Enhancing Personalized Healthcare via Capturing Disease Severity, Interaction, and Progression

Yanchao Tan[†], Zihao Zhou[†], Leisheng Yu[‡], Weiming Liu[¶], Chaochao Chen[¶],
Guofang Ma[◇], Xiao Hu[§], Vicki S Hertzberg[§], Carl Yang[♠*]
[†]*College of Computer and Data Science, Fuzhou University, Fuzhou, China*
[‡]*Department of Computer Science, Rice University, Houston, United States*
[¶]*College of Computer Science, Zhejiang University, Hangzhou, China*
[◇]*School of Computer Science, Zhejiang Gongshang University, Hangzhou, China*
[§]*Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, United States*
[♠*]*Department of Computer Science, Emory University, Atlanta, United States*
yctan@fzu.edu.cn, reviverkey@gmail.com, ly50@rice.edu, {21831010, zjuccc}@zju.edu.cn,
maguofang@zjgsu.edu.cn, {xiao.hu, vhertzb, j.carlyang}@emory.edu

***Abstract*—Personalized diagnosis prediction based on electronic health records (EHR) of patients is a promising yet challenging task for AI in healthcare. Existing studies typically ignore the heterogeneity of diseases across different patients. For example, diabetes can have different complications across different patients (e.g., hyperlipidemia and circulatory disorder), which requires personalized diagnoses and treatments. Specifically, existing models fail to consider 1) *varying severity* of the same diseases for different patients, 2) *complex interactions* among syndromic diseases, and 3) *dynamic progression* of chronic diseases. In this work, we propose to perform personalized diagnosis prediction based on EHR data via capturing disease severity, interaction, and progression. In particular, we enable personalized disease representations via severity-driven embeddings at the disease level. Then, at the visit level, we propose to capture higher-order interactions among diseases that can collectively affect patients' health status via hypergraph-based aggregation; at the patient level, we devise a personalized generative model based on neural ordinary differential equations to capture the continuous-time disease progressions underlying discrete and incomplete visits. Extensive experiments on two real-world EHR datasets show significant performance gains brought by our approach, yielding average improvements of 10.70% for diagnosis prediction over state-of-the-art competitors.**

I. Introduction

Electronic health records (EHR) containing patients' hospital visit information have been widely analyzed to help researchers and doctors build predictive models for clinical decision-making [3], [10]. Among them, diagnosis prediction has been successfully deployed on many industrial platforms to assist doctors in improving diagnostic efficiency (e.g., Medical AI of Tencent[1], and Microsoft Cloud for Healthcare[2]). Many of them leverage deep learning methods to model the hospital visits, such as recurrent neural networks [3], [12], attention-based mechanisms [3], [14], and graph neural networks [4], [22].

*The corresponding author is Carl Yang.

[1]https://healthcare.tencent.com/

[2]https://www.microsoft.com/en-us/industry/health/microsoft-cloud-for-healthcare/

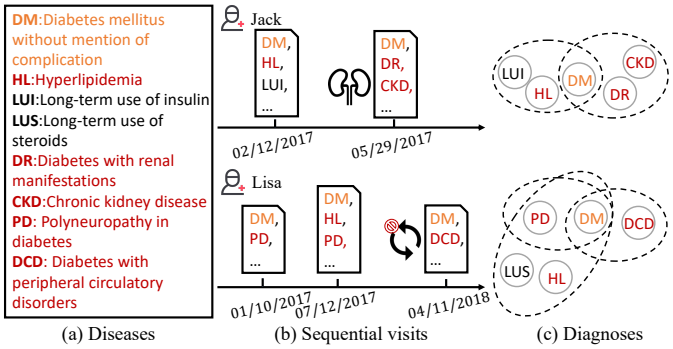


Fig. 1: A toy example of two diabetic patients, each modeled as a sequence of multiple hospital visits, where each visit is modeled as a set of diagnosed diseases.

Although deep learning models have shown great success in diagnosis predictions, they do not fully consider the heterogeneity of diseases, where one disease can have different phenotypes and should be treated in different ways [26]. As shown in Figure 1, when Jack and Lisa are first diagnosed with diabetes (DM), they can have different complications with different severities, namely hyperlipidemia (HL) for Jack and polyneuropathy in diabetes (PD) for Lisa. Moreover, both the interactions among syndromic diseases diagnosed in one visit and the progression of chronic diseases across multiple visits can imply different disease phenotypes and adverse reactions. For example, the disease interactions among {DM, HL, LUI, ...} can affect the disease progression of DM, and lead to comorbidities in the genitourinary system (e.g., CKD) for Jack, while the interactions among {DM, PD, ...} can result in different comorbidities in the endocrine glands (e.g, DCD) for Lisa. To provide personalized treatment for patients according to different disease phenotypes, an ideal model should consider heterogeneous disease severity, interaction, and progression based on available EHR data.

In this work, we consider a personalized diagnosis prediction task. The key challenges can boil down to the appropriate representation learning of diseases, visits, and patients.

