SHIELD: Semantic Hierarchy-Enhanced Hypergraph Learning for Diagnosis Prediction

Xusheng Yu[1], Hang Lv[1], Guofang Ma[2], Yanchao Tan[1†], Xing Chen[1], Carl Yang[3]
[1] Fuzhou University, Fuzhou, China
[2] Zhejiang Gongshang University, Hangzhou, China
[3] Emory University, Atlanta, GA, United States
exungsh@foxmail.com, lvhangkenn@gmail.com, {yctan, chenxing}@fzu.edu.cn,
maguofang@zjgsu.edu.cn, j.carlyang@emory.edu

***Abstract*—Electronic Health Records (EHRs) provide vital longitudinal data for clinical decision support. However, existing methods often fail to unify semantic richness with hierarchical disease structures and struggle to capture complex higher-order interactions beyond pairwise modeling. We propose `SHIELD`, which integrates semantic quantization with dual-view hypergraph learning. Specifically, we employ residual quantization to derive a semantic hierarchy from disease embeddings for interpretable modeling. A dual-view hypergraph convolution module then captures many-to-many clinical dependencies by evolving patient-disease and disease-patient hypergraphs in parallel. Finally, a GRU-based attention mechanism captures temporal dynamics for diagnosis prediction. Experiments on MIMIC-III and eICU datasets show that `SHIELD` significantly outperforms state-of-the-art baselines, improving Precision@10 by up to 17.83% and 9.10%, respectively.**

***Index Terms*—Diagnosis prediction, Hierarchical disease representation, Residual Quantization, Hypergraph Neural Networks, Electronic Health Records.**

I. Introduction

The digitalization of healthcare has established Electronic Health Records (EHRs) as a cornerstone for personalized clinical decision support. EHRs provide longitudinal and heterogeneous patient data, the modeling of which is vital for tasks like diagnosis prediction. Recent deep learning advances have utilized RNNs for temporal sequences [1], [2], CNNs for local patterns [3], attention mechanisms for interpretability [4], and pre-trained language models for semantic enrichment [5], [6]. Despite this, two fundamental challenges remain that limit predictive performance and model interpretability:

Challenge 1: Constructing a unified disease representation that integrates semantic richness and hierarchical structure. Disease concepts like *Hypertension* and *Diabetes Mellitus* exhibit both semantic proximity (e.g., shared insulin resistance mechanisms [7]) and taxonomic hierarchy. Semantic-based models [5] identify contextual similarities (Fig. 1(a), left) but produce flat representations that fail to distinguish parent-child relations (e.g., *Diabetes* vs. *Type 2 Diabetes*), hindering structured reasoning. Hierarchy-based models [8]–[10] rely on static ICD trees (Fig. 1(a), right), which strictly follow taxonomic branches but overlook crucial cross-branch clinical correlations (e.g., *Hypertension* and *Type 2 Diabetes* occupy distal branches despite frequent co-occurrence). Thus, a unified mechanism is needed to integrate semantic richness with adaptive hierarchical modeling.

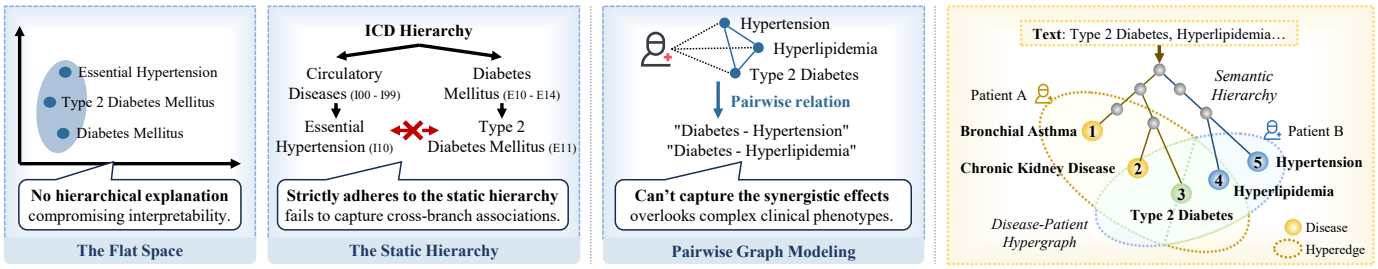
Challenge 2: Modeling higher-order interactions among patient and disease. EHR data involves many-to-many interactions where multiple diagnoses frequently co-occur within a single visit. Traditional graph methods (Fig. 1(b)) simplify these into independent pairwise links (e.g., *diabetes-hypertension*), fragmenting the clinical context and failing to capture synergistic effects characteristic of complex phenotypes like *metabolic syndrome*. Furthermore, static representations ignore the mutual refinement between patient cohorts and disease patterns. Moving from fragmented pairwise links to holistic higher-order interaction modeling is essential to capture deep clinical associations across the entire population.

