) \in D_k$ to denote a visit from physician $p_k$, with $x$ and $c$ being its features and target diseases (1: positive, 0: unlabeled), respectively. We use $y$ to denote the unknown real disease existence (1: positive, 0: negative), and $y_k$ to denote the subjective decision that $p_k$ would make under condition $x$. The goal of this study is to use multiple personal decision patterns $\mathcal{F} = \{f_1, \dots f_K\}$ to derive a more objective global CDS model $f$.

### *2.2. Proposed Approach*

In this section, we propose a two-step learning method called PhyC, which includes steps of decision pattern learning and global model learning, to train an objective CDS model.

The first step aims to model physicians' subjective decision patterns $\mathcal{F}$. A direct way is to train $K$ basic models/networks (we regard the multi-layer perceptron, MLP, as basic networks in this study, Figure 1(a)) using one's own records $f_k: c \hookleftarrow x, v \in D_k$. In our practice, we use a multi-branch network to train $f_k$, with a gate input to ensure that only $v \in D_k$ are used in optimizing the $k^{\text{th}}$ branch, as shown in Figure 1(b).

In the second step, we propose a Bayes-based method to learn a global objective model $f$. An important property for PU problems is that conditional probability $P(y = 1|y_k = 1, x) = 1$ always satisfies [10]. Putting this property into a Bayesian equation, we can then derive the relationship between a physician's decision pattern $f_k(x) = P(y_k = 1|x)$ and the global model $f(x) = P(y = 1|x)$:

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

In the equation, a physician's decision pattern $f_k$ is a multi-dimensional probability for diseases being diagnosed by $p_k$ under condition $x$, and the wanted global model $f(x)$ is the objective probability for diseases to exist. By defining a personal propensity score $g_k(x) = P(y_k = 1|y = 1, x)$ representing the possibility for a physician to diagnose the latent positive diseases $y$ under condition $x$, the equation is then $f_k = f \cdot g_k$. Because all values $f_k(x), k = 1..K, x \in D$ can be output using $\mathcal{F}$, the functions of $f$ and $g_k$ can be separated by minimizing the KL divergence $L_{\mathrm{KL}} = \sum_{k=1}^{K} \mathrm{KL}(f_k, f \cdot g_k)/K$ as the loss of the global network. During optimization, since the common model $f$ affects all $K$ components, but each $g_k$ affects only one, the model's continuous requirement will lead personal propensity into $g_k$, leaving the objective component inside $f$.

In modeling the above separation process, we construct a network containing two subnetworks, shown in Figure 1(c). The $f$ branch uses a basic MLP for the fairness of experimental comparison, while the $g$ branch is a radial basis function (RBF) network to learn physicians' propensity locally. Notice that the $g$ branch and the values of $f_k$ are only used in loss optimization. In other words, the $f$ branch with a basic MLP network has learned all the needed weights to map for prediction.

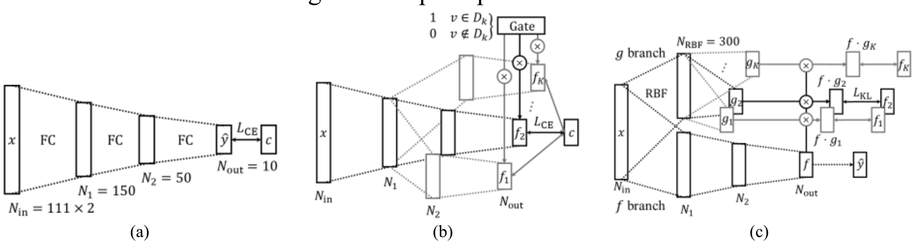


**Figure 1.** a) Basic MLP network b) decision pattern learning network c) global model learning network.

## 2.3. Experimental Setups

We experiment with real-world EHR data from the Cardiology Department, The First Affiliated Hospital, Medical School of Zhejiang University. The dataset contains 839,694 visits and is recorded by 16 physicians. 111 measurements that satisfy existent rates >5% are used as input features. Each feature dimension is normalized, zero-filled, and uses an additional binary dimension to flag the absence. 10 diseases that satisfy positive rates >1% are used as output targets with binary labels.

The training set includes all physicians whose diagnosis rate reaches the maximum on any target disease. The final training set contains 228,154 visits from 7 physicians. In this way, we make sure that 1) each target disease has one or more physicians who major in it; 2) the physicians split into the training and testing sets are not overlapped.

Since it is a PU problem, we use two testing sets to evaluate the performance. Set A is designed to reduce unlabeled positive samples best: for each disease, we use randomly selected 30,000 visits from the physicians who have the maximum diagnosis rate on the disease and are not included in the training set for testing, where only the performance on this dimension counts. Set B uses all records not used in training for testing, which can help avoid the potential subjective effects of using records from one physician.

Comparison is performed between PhyC, MLP, and RBF. MLP and the $f$ branch of PhyC use the same shaped network to derive predictions. The hyperparameters for MLP, the decision pattern learning network, and the $f$ branch are tuned using MLP. The hyperparameters for RBF and the $g$ branch are tuned using RBF. The area under the ROC curve ($\mathrm{AUC_{ROC}}$) is used to evaluate, and 10 repetitive tests are applied.

## 3. Results

The evaluation results for the three methods are listed in Table 1. Each value is reported in a mean±std form, with the best results marked in bold.