Challenge I (Disease level): *How to properly obtain personalized initial disease representations?* The effects of the disease can be different due to patients' own health status (e.g., different ages and lifestyles). However, to protect the privacy of patients, it is not always possible to directly obtain patients' personal information. In the example of Figure 1, after first diagnosed with DM, the subsequent visit of Jack is 3.5 months while that of Lisa is 6 months, which can indicate different disease severities, but the signal is noisy and it remains unknown as how to leverage such weak signals to enhance personalized disease modeling.

Challenge II (Visit level): *How to comprehensively model complex disease interactions inside one visit?* Visits contain variable numbers of diagnosed diseases and diseases in a visit can have complex syndromic interactions, which can result in complicated adverse reactions. As shown in Figure 1, both the first-order interactions (e.g., DM $\rightarrow$ CKD) and higher-order interactions (e.g., DM $\times$ HL $\rightarrow$ CKD) are potential combinations that can lead to Jack's next diagnosed CKD. However, it is computationally inhibitive to traverse all combinations of diseases, and it is unclear how to provide an accurate visit representation based on the set of diseases in each visit.

Challenge III (Patient level): *How to continuously model visits with discrete timestamps and irregular intervals for patients?* Most progressions of diseases and changes in patient body status are continuous in nature, but the observed clinical records are discrete and incomplete due to the patients' irregular visits. Existing algorithms [9], [10] only consider the order of discrete visits and ignore informative diagnosed timestamps, and thus fail to accurately capture latent disease dynamics over time. How to fully leverage partially observed patient visits with irregular intervals for continuous modeling of disease progression remains unknown.

To address the above challenges, we propose ProCare in Section III, which includes three pivotal levels: (i) *disease level*, we combine ontology-aware and frequency-aware encodings to enhance the severity information in disease representations; (ii) *visit level*, we design a hypergraph convolution mechanism to model first- and higher-order combinatorial disease interactions inside each single visit; (iii) *patient level*, we leverage neural ordinary differential equations to continuously model disease progression under discrete and incomplete observations of patients' irregular visits.

Extensive experiments on real benchmark EHR datasets towards personalized diagnosis prediction tasks demonstrate the superiority of ProCare against state-of-the-art approaches (e.g., with up to 16.24% relative improvements in Recall@10 on MIMIC-III over the best baseline). More comprehensive results and discussion as well as ablation studies, and hyperparameter studies are presented and analyzed in Section IV.

II. Related Work

A. Modeling Interactions for Diagnosis Prediction

Since there exist syndromic relations among diseases that can cause adverse patient reactions, many studies propose to capture disease interactions for accurate diagnosis prediction [5], [24]. For example, Marble et al. [5] based on tensor factorization for high throughput phenotyping to model disease interactions. To effectively capture disease relations based on sparse interactive data, recent studies leverage deep learning methods for modeling structural information [4], [10], [25]. For example, GRAM [4] constructed a disease graph from ontological medical knowledge to represent medical concepts as combinations of their ancestors. CGL [10] utilized a hierarchical structure of medical domain knowledge and introduced an ontology weight to capture hidden disease interactions. However, they rely on pre-given graph structures that have to be constructed manually and cannot work when there are no explicit pair-wise relations between diseases in a visit.

Besides modeling pair-wise (first-order) relations, existing studies also capture the higher-order relations. For example, Cache [23] proposed to provide clinical predictions based on hypergraph representation learning. These methods do not consider the progression of diseases.

B. Modeling Dynamics for Diagnosis Prediction

Most EHR data can be formed as a sequence of patient examination records at hospital visits for diagnosis prediction. [8], [27]. For example, Liu et al. [8] utilized a temporal graph to capture temporal relations of the medical events in each clinical sequence. Recently, deep learning has been widely adopted to model the dynamics in EHR data for diagnosis prediction. For example, RETAIN [3] employed an attention process on recurrent neural networks (RNNs) to model the order of visits for the disease prediction task. Dipole [12] applied bidirectional long-short-term memory networks and attention mechanisms to predict patient visit information. Timeline [1] utilized time-aware attention mechanisms in RNNs for health event predictions. Chet [9] designed a context-aware dynamic graph learning method to learn disease combinations and disease development schemes. However, they ignore the timestamps of the visits and the irregular gaps between them, thus failing to capture the real continuous dynamics of disease progression.

To better leverage diagnosis timestamps for diagnosis predictions, recent studies [11], [15] proposed to capture temporal dependencies via learning time embeddings. HiTANet [11] designed time interval vectors to model irregular time gaps between successive visits. As closest to us, [16] used ordinary differential equations to describe the continuous dynamics of disease progression. However, they do not consider complex interactions among syndromic diseases.

III. The ProCare Framework

A. Problem Statement

Our goal is to provide personalized diagnosis predictions based on EHR via modeling disease severity, interaction, and progression in an end-to-end framework (shown in Figure 2).