To address these challenges, we propose `SHIELD` (**S**emantic **HI**erarchy-**E**nhanced Hypergraph **L**earning for **D**iagnosis Prediction). As shown in Fig. 1(c), `SHIELD` comprises three modules: **(1) Hierarchical Semantic Quantization (HSQ):** Inspired by RQ-VAE [11], HSQ uses residual quantization to map flat embeddings into tree-like discrete codes, capturing both broad categories and specific subtypes without predefined rules. **(2) Dual-View Hypergraph Convolution (DHC):** DHC constructs complementary Patient-Disease and Disease-Patient hypergraphs to model high-order comorbidity and phenotypic cohorts, refining representations via parallel convolutions. **(3) Temporal Attention Predictor (TAP):** TAP integrates refined features into a GRU with location-based attention to capture longitudinal trajectories for clinical prediction.

Our main contributions are: (1) **Unified Framework**: A novel architecture that unifies semantic hierarchy construction with many-to-many patient-disease interaction modeling. (2) **Effective Design**: An efficient fusion of RQ-VAE and dual-view hypergraphs that balances predictive performance with clinical interpretability. (3) **Empirical Results**: Extensive experiments on real-world EHR datasets where `SHIELD` con-

† Yanchao Tan is the corresponding author.

This work was supported in part by the National Natural Science Foundation of China under Grants (62302098), Fujian Provincial Natural Science Foundation of China under Grants (2025J01540), and Fujian Provincial Artificial Intelligence Industry Development Technology Project under Grant (2025H0042). Carl Yang was not supported by any funds from China.



(a) Challenge 1: Constructing a Unified Disease Representation that Integrates Semantic Richness and Hierarchical Structure

(b) Challenge 2: Modeling Higher-Order Interactions among Patient and Disease

(c) Jointly Learns Semantic Hierarchy Representations and Higher-Order Patient-Disease Interactions

Fig. 1. (a) Flat semantic spaces lack hierarchical structure, and the static ICD tree offers limited flexibility. (b) Pairwise graph models are insufficient to capture synergistic effects among multiple co-occurring diseases. (c) SHIELD leverages quantization to construct a learnable and interpretable disease hierarchy, and employs hypergraphs to model higher-order patient-disease relations.

sistently outperforms state-of-the-art baselines, particularly in low-resource settings.

II. Related Work

A. Diagnosis Prediction

Predictive modeling using Electronic Health Records (EHRs) is fundamental to clinical decision support and risk prediction. Early research centered on capturing longitudinal dependencies through recurrent architectures, notably Doctor AI [1] and RETAIN [2]. Subsequent works [12]–[14] further advanced temporal modeling by integrating attention mechanisms, memory networks, and transformer architectures.

Beyond temporal dynamics, recent studies have shifted toward structural learning to model medical concept relationships. Graph-based approaches [15]–[17] leverage disease ontologies and co-occurrence patterns to learn structured representations via Graph Neural Networks (GNNs). Despite their success, these methods remain sensitive to node embedding quality and mainly model pairwise relations, limiting their ability to capture higher-order clinical interactions. To address this, we propose a structured representation framework based on residual quantization to model hierarchical semantics and complex disease correlations in EHR data.

B. Residual Quantization

Vector Quantization (VQ) facilitates compact and interpretable representations by mapping continuous latent vectors into discrete codebooks. To enhance quantization capacity, Residual Quantization (RQ) progressively approximates high-dimensional vectors by quantizing residual errors across multiple stages [18]. Compared to traditional VQ, RQ excels at modeling complex feature distributions and has been widely adopted in computer vision [19], [20] and recommendation systems [11], [21] to generate high-quality semantic IDs.

Despite these advances, the application of RQ in healthcare remains underexplored. Conventional EHR-based models primarily rely on continuous disease embeddings, which fail to explicitly capture the inherent hierarchical taxonomies of medical codes (e.g., ICD chapters $\rightarrow$ blocks $\rightarrow$ categories). RQ's multi-stage decomposition naturally aligns with this medical hierarchy, providing a principled framework to bridge continuous latent features with discrete medical structures.

III. Method

A. Problem Statement

Let $\mathcal{P} = \{p_n\}_{n=1}^{N}$ denote a cohort of N patients. The medical history of patient p_n is represented as a longitudinal sequence of T_n visits: $\mathcal{V}_n = (v_{(n,1)}, \ldots, v_{(n,T_n)})$. Each visit $v_{(n,t)}$ consists of a set of diagnoses $\mathcal{D}_{n,t} = \{d_{(n,t,m)}\}_{m=1}^{M_{(n,t)}}$, where $d_{(n,t,m)} \in \mathcal{D}$ belongs to the universe of all medical conditions. To facilitate clinical analysis, these diagnoses are mapped to their corresponding hierarchical Clinical Classifications Software (CCS) categories. Given the historical sequence $\mathcal{V}_n$, the primary objective of SHIELD is to predict the set of relevant CCS codes for the patient's $(T_n + 1)$-th visit.