**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** |

A graphic comparison of target diseases is shown in Figure 2. Signs for significance levels derived by t-tests on each disease dimension are also presented in the figure, with signs * for p-value<0.05, ** for p-value<0.01, and *** for p-value<0.001.

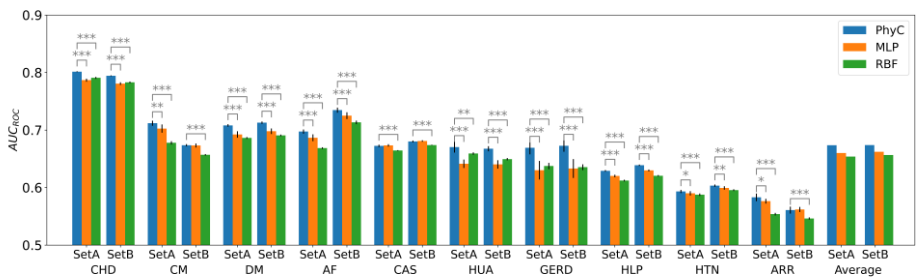


**Figure 2.** Comparative graph on $AUC_{ROC}$ scores.

## 4. Discussion

The results suggest that PhyC performs significantly better than MLP and RBF networks on most dimensions within the target diseases. As shown in Figure 1, the outputs of PhyC are solely decided by its $f$ branch, which uses the same network structure as MLP, meaning such improvements are irrelevant to network complexity. Therefore, it can only be caused by the optimization of $f_k = f \cdot g_k$.

We focus on discussing what $f$ learned in optimization. An $f$ network containing either objective or subjective components has the potential to make PhyC evaluated better on set A, since there could be similar subjective patterns between physicians with top diagnosis rates. On set B, however, a subjective pattern is likely to harm the overall performance, since the subjective component can only exist in the minority of physicians. Therefore, a potential subjective component would make PhyC less effective on set A, while a potential objective component would not change the relative performance much between the two testing sets. The total and dimensional comparisons fit more to the latter assumptions, confirming that $f$ has learned a global objective component.

The PhyC approach is a pilot methodologic implementation of the physician-centered utilization concept. In future works, efforts can be put into 1) the utilization scenarios for the personal propensity scores $g_k$, 2) mutual evaluation between physicians, and 3) multi-department and multi-center aggregation.

## 5. Conclusions

In this paper, we mainly introduce a novel concept to utilize EHR data by placing data providers, the physicians, in a central place. We explain clinical reasons physicians have different decision patterns stored in EHR and how the decision patterns can help derive a more objective CDS model by identifying the objective component and separating the subjective components. To endorse our proposal, we design the PhyC approach and test the approach in a real-world EHR-based dataset. The results suggest that our approach can help train more objective CDS models and significantly enhance final CDS performances.

## Acknowledgements

This work was supported by China Postdoctoral Science Foundation (No. 2022TQ0312), the National Key Research and Development Program of China (No.2022YFC2504605), and Key Research Project of Zhejiang Laboratory (No.2022ND0AC01).

## References

[1] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for Electronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform. 2018 Sep;22(5):1589-604, doi: 10.1109/JBHI.2017.2767063.

[2] Sarwar T, Seifollahi S, Chan J, Zhang X, Aksakalli V, Hudson I, Verspoor K, Cavedon L. The secondary use of electronic health records for data mining: data characteristics and challenges. ACM Comput Surv. 2022 Jan;55(2):1-40, doi: 10.1145/3490234.

[3] Bernardini M, Romeo L, Frontoni E, Amini MR. A semi-supervised multi-task learning approach for predicting short-term kidney disease evolution. IEEE J Biomed Health Inform. 2021 Oct;25(10):3983-94, doi: 10.1109/JBHI.2021.3074206.

[4] Zhang L, Ding X, Ma Y, Muthu N, Ajmal I, Moore JH, Herman DS, Chen J. A maximum likelihood approach to electronic health record phenotyping using positive and unlabeled patients. J Am Med Inform Assoc. 2020 Jan;27(1):119-26, doi: 10.1093/jamia/ocz170.

[5] Hong N, Liu C, Gao J, Han L, Chang F, Gong M, Su L. State of the art of machine learning-enabled clinical decision support in intensive care units: literature review. JMIR Med Inform. 2022 Mar;10(3):e28781, doi: 10.2196/28781.

[6] Kamel Rahimi A, Canfell OJ, Chan W, Sly B, Pole JD, Sullivan C, Shrapnel S. Machine learning models for diabetes management in acute care using electronic medical records: a systematic review. Int J Med Inform. 2022 Apr;162:104758, doi: 10.1016/j.ijmedinf.2022.104758.

[7] Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. BMC Med Inform Decis Mak. 2016 Nov;16(1):138, doi: 10.1186/s12911-016-0377-1.

[8] Norton WE, Chambers DA. Unpacking the complexities of de-implementing inappropriate health interventions. Implement Sci. 2020 Jan;15(1): 2, doi: 10.1186/s13012-019-0960-9.

[9] Sohn S, Wi CI, Juhn YJ, Liu H. Analysis of clinical variations in asthma care documented in electronic health records between staff and resident physicians. Stud Health Technol Inform. 2017;245:1170-4.

[10] Bekker J, Davis J. Learning from positive and unlabeled data: a survey. Mach Learn. 2020 Apr;109(4):719-60, doi: 10.1007/s10994-020-05877-5.