We first denote the disease dataset as $\mathcal{D} = \{d_1, d_2, \ldots, d_N\}$, where N denotes the number of diseases. For a patient $u_i \in \mathcal{U}$, we denote a sequence of historical visits with timestamps as

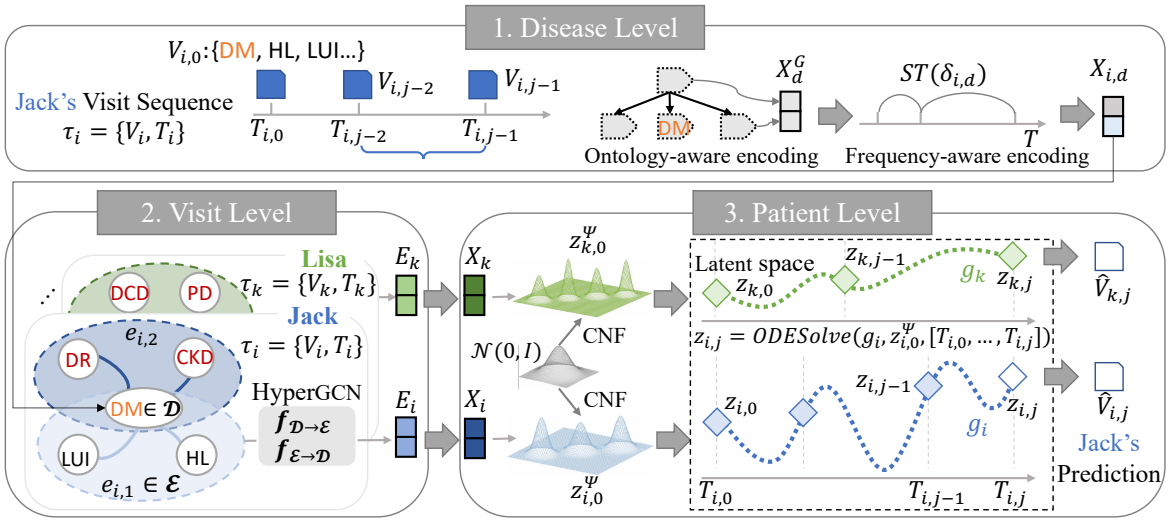


Fig. 2: The overall framework of ProCare.

$\tau_i = \{\boldsymbol{V}_{i,a}, T_{i,a}\}_{a=0}^{M_i-1}$, where M_i denotes the number of visits of u_i. $\boldsymbol{V}_{i,a} \in \{0,1\}^N$ is a multi-hot column vector for the diagnoses in the a-th visit and $T_{i,a}$ is the diagnosed timestamp. The task of personalized healthcare prediction for each patient u_i is to predict the last diagnosed diseases in $\boldsymbol{V}_{i,j}$ at $T_{i,j}$ based on previous clinical visits of u_i, i.e., $\{(\boldsymbol{V}_{i,0}, T_{i,0}), (\boldsymbol{V}_{i,1}, T_{i,1}), \ldots, (\boldsymbol{V}_{i,j-1}, T_{i,j-1})\}$.

B. Disease Level

The disease severity is determined by disease subtypes and patient status. Our idea here is to combine the information about disease subtypes from known disease ontology and weak signals of patient status regarding visiting frequency to provide informative bases in disease initial representation, which can be further leveraged by the subsequent model to learn personalized disease severity.

Since the official ICD-9 disease ontology provides hierarchical classifications of medical codes into different types of (sub)diseases, it can be naturally adopted to learn the embedding of each disease [10], [18]. To fully capture the hierarchical relations among diseases, we propose to initialize our personalized disease embedding $\boldsymbol{X}_{i,d}$ via recursively concatenating the embedding of each sub-disease to its parent from the leaf level of the hierarchy. Thus, we define the ontology-aware encoding $\mathbf{X}_d^G \in \mathbb{R}^{D\times 1}$ as follows:

$$\mathbf{X}_d^G = \mathbf{X}_d^{G,1} || \mathbf{X}_d^{G,2} || \ldots || \mathbf{X}_d^{G,k} || \ldots || \mathbf{X}_d^{G,K}, \tag{1}$$

where $||$ denotes the concatenation operation and $\mathbf{X}_d^{G,k}$ denotes the ontology-aware representation in the k-th level. K is a hyperparameter that we empirically set to 4 by default. D is the embedding dimension.

Without risking the exposure of patients' privacy, we propose to leverage the visit times of each patient and compute a frequency-aware encoding. The insight is: the shorter the average time of revisit, the more help the patient needs from the doctor, and the worse status the patient might have. As shown in Figure 1, to measure different severities of DM for Lisa and Jack, we propose to extract their visits that include DM in their historical visits. Since Jack was diagnosed two times with DM (i.e., 05/29/2017 and 02/12/2017) while Lisa has three times (i.e., 04/11/2018, 07/12/2017, and 01/10/2017 for Lisa), $\delta_{i,d}$ for Jack's DM is 106 days and the one for Lisa is 228 days.