B. Hierarchical Semantic Quantization (HSQ)

1) RQ-VAE Pretraining: To obtain high-quality and interpretable disease representations, we employ a Residual Quantization Variational Autoencoder. This architecture leverages residual quantization to preserve fine-grained semantic information while explicitly capturing the hierarchical structure of diseases through discrete **Semantic IDs**. The RQ-VAE framework consists of three primary components: an encoder, a hierarchical residual-quantization module, and a decoder.

The encoder $\text{Encoder}_{\text{RQ}} : \mathbb{R}^{768} \rightarrow \mathbb{R}^{d}$ takes initial disease-name embeddings $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_M] \in \mathbb{R}^{M \times 768}$ from pre-trained BioBERT [5] as input, projecting them into a latent space via $\mathbf{Z} = \text{Encoder}_{\text{RQ}}(\mathbf{S}) = \mathbf{S}\mathbf{W}_e + \mathbf{b}_e \in \mathbb{R}^{M \times d}$. Here, $\mathbf{W}_e \in \mathbb{R}^{768 \times d}$ and $\mathbf{b}_e \in \mathbb{R}^{d}$ are the learnable weight matrix and bias term, respectively, and d is the latent dimension.

To prevent the codebook collapse issue in vector quantization [22] and to model the inherent taxonomy of diseases, we construct an L-layer residual codebook hierarchy $\{\mathcal{C}^{(l)}\}_{l=1}^{L}$ using hierarchical K-means clustering [11].

Specifically, for the first layer ($l = 1$), the codebook $\mathcal{C}^{(1)} = \{\boldsymbol{c}_k^{(1)}\}_{k=1}^{K}$ is initialized by performing K-means clustering directly on the latent representations $\boldsymbol{Z}$. For subsequent layers $l \in \{2, \ldots, L\}$, the codebook $\mathcal{C}^{(l)}$ is constructed by clustering the residuals from the preceding layer. Let $\boldsymbol{r}_i^{(l-1)}$ be the residual of the i-th disease after $l-1$ stages of quantization. The codebook for layer l is defined as:

$$\mathcal{C}^{(l)} = \{\boldsymbol{c}_k^{(l)}\}_{k=1}^{K} = \text{K-means}\big(\{\boldsymbol{r}_i^{(l-1)}\}_{i=1}^{M},\ K\big). \quad (1)$$

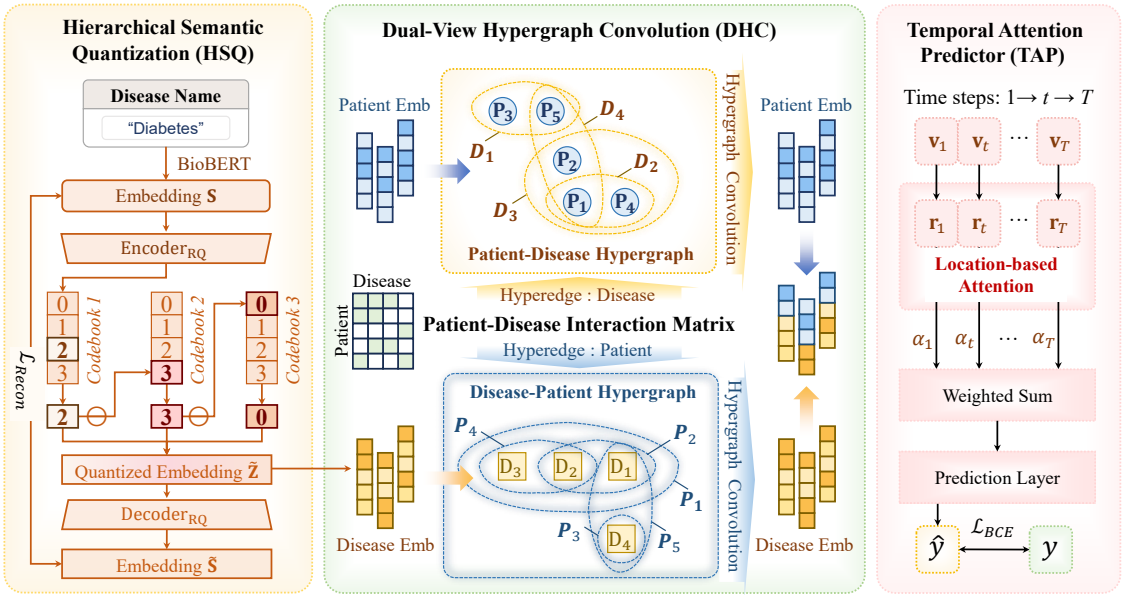


Fig. 2. The overall framework of SHIELD. (1) HSQ: Uses BioBERT and RQ-VAE to generate discrete, interpretable disease embeddings via codebook quantization. (2) DHC: Constructs dual hypergraphs (patient-disease & disease-patient) to model higher-order interactions, refined via hypergraph convolution. (3) TAP: Aggregates time-step representations using location-based attention, followed by a prediction layer optimized with BCE loss.