With the definition of average revisit time $\delta_{i,d}$, we can obtain the frequency-aware encoding $ST(\delta_{i,d})$ as follows:

$$ST(\delta_{i,d})_k = \sin\left(\frac{\delta_{i,d} * k}{T_{i,m} * D}\right), \; ST(\delta_{i,d})_{k+D} = \cos\left(\frac{\delta_{i,d} * k}{T_{i,m} * D}\right), \tag{2}$$

where $0 \le k < D$ denotes the k-th order of dimension. By mapping each average revisit time value into a vector, we obtain the embedding of $\delta_{i,d}$, which can reflect patient u_i's personalized health status, and ensure similar values (average time of revisit) can be embedded into similar embeddings.

Based on the ontology-aware encoding $\boldsymbol{X}_d^G$ and the frequency-aware encoding $ST(\delta_{i,d})$, we can integrate them to obtain $\boldsymbol{X}_{i,d}$, which denotes the personalized initial representation of disease d on patient u_i as follows:

$$\boldsymbol{X}_{i,d} = \sigma\left(\boldsymbol{W}_t\left[\boldsymbol{X}_d^G \| ST(\delta_{i,d})\right]\right) \in \mathbb{R}^{D\times 1}, \tag{3}$$

where $\mathbf{W}_t \in \mathbb{R}^{D\times 3D}$ is a learnable weight matrix.

C. Visit Level

Past clinical studies [13], [25] have mainly focused on repeated interaction with a single disease, where they select one disease to observe across patients with different health status, and the task is to predict this disease. However, since a visit can contain multiple diseases and disease can interact with several diseases simultaneously, the above interactions are insufficient to characterize the complex relations within a diagnosis. As shown in Figure 1, the disease CKD can be affected by the first-order interaction from other diseases (e.g., DM $\to$ CKD and DR $\to$ CKD), the second-order interaction (e.g., DM $\times$ HL $\to$ CKD), as well as the third-order interaction (e.g., DM $\times$ HL $\times$ LUI $\to$ CKD). The possible number of higher-order interactions grows exponentially as the size of a visit increases, leading to large computational burdens.

To properly model the disease interaction inside one visit, we propose to regard each patient as a hypergraph (each visit as a hyperedge) and adopt hypergraph convolutional networks (HyperGCN) to obtain visit representations. Specifically, we define a two-stage aggregation transformation upon the hypergraph structure, where each hyperedge denotes the patient's visit and each node denotes the diagnosed diseases. The processes of the HyperGCN at the l-th layer are formulated as follows:

$$\boldsymbol{E}_{i,j}^{l} = f_{\mathcal{D}\to\mathcal{E}}(\{\mathbf{X}_{i,d}^{l-1}\}_{d\in e_{i,j}}), \mathbf{X}_{i,d}^{l} = f_{\mathcal{E}\to\mathcal{D}}(\{\boldsymbol{E}_{i,j}^{l}\}_{e_{i,j}\in\mathcal{E}_d}). \tag{4}$$

Here $\boldsymbol{E}_{i,j}$ and $\boldsymbol{X}_{i,d}$ stand for patient u_i's embeddings of hyperedge j and node d, respectively. $\{\mathbf{X}_{i,d}\}_{d\in e_{i,j}}$ is the hidden representation of node that is contained in the hyperedge $e_{i,j}$, and $\{\boldsymbol{E}_{i,j}\}_{e_{i,j}\in\mathcal{E}_d}$ is the hidden representations of hyperedges that contain the node d.

To ensure expressive representations of sets and identify the most relevant elements within the set for message passing, we adopt the self-attention function [9], [20], [21] for both $f_{\mathcal{D}\to\mathcal{E}}(\cdot)$ and $f_{\mathcal{E}\to\mathcal{D}}(\cdot)$. Thus, we have

$$f_{\mathcal{D}\to\mathcal{E}}(\boldsymbol{S}) = f_{\mathcal{E}\to\mathcal{D}}(\boldsymbol{S}) = \text{Self-Att}(\boldsymbol{S}), \tag{5}$$

where $\boldsymbol{S}$ is the embedding of the input set. In this way, we can obtain the visit representation $\boldsymbol{E}_{i,j}^{l}$ by setting $\boldsymbol{S} = \{\mathbf{X}_{i,d}^{l-1}\}_{d\in e_{i,j}}$ and obtain the disease embedding $\mathbf{X}_{i,d}^{l}$ by setting $\boldsymbol{S} = \{\mathbf{X}_{i,d}^{l-1}\}_{d\in e_{i,j}}$, so as to perform different stages of aggregation, respectively. The mathematical formulation of self-attention Self-Att is written as

$$\text{Self-Att}(\boldsymbol{S}) = \text{softmax}\left((\boldsymbol{S}\mathbf{W}^{Q})(\boldsymbol{S}\mathbf{W}^{K})^{T}/\sqrt{D}\right)\boldsymbol{S}\mathbf{W}^{V}, \tag{6}$$

where the matrices $\mathbf{W}^{Q}$, $\mathbf{W}^{K}$, and $\mathbf{W}^{V}$ are learnable.