This progressive refinement ensures that each layer effectively captures the information gap left by its predecessor.

Formally, the residual quantization function $\mathrm{RQ}(\cdot)$ maps a latent vector $\boldsymbol{z}$ to a sequence of discrete indices and a reconstructed vector. For each disease, the quantization indices $\{q^{(l)}\}_{l=1}^{L}$ and residuals $\{\boldsymbol{r}^{(l)}\}_{l=0}^{L}$ are computed recursively:

$$\boldsymbol{r}^{(0)} = \boldsymbol{z}, \tag{2}$$

$$q^{(l)} = \arg\min_{k\in\{1,\dots,K\}} \|\boldsymbol{r}^{(l-1)} - \boldsymbol{c}_k^{(l)}\|_2, \tag{3}$$

$$\boldsymbol{r}^{(l)} = \boldsymbol{r}^{(l-1)} - \boldsymbol{c}_{q^{(l)}}^{(l)}. \tag{4}$$

The resulting tuple of indices $\mathcal{I} = [q^{(1)}, q^{(2)}, \dots, q^{(L)}]$ is defined as the **Semantic ID** of the disease. These IDs reflect the disease's position within the learned hierarchical structure, where $q^{(1)}$ represents a coarse-grained category and subsequent indices denote increasingly fine-grained sub-categories.

The quantized embedding $\tilde{\mathbf{z}}$ is reconstructed as $\tilde{\mathbf{z}} = \sum_{l=1}^{L} \mathbf{c}_{q^{(l)}}^{(l)}$, which is then fed into the decoder $\mathrm{Decoder}_{\mathrm{RQ}}$ to reconstruct the original space: $\tilde{\mathbf{S}} = \tilde{\mathbf{Z}}\mathbf{W}_d + \mathbf{b}_d \in \mathbb{R}^{M\times 768}$. To preserve semantic information, we minimize the reconstruction loss: $\mathcal{L}_{\mathrm{recon}} = \frac{1}{M}\sum_{i=1}^{M} \|\mathbf{s}_i - \tilde{\mathbf{s}}_i\|_2^2$.

To optimize the encoder, decoder, and codebooks jointly while minimizing multi-level residual errors, the quantization loss is defined as:

$$\mathcal{L}_{\mathrm{quant}} = \sum_{l=1}^{L} \left(\|\mathrm{sg}[\mathbf{r}^{(l-1)}] - \mathbf{c}_{q^{(l)}}^{(l)}\|_2^2 + \beta \|\mathbf{r}^{(l-1)} - \mathrm{sg}[\mathbf{c}_{q^{(l)}}^{(l)}]\|_2^2 \right) \tag{5}$$

where $\mathrm{sg}[\cdot]$ is the stop-gradient operator, and $\beta = 0.25$ balances codebook and encoder optimization [11]. Furthermore, to prevent codebook collisions and encourage diversity, we introduce a diversity loss [23]:

$$\mathcal{L}_{\mathrm{div}} = \frac{1}{M}\sum_{k=1}^{K} \left| \mathrm{count}_k - \frac{M}{K} \right| + \frac{1}{K(K-1)} \sum_{i\neq j} \|\mathbf{c}_i - \mathbf{c}_j\|_2^2 \tag{6}$$

where K is the codebook size per layer. The total pre-training objective is formulated as: $\mathcal{L}_{\text{RQ-VAE}} = \mathcal{L}_{\mathrm{recon}} + \mu\mathcal{L}_{\mathrm{quant}} + \lambda\mathcal{L}_{\mathrm{div}}$, with hyperparameters μ and λ controlling loss weights.

2) RQ-VAE Inference in Formal training: During downstream training, the latent disease representation is obtained using the pre-trained RQ-VAE encoder and codebooks. Each disease embedding is first encoded into a latent vector z_i, then mapped to the nearest discrete code index through vector quantization, and subsequently reconstructed as $\tilde{z}_i$. The final representation aggregates these quantized embeddings into a unified matrix $\boldsymbol{Q} = \{\tilde{z}_1, \dots, \tilde{z}_M\}$, which serves as input for the *Dual-View Hypergraph Convolution* module.

C. Dual-View Hypergraph Convolution (DHC)

1) Hypergraph Preliminaries: A hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ consists of vertices $\mathcal{V}$, hyperedges $\mathcal{E}$, and a diagonal weight matrix $\mathbf{W}$. It is characterized by an incidence matrix $\mathbf{H} \in \{0,1\}^{|\mathcal{V}|\times|\mathcal{E}|}$, where $h(v,e) = 1$ if $v \in e$ and 0 otherwise. Vertex and hyperedge degrees are defined as $d(v) = \sum_{e\in\mathcal{E}} w(e)h(v,e)$ and $\delta(e) = \sum_{v\in\mathcal{V}} h(v,e)$, respectively. $\mathbf{D}_v$ and $\mathbf{D}_e$ denote the corresponding diagonal degree matrices.

2) Dual-View Hypergraph Construction: From longitudinal EHR data, we construct dual hypergraphs to model patient similarity and disease comorbidity. Let $\mathcal{P}$ and $\mathcal{D}$ be the sets of patients and diseases.

Patient-Disease Hypergraph ($\mathcal{H}_{pd}$): We define nodes as patients $p \in \mathcal{P}$ and hyperedges as diseases $d \in \mathcal{D}$. The incidence matrix $\mathbf{H}_{pd} \in \{0,1\}^{|\mathcal{P}|\times|\mathcal{D}|}$ is set to $\mathbf{H}_{pd}[p][d] = 1$ if patient p was diagnosed with disease d, grouping phenotypically similar patients.

Disease-Patient Hypergraph ($\mathcal{H}_{dp}$): As the dual of $\mathcal{H}_{pd}$, we define diseases as nodes and patients as hyperedges, with incidence matrix $\mathbf{H}_{dp} = \mathbf{H}_{pd}^{\top}$. Each hyperedge in $\mathcal{H}_{dp}$ captures the full diagnosis history of a patient, modeling higher-order comorbidity patterns.

3) Dual-View Hypergraph Convolution: We propose a dual-view framework to evolve representations by capturing higher-order dependencies. We define the hypergraph convolution operator as:

$$\text{HGConv}(\mathbf{E}, \mathbf{H}, \mathbf{W}, \mathbf{\Theta}) = \sigma\left(\mathbf{D}_v^{-1/2}\mathbf{H}\mathbf{W}\mathbf{D}_e^{-1}\mathbf{H}^{\top}\mathbf{D}_v^{-1/2}\mathbf{E}\mathbf{\Theta}\right) \tag{7}$$

Patient-Centric Evolution ($\mathcal{H}_{pd}$): Let $\mathbf{E}_p^{(0)} \in \mathbb{R}^{N\times d}$ be the initial patient embeddings. We iteratively update the representations for $\hat{l} = 1, \ldots, \hat{L}$ to capture patient correlations:

$$\mathbf{E}_p^{(\hat{l})} = \text{HGConv}(\mathbf{E}_p^{(\hat{l}-1)}, \mathbf{H}_{pd}, \mathbf{W}_{pd}^{(\hat{l})}, \mathbf{\Theta}_p^{(\hat{l})}) \tag{8}$$

The final patient-cohort context is $\mathbf{Z}_p = \mathbf{E}_p^{(\hat{L})}$.

Disease-Centric Evolution ($\mathcal{H}_{dp}$): Concurrently, disease embeddings $\mathbf{E}_d^{(0)} = \mathbf{Q}$ are refined via $\mathcal{H}_{dp}$ to model comorbidity patterns:

$$\mathbf{E}_d^{(\hat{l})} = \text{HGConv}(\mathbf{E}_d^{(\hat{l}-1)}, \mathbf{H}_{dp}, \mathbf{W}_{dp}^{(\hat{l})}, \mathbf{\Theta}_d^{(\hat{l})}) \tag{9}$$

where $\mathbf{Z}_d = \mathbf{E}_d^{(\hat{L})}$ encodes high-order co-occurrence features across the population.

4) Cross-View Feature Fusion.: The final stage integrates the patient-level clinical context into the disease representations. We argue that a disease's representation should not only depend on its global comorbidity ($\boldsymbol{Z}_d$) but also on the specific characteristics of the patient population it affects ($\boldsymbol{Z}_p$).

Using the disease-patient incidence matrix $\boldsymbol{H}_{dp}$ as a spatial transformation operator, we aggregate the refined patient features $\boldsymbol{Z}_p$ into the disease space. The final disease representation $\boldsymbol{E}_d$ is formulated as:

$$\boldsymbol{E}_d = \boldsymbol{Z}_d + \boldsymbol{H}_{dp}\boldsymbol{Z}_p\boldsymbol{W}_{fusion} \tag{10}$$

where $\boldsymbol{W}_{fusion} \in \mathbb{R}^{d\times d}$ is a learnable weight matrix for feature alignment.