D. Patient Level

With the visit representations $\boldsymbol{E}$ learned, it is important to obtain accurate patient representation via further aggregating the visit information. However, two challenges remain when modeling a patient's latent disease progression. 1) Discreteness: Although the disease progressions should be continuous in nature [16], we can only observe patients' visits at discrete timestamps with irregular intervals in real EHR data. 2) Incompleteness: Patients can have multiple diagnosed diseases at each visit, and some diseases can be missed since they are not in the primary diagnosis recorded by doctors. Therefore, existing algorithms that only consider discrete visits and assume all diagnoses are recorded can even backfire in personalized diagnosis predictions.

To address the above two challenges, we propose a personalized generative model based on neural ordinary differential equations for capturing patients' continuous-time trajectories of disease progression. As shown in the bottom right of Figure 2, rather than modeling Jack via aggregating three blue marks, we can represent his health status through $\boldsymbol{z}_{i,j}$ at $T_{i,j}$ on the smooth blue trajectory and predict the $\boldsymbol{V}_{i,j}$ for Jack.

Specifically, we fisrt propose to initialize them via aggregating patient's historical visit embeddings (i.e., $\mathbf{E}_i = (\mathbf{E}_{i,0}, \mathbf{E}_{i,1}, \ldots, \mathbf{E}_{i,j-1})$) via a widely adopted self-attention mechanism. In this way, the visit-driven representation $\boldsymbol{z}_{i,0}$ of patient i is calculated through:

$$\boldsymbol{z}_{i,0} = \text{Self-Att}(\boldsymbol{E}_i), \tag{7}$$

where the calculation of $\text{Self-Att}(\mathcal{S})$ can be referred in Eq. 6 in Section III-C and we can simple replace $\mathcal{S}$ by $\boldsymbol{E}_i$.

Furthermore, to capture the patient's latest health status and consider the latent disease progress, we propose to infer the patient u_i's representation $\boldsymbol{z}_{i,j}$ at time $T_{i,j}$ by solving the neural ordinary differential equations:

$$\begin{aligned}\boldsymbol{z}_{i,j} &= \boldsymbol{z}_{i,0} + \int_{t=T_{i,0}}^{T_{i,j}} g_i\left(\boldsymbol{z}_{i,t}\right) dt \\ &= \text{ODESolve}_{\eta}\left(g_i, \mathrm{z}_{i,0}, [T_{i,0}, T_{i,1}, \ldots, T_{i,j}]\right),\end{aligned} \tag{8}$$

where g_i is a learnable neural network for each patient u_i and we adopt the fourth-order Runge–Kutta for an accurate and efficient approximation [2], [7]. η is a hyperparameter to control the step size of the network g_i.

Since personalized diagnosis prediction is a multi-label classification task, we use a dense layer with a softmax function to calculate the predicted probability. Specifically, the prediction $\hat{\mathbf{V}}_{i,j}$ is based on the inferred patient status $z_{i,j}$ and $\mathbf{V}_{i,j}$ is the ground-truth of patient u_i's (j+1)-th diagnosis. The predictive objective function $\mathcal{L}_P$ are listed as follows:

$$p(\hat{\mathbf{V}}_{i,j}|\tau_i, \boldsymbol{z}_{i,0}) = \text{softmax}(\text{MLP}(\mathrm{z}_{i,j})), \tag{9}$$

Finally, the overall objective function of ProCare is:

$$\max \sum_{i=1}^{N} \mathbf{V}_{i,j} \log\left(\hat{\mathbf{V}}_{i,j}\right) + (1-\mathbf{V}_{i,j}) \log\left(1-\hat{\mathbf{V}}_{i,j}\right). \tag{10}$$

IV. Experiment

In this section, we evaluate our proposed ProCare framework focusing on the following three research questions:

- **RQ1:** How does ProCare perform in comparison to state-of-the-art diagnosis prediction methods?
- **RQ2:** What are the effects of different model components?
- **RQ3:** How do the hyperparameters affect the prediction performance and how to choose optimal values?

TABLE I: Statistics of the datasets used in our experiments.

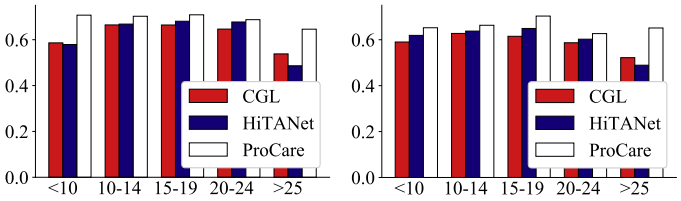
Dataset	MIMIC-III	NELL
# of patients	2,371	3,907
# of visits	7,279	68,969
Avg. visits per patient	3.07	17.65
# of unique ICD9 codes	4,880	5,630
Avg. # of diagnosis codes per visit	13.39	2.25
Max # of diagnosis codes per visit	39.0	34.0

A. Experimental Setup

1) Datasets and Evaluation Protocols.: We use two real-world EHR datasets to verify the effectiveness of compared methods, i.e., MIMIC-III [6] and NELL. NELL is a large-scale real-world clinical data collected by the Nell Hodgson Woodruff School of Nursing at Emory University. Both datasets are fully anonymized and carefully sanitized before our access. We chose patients who made at least three visits for both datasets. Both datasets have K=4 levels in the ICD-9 hierarchy. The statistics are summarized in Table I. For evaluation metrics, we use Recall@k and NDCG@k that are consistent with [9], [10], [22].