D. Temporal Attention Predictor (TAP)

1) Visit Representation and Diagnosis Prediction: For a patient p_n, we first aggregate the fused disease embeddings $\boldsymbol{E}_d^i$ within each visit to compute the visit embedding $\boldsymbol{v}_{(n,t)}$:

$$\boldsymbol{v}_{(n,t)} = \frac{1}{|\mathcal{D}_{(n,t)}|}\sum_{d_{(n,t,i)}\in\mathcal{D}_{(n,t)}} \boldsymbol{E}_d^i. \tag{11}$$

To capture temporal dependencies across multiple visits, we employ a Gated Recurrent Unit (GRU) to obtain the hidden state sequence $\boldsymbol{R}_n$:

$$\boldsymbol{R}_n = \text{GRU}(\boldsymbol{v}_{(n,1)}, \boldsymbol{v}_{(n,2)}, \ldots, \boldsymbol{v}_{(n,T_n)}) \in \mathbb{R}^{T_n\times h}, \tag{12}$$

TABLE I
STATISTICS OF THE EICU AND MIMIC-III DATASETS.

Dataset	eICU	MIMIC-III
# of patients	18,836	5,449
# of visits	46,116	14,141
Avg. # visits per patient	2.45	2.60
Max # visits per patient	26	29
# of unique diagnoses	763	3,874
# of CCS codes	172	285

where h is the hidden size. Subsequently, a location-based attention mechanism is applied to aggregate these temporal features into a global patient representation $\boldsymbol{o}_n$:

$$\boldsymbol{\alpha} = \text{softmax}\left(\boldsymbol{R}_n\boldsymbol{w}_\alpha\right), \quad \boldsymbol{o}_n = \boldsymbol{\alpha}^{\top}\boldsymbol{R}_n \in \mathbb{R}^h, \tag{13}$$

where $\boldsymbol{w}_\alpha \in \mathbb{R}^h$ is a learnable context vector.

Finally, $\boldsymbol{o}_n$ is passed through a dense layer with sigmoid activation to predict the probability $\hat{\boldsymbol{y}}$ of CCS codes. The diagnostic loss $\mathcal{L}_c$ is defined using binary cross-entropy:

$$\mathcal{L}_c = -\sum\left(\boldsymbol{y}\log\hat{\boldsymbol{y}} + (1-\boldsymbol{y})\log(1-\hat{\boldsymbol{y}})\right). \tag{14}$$

The total multi-task objective $\mathcal{L}$ of our framework is:

$$\mathcal{L} = \mathcal{L}_{\text{RQ-VAE}} + \mathcal{L}_c. \tag{15}$$

IV. EXPERIMENTS

In this section, we assess the SHIELD framework by addressing four key research questions: **RQ1:** How does SHIELD compare in performance to other state-of-the-art recommendation models? **RQ2:** What is the contribution of each component in the SHIELD model to performance enhancement? **RQ3:** How do the hyperparameters affect the prediction performance, and how to choose optimal values? **RQ4:** Does SHIELD construct a semantic hierarchy structure that captures relationships among diseases?

A. Experimental Settings

1) Dataset Descriptions: We evaluate our model on two widely used real-world EHR datasets: eICU [24] and MIMIC-III [25]. Following prior work [10], we retain patients with at least two clinical visits to ensure meaningful temporal modeling, and aim to predict the CCS codes of their next visit. The detailed dataset statistics are summarized in Table I.

2) Evaluation Metrics: To comprehensively assess model performance, we adopt both visit-level Precision@k (P@k) and code-level Accuracy@k (Acc@k), providing coarse- and fine-grained evaluation, respectively. These metrics are widely used in prior studies [10], [13], [26].

3) Baselines: To validate the effectiveness of SHIELD, we compare it against representative state-of-the-art methods from three categories: (1) **Semantic-aware methods:** BERT [27], BERT*, BioBERT [5], BioBERT*, and VecoCare [28], where * denotes models that incorporate patients' historical visits during training.(2) **Hierarchy-aware methods:** KAME [10],

TABLE II

EXPERIMENTAL RESULTS FOR DIAGNOSIS PREDICTION (%) ON THE EICU AND MIMIC-III DATASETS WITH 5% TRAINING DATA. THE BEST PERFORMANCES ARE HIGHLIGHTED IN **BOLDFACE** AND THE SECOND RUNNERS ARE UNDERLINED. *IMPROV.* DENOTES THE RELATIVE IMPROVEMENTS OF OUR PROPOSED SHIELD OVER THE SECOND RUNNERS.

Dataset	eICU				MIMIC-III			
Metric	Visit-Level		Code-Level		Visit-Level		Code-Level	
	P@10	P@20	Acc@10	Acc@20	P@10	P@20	Acc@10	Acc@20
BERT	$33.72_{\pm 0.12}$	$45.64_{\pm 0.13}$	$27.22_{\pm 0.12}$	$41.07_{\pm 0.11}$	$21.77_{\pm 0.22}$	$32.19_{\pm 0.23}$	$16.14_{\pm 0.22}$	$31.88_{\pm 0.19}$
BERT*	$36.34_{\pm 0.19}$	$50.29_{\pm 0.24}$	$29.87_{\pm 0.25}$	$44.75_{\pm 0.25}$	$23.81_{\pm 0.19}$	$32.26_{\pm 0.24}$	$17.84_{\pm 0.25}$	$32.15_{\pm 0.25}$
BioBERT	$43.76_{\pm 0.09}$	$58.98_{\pm 0.12}$	$35.02_{\pm 0.10}$	$51.96_{\pm 0.12}$	$34.17_{\pm 0.21}$	$43.31_{\pm 0.26}$	$25.35_{\pm 0.19}$	$42.06_{\pm 0.20}$
BioBERT*	$43.88_{\pm 0.11}$	$59.07_{\pm 0.14}$	$35.12_{\pm 0.13}$	$52.09_{\pm 0.13}$	$34.42_{\pm 0.25}$	$43.44_{\pm 0.29}$	$25.49_{\pm 0.18}$	$42.27_{\pm 0.21}$
VecoCare	$45.63_{\pm 0.10}$	$61.08_{\pm 0.12}$	$36.20_{\pm 0.09}$	$53.84_{\pm 0.12}$	$35.27_{\pm 0.20}$	$43.54_{\pm 0.19}$	$25.98_{\pm 0.21}$	$42.30_{\pm 0.23}$
KAME	$45.63_{\pm 0.08}$	$60.94_{\pm 0.11}$	$36.01_{\pm 0.09}$	$53.66_{\pm 0.08}$	$35.01_{\pm 0.17}$	$43.17_{\pm 0.20}$	$25.95_{\pm 0.21}$	$42.16_{\pm 0.18}$
CGL	$45.97_{\pm 0.13}$	$61.38_{\pm 0.09}$	$36.13_{\pm 0.11}$	$53.67_{\pm 0.12}$	$35.54_{\pm 0.22}$	$43.78_{\pm 0.25}$	$26.15_{\pm 0.19}$	$42.49_{\pm 0.20}$
HiTANet	$46.32_{\pm 0.09}$	$61.77_{\pm 0.12}$	$36.26_{\pm 0.12}$	$53.79_{\pm 0.10}$	$36.23_{\pm 0.15}$	$44.32_{\pm 0.19}$	$26.56_{\pm 0.17}$	$43.02_{\pm 0.16}$
DoctorAI	$45.83_{\pm 0.05}$	$61.12_{\pm 0.07}$	$36.12_{\pm 0.06}$	$53.88_{\pm 0.05}$	$35.72_{\pm 0.16}$	$43.70_{\pm 0.20}$	$26.31_{\pm 0.19}$	$42.64_{\pm 0.21}$
RETAIN	$45.54_{\pm 0.04}$	$60.89_{\pm 0.05}$	$35.83_{\pm 0.03}$	$53.51_{\pm 0.06}$	$34.73_{\pm 0.18}$	$43.10_{\pm 0.18}$	$25.73_{\pm 0.17}$	$42.00_{\pm 0.19}$
StageNet	$46.05_{\pm 0.07}$	$61.34_{\pm 0.08}$	$36.11_{\pm 0.06}$	$53.36_{\pm 0.07}$	$36.02_{\pm 0.19}$	$43.91_{\pm 0.21}$	$26.47_{\pm 0.16}$	$42.74_{\pm 0.13}$
TRANS	$46.34_{\pm 0.05}$	$62.00_{\pm 0.05}$	$36.35_{\pm 0.06}$	$54.01_{\pm 0.04}$	$36.64_{\pm 0.22}$	$44.85_{\pm 0.23}$	$26.96_{\pm 0.19}$	$43.35_{\pm 0.22}$
SepsisCalc	$46.44_{\pm 0.05}$	$62.12_{\pm 0.04}$	$36.48_{\pm 0.06}$	$54.32_{\pm 0.03}$	$36.82_{\pm 0.21}$	$45.01_{\pm 0.22}$	$27.08_{\pm 0.18}$	$43.44_{\pm 0.20}$
SHIELD	$\mathbf{54.72_{\pm 0.04}}$	$\mathbf{68.15_{\pm 0.08}}$	$\mathbf{48.85_{\pm 0.06}}$	$\mathbf{65.57_{\pm 0.08}}$	$\mathbf{40.17_{\pm 0.23}}$	$\mathbf{46.55_{\pm 0.25}}$	$\mathbf{29.27_{\pm 0.21}}$	$\mathbf{45.23_{\pm 0.19}}$
Improv.	17.83%	9.71%	33.91%	20.71%	9.10%	3.42%	8.09%	4.12%

CGL [17], and HiTANet [29], which exploit medical ontologies.(3) **Temporal-aware methods:** DoctorAI [1], RETAIN [2], StageNet [30], TRANS [13], and SepsisCale [31]. Note that KAME, CGL, TRANS, and SepsisCale leverage graph neural networks for medical relationships.