TABLE II: Experimental results on two benchmark EHR datasets. The best performances are highlighted in boldface and the second runners are underlined, where * denotes significant improvements based on the Wilcoxon signed-rank test.

Method	Recall@5	NDCG@5	Recall@10	NDCG@10	Recall@5	NDCG@5	Recall@10	NDCG@10
	MIMIC-III				NELL			
RATAIN	0.1510±0.18%	0.4188±0.16%	0.2134±0.13%	0.3537±0.12%	0.6272±0.21%	0.5974±0.16%	0.7535±0.16%	0.6227±0.13%
Dipole	0.1442±0.24%	0.3999±0.18%	0.2038±0.28%	0.3378±0.18%	0.5989±0.21%	0.5705±0.18%	0.7195±0.17%	0.5946±0.15%
GRAM	0.1429±0.13%	0.4059±0.10%	0.2112±0.14%	0.3510±0.12%	0.6394±0.15%	0.6118±0.12%	0.7277±0.16%	0.6325±0.13%
Timeline	0.1487±0.15%	0.4123±0.13%	0.2100±0.12%	0.3482±0.10%	0.6174±0.15%	0.5881±0.15%	0.7417±0.16%	0.6129±0.13%
KAME	0.1353±0.14%	0.3992±0.13%	0.2055±0.13%	0.3070±0.11%	0.5620±0.12%	0.5353±0.10%	0.6751±0.15%	0.5579±0.13%
MHM	0.1383±0.14%	0.4080±0.13%	0.2128±0.13%	0.3481±0.10%	0.5745±0.15%	0.5472±0.12%	0.6902±0.13%	0.5704±0.11%
TAdaNet	0.1433±0.19%	0.4114±0.15%	0.2172±0.17%	0.3568±0.13%	0.5952±0.16%	0.5669±0.15%	0.7150±0.12%	0.5909±0.10%
HiTANet	0.1502±0.21%	0.4166±0.17%	0.2122±0.18%	0.3518±0.16%	0.6446±0.18%	0.6186±0.15%	0.7701±0.15%	0.6502±0.12%
CGL	0.1538±0.22%	0.4265±0.19%	0.2173±0.26%	0.3602±0.21%	0.6387±0.18%	0.6084±0.15%	0.7673±0.13%	0.6341±0.10%
Chet	0.1636±0.13%	0.4403±0.08%	0.2312±0.12%	0.3719±0.10%	0.6182±0.15%	0.5913±0.12%	0.7381±0.13%	0.6181±0.13%
ProCare	**0.1885±0.18%***	**0.5004±0.16%***	**0.2687±0.21%***	**0.4271±0.18%***	**0.6940±0.22%***	**0.6632±0.21%***	**0.8101±0.17%***	**0.6863±0.16%***



(a) Recall@5 on NELL (b) NDCG@5 on NELL

Fig. 3: Performance with different lengths of visit sequence.

TABLE III: Ablation analysis of our proposed ProCare.

Method	Recall@5	NDCG@5	Recall@10	NDCG@10
	MIMIC-III			
ProCare w/o. DL	0.1868	0.4538	0.2660	0.4017
ProCare w/o. VL	0.1864	0.4646	0.2612	0.4003
ProCare w/o. PL	0.1810	0.4379	0.2607	0.3830
ProCare	**0.1885**	**0.5004**	**0.2687**	**0.4271**
	NELL			
ProCare w/o. DL	0.6907	0.6600	0.7948	0.6828
ProCare w/o. VL	0.6823	0.6570	0.7842	0.6731
ProCare w/o. PL	0.6619	0.6443	0.7632	0.6661
ProCare	**0.6940**	**0.6632**	**0.8101**	**0.6863**

2) Methods for Comparison: We adopt 10 representative state-of-the-art methods as baselines for the performance comparison with ProCare: (1) interaction modeling methods: **GRAM** [4], **KAME** [14], **MHM** [17], **TAdaNet** [19], and **CGL** [10]; (2) dynamic modeling methods: **RETAIN** [3], **Dipole** [12], **Timeline** [1], **HiTANet** [11], and **Chet** [9].

3) Implementation Details: We split dataset randomly according to patients into training/validation/test sets (i.e., 2100/61/210 on MIMIC-III and 3125/391/391 on NELL), which is consistent with [9], [10]. We optimize the compared baselines with standard Adam and tune all hyperparameters on training sets through grid search. In particular, L in $\{1, 2, 3, 4\}$, η in $\{0.001, 0.05, 0.1, 0.5\}$. We set the embedding dimension D as 128 and the batch size as 128 for all compared methods on MIMIC-III and NELL datasets. We carefully tune the hyperparameters of baselines as suggested in the original papers to achieve their best performance.

B. Overall Performance Comparison (RQ1)

We compare the personalized diagnosis prediction results of the proposed ProCare framework to those of the baseline models. Table II shows the Recall@k and NDCG@k on MIMIC-III and NELL datasets with k=$\{5, 10\}$. We have the following observations.

In general, ProCare outperforms all 10 baselines across all evaluation metrics on both datasets. It gains a substantial lead over the second-best models, Chet and HiTANet, with performance improvements ranging from 5.19% in Recall@10 on NELL to a significant 16.24% on MIMIC-III.

In particular, compared with Chet, ProCare exploits hypergraphs for better modeling the interactive disease data, which can mine the joint impact of two or more diseases on patients. Therefore, ProCare outperforms Chet by up to 16.24% in Recall@10 on MIMIC-III and up to 12.26% in Recall@5 on NELL. Compared with HiTAnet, ProCare can further capture the continuous disease progression between irregular clinical visits and complicated relations among syndromic diseases. Therefore, ProCare can outperform HiTANet by up to 26.60% in Recall@10 on MIMIC-III and up to 7.66% on NELL.

Moreover, Figure 3 shows that ProCare consistently outperforms CGL and HiTANet across varying visit sequence lengths on the NELL dataset. Notably, when the sequence length exceeds 25, ProCare excels by up to 33.14% and 24.77% over HiTANet and CGL respectively. This demonstrates ProCare's effectiveness in long-term clinical predictions, crucial for tracking chronic disease progression.

C. Model Ablations (RQ2)

To better understand our proposed techniques, i.e., personalized disease representation learning (Disease Level), higher-order interaction modeling of diseases (Visit Level), and continuous patients' latent health status learning with irregular observed visits (Patient Level), we study ProCare from three pivotal levels (shown in Table III) as follows:

Compared with ProCare w/o. DL, ProCare leads to performance gains ranging from 0.48% (achieved in Recall@5 on NELL) to 10.26% (achieved in NDCG@5 on MIMIC-III). Since different disease severity can cause heterogeneous disease interaction and varied progression rates, ProCare w/o. DL has a decrease in performance, especially the NDCG metric on the MIMIC-III dataset.

Furthermore, the performance gains of ProCare over ProCare w/o. VL ranges from 0.94% (with NDCG@5 on NELL) to 7.71% (with NDCG@5 on MIMIC-III), where ProCare explicitly models higher-order disease interaction among two

TABLE IV: Hyperparameter Studies.

Param.	Recall@5	NDCG@5	Recall@5	NDCG@5
	MIMIC-III		NELL	
L = 1	0.1871	0.4973	0.6732	0.6533
L = 2	**0.1885**	**0.5004**	0.6801	0.6561
L = 3	0.1869	0.4970	**0.6940**	**0.6632**
L = 4	0.1835	0.4790	0.6921	0.6615
η = 0.01	0.1845	0.4911	0.6912	0.6598
η = 0.05	0.1872	0.4896	**0.6940**	**0.6632**
η = 0.1	**0.1885**	**0.5004**	0.6887	0.6571
η = 0.5	0.1828	0.4864	0.6762	0.6517

or more diseases inside a visit and further improves the performance of personalized diagnosis prediction.

The ProCare outperforms ProCare w/o. PL with gains ranging from 2.93% to 14.27% in NDCG@5 on different datasets. This is mainly because the PL module helps capture latent continuous-time dynamics in patient data. The results also affirm the effectiveness of our novel generative model for handling incomplete patient visits.

D. Effect of Hyperparameters (RQ3)

As shown in Table IV, ProCare attains optimal performance with $L = 2$ layers on MIMIC-III and $L = 3$ on NELL, varying with the average number of visits in each dataset. Exceeding these optimal layer counts causes over-smoothing issues, leading to performance degradation on both datasets. η controls the step size of the ordinary differential equations, which denotes the sampling frequency of the personalized ordinary differential equations generation model. The optimal η value on MIMIC-III is about 0.1, and the optimal η value on NELL is about 0.05. Since the average number of visits is larger on NELL, it's reasonable to use a smaller sampling frequency for modeling the health status trajectory of patients.

V. Conclusion

In this paper, we propose to make personalized diagnosis predictions based on the EHR data of patients. Specifically, we propose a novel personalized diagnosis model (ProCare) with three pivotal levels, which capture the higher-order disease interactions and continuous-time disease progression based on severity-driven disease encodings. Extensive quantitative experiments demonstrate the clear advantages of our ProCare over state-of-the-art towards diagnosis prediction, especially for long visit sequences.

VI. Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 6230071268) and the Natural Science Foundation of Zhejiang Province (LQ23F020007). Carl Yang was supported by the National Institute Of Diabetes And Digestive And Kidney Diseases of the National Institutes of Health under Award Number K25DK135913.

References

[1] T. Bai, S. Zhang, B. L. Egleston, and S. Vucetic. Interpretable representation learning for healthcare via capturing disease progression through time. In *SIGKDD*, pages 43–51, 2018.

[2] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations. *arXiv preprint arXiv:1806.07366*, 2018.

[3] E. Choi, M. T. Bahadori, J. A. Kulas, A. Schuetz, W. F. Stewart, and J. Sun. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In *NeurIPS*, 2016.

[4] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun. Gram: graph-based attention model for healthcare representation learning. In *SIGKDD*, pages 787–795, 2017.

[5] J. C. Ho, J. Ghosh, and J. Sun. Marble: high-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization. In *SIGKDD*, pages 115–124, 2014.

[6] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-Wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 2016.

[7] Y. Li, H. Yi, C. Bender, S. Shan, and J. B. Oliva. Exchangeable neural ode for set modeling. *NeurIPS*, 33:6936–6946, 2020.

[8] C. Liu, F. Wang, J. Hu, and H. Xiong. Temporal phenotyping from longitudinal electronic health records: A graph based framework. In *SIGKDD*, pages 705–714, 2015.

[9] C. Lu, T. Han, and Y. Ning. Context-aware health event prediction via transition functions on dynamic disease graphs. In *AAAI*, volume 36, pages 4567–4574, 2022.

[10] C. Lu, C. K. Reddy, P. Chakraborty, S. Kleinberg, and Y. Ning. Collaborative graph learning with auxiliary text for temporal event prediction in healthcare. In *IJCAI*, 2021.

[11] J. Luo, M. Ye, C. Xiao, and F. Ma. Hitanet: Hierarchical time-aware attention networks for risk prediction on electronic health records. In *SIGKDD*, pages 647–656, 2020.

[12] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In *SIGKDD*, pages 1903–1911, 2017.

[13] F. Ma, J. Gao, Q. Suo, Q. You, J. Zhou, and A. Zhang. Risk prediction on electronic health records with prior medical knowledge. In *SIGKDD*, pages 1910–1919, 2018.

[14] F. Ma, Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao. Kame: Knowledge-based attention model for diagnosis prediction in healthcare. In *CIKM*, 2018.

[15] L. Ma, C. Zhang, Y. Wang, W. Ruan, J. Wang, W. Tang, X. Ma, X. Gao, and J. Gao. Concare: Personalized clinical feature embedding via capturing the healthcare context. In *AAAI*, 2020.

[16] Z. Qian, W. R. Zame, M. van der Schaar, L. M. Fleuren, and P. Elbers. Integrating expert odes into neural odes: Pharmacology and disease progression. In *NeurIPS*, 2021.

[17] Z. Qiao, Z. Zhang, X. Wu, S. Ge, and W. Fan. Mhm: Multi-modal clinical data based hierarchical multi-label diagnosis prediction. In *SIGIR*, pages 1841–1844, 2020.

[18] L. Song, C. W. Cheong, K. Yin, W. K. Cheung, B. C. Fung, and J. Poon. Medical concept embedding with multiple ontological representations. In *IJCAI*, 2019.

[19] Q. Suo, J. Chou, W. Zhong, and A. Zhang. Tadanet: Task-adaptive network for graph-enriched meta-learning. In *SIGKDD*, pages 1789–1799, 2020.

[20] M. Usama, B. Ahmad, W. Xiao, M. S. Hossain, and G. Muhammad. Self-attention based recurrent convolutional neural network for disease prediction using healthcare data. *Computer methods and programs in biomedicine*, 190:105191, 2020.

[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *NeurIPS*, 30, 2017.

[22] Z. Wang, R. Wen, X. Chen, S. Cao, S.-L. Huang, B. Qian, and Y. Zheng. Online disease diagnosis with inductive heterogeneous graph convolutional networks. In *WWW*, pages 3349–3358, 2021.

[23] R. Xu, Y. Yu, C. Zhang, M. K. Ali, J. C. Ho, and C. Yang. Counterfactual and factual reasoning over hypergraphs for interpretable clinical predictions on ehr. In *Machine Learning for Health*, pages 259–278. PMLR, 2022.

[24] K. Yang, X. Li, H. Liu, J. Mei, G. Xie, J. Zhao, B. Xie, and F. Wang. Tagited: Predictive task guided tensor decomposition for representation learning from electronic health records. In *AAAI*, volume 31, 2017.

[25] M. Ye, S. Cui, Y. Wang, J. Luo, C. Xiao, and F. Ma. Medpath: Augmenting health risk prediction via medical knowledge paths. In *WWW*, 2021.

[26] Y. Zheng, S. H. Ley, and F. B. Hu. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nature reviews endocrinology*, 14(2):88–98, 2018.

[27] J. Zhou, F. Wang, J. Hu, and J. Ye. From micro to macro: data driven phenotyping by densification of longitudinal electronic medical records. In *SIGKDD*, pages 135–144, 2014.