4) Implementation Details: We split both datasets into training, validation, and test sets with a ratio of 7:1:2 at the patient level, consistent with [6], [17], [32]. All baseline models are optimized using the Adam optimizer, with hyperparameters tuned according to their original implementations. The embedding dimension for both diseases and patients is set to $d = 64$. Our model is trained with the AdamW optimizer with L2 weight decay and a learning rate of 10^{-3}. The number of RQ-VAE layers is set to $L = 4$, which aligns with the hierarchical structure of the ICD codes. The number of centers K in the K-means clustering algorithm is set to 128. For the quantization loss, we set $\beta = 0.25$ following [11], [23]. The parameters $\mu = 1$, $\lambda = 0.15$ (for the eICU), and $\lambda = 0.25$ (for the MIMIC-III). The number of layers for the hypergraph convolution is set to $\hat{L} = 2$. All experiments are conducted using a NVIDIA GTX 3090 Ti GPU. The full code for this work is available[1].

B. Overall Performance (RQ1)

Table II summarizes the performance on eICU and MIMIC-III using 5% training data. SHIELD consistently outperforms all baselines, achieving up to 17.83% and 9.10% gains in P@10 over the second-best method. Key observations include:

Superiority over Semantic-aware Methods: While BioBERT and VecoCare focus on disease semantics, their flat representations fail to capture hierarchies. By leveraging RQ-VAE-based quantization, SHIELD surpasses VecoCare by 19.92% (eICU) and 13.89% (MIMIC-III), demonstrating the power of our interpretable semantic hierarchy.

[1]https://anonymous.4open.science/r/SHIELD

TABLE III

ABLATION STUDY RESULTS (%) ON EICU AND MIMIC-III DATASETS.

Metric	P@10	P@20	Acc@10	Acc@20
eICU				
SHIELD w/o. HSQ	53.82	67.48	48.14	64.97
SHIELD w/o. DHC	51.27	65.87	45.43	63.09
SHIELD w/o. TAP	53.55	67.24	47.11	64.14
SHIELD (our method)	**54.72**	**68.15**	**48.85**	**65.57**
MIMIC-III				
SHIELD w/o. HSQ	39.60	46.24	29.03	44.92
SHIELD w/o. DHC	37.53	45.33	27.53	43.73
SHIELD w/o. TAP	39.41	46.42	28.91	44.94
SHIELD (our method)	**40.17**	**46.55**	**29.27**	**45.23**

Flexibility over Hierarchy-aware Methods: Unlike KAME and HiTANet, which rely on rigid, predefined ICD ontologies, SHIELD learns dynamic hierarchical representations from data. This adaptability leads to significant improvements over HiTANet (18.13% on eICU and 10.87% on MIMIC-III).

Advantage of Hypergraph Modeling: Compared to graph-based methods (e.g., CGL, SepsisCalc) that model only pairwise links, SHIELD's dual-view hypergraph captures higher-order interactions. Consequently, it outperforms SepsisCalc by notable margins on both datasets.

The pronounced improvements in Accuracy further highlight SHIELD's efficacy in fine-grained prediction. The multi-level residual quantization effectively captures both coarse- and fine-grained semantics, enabling precise clinical inference by unifying semantic hierarchy, higher-order relations, and temporal dynamics.

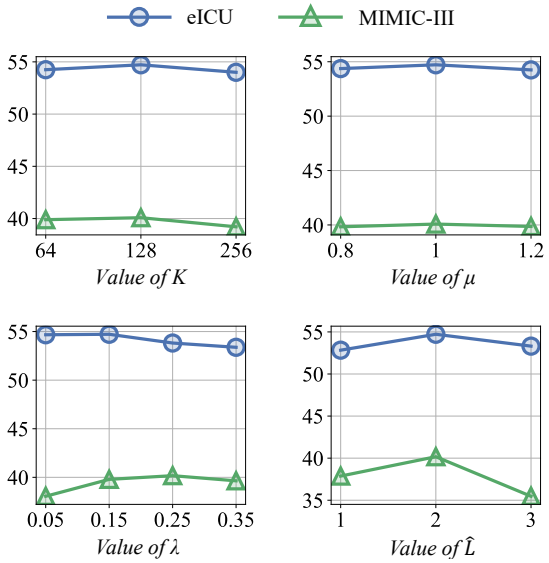


Fig. 3. Performance (Precision@10) of SHIELD with Varying Hyperparameters on eICU and MIMIC-III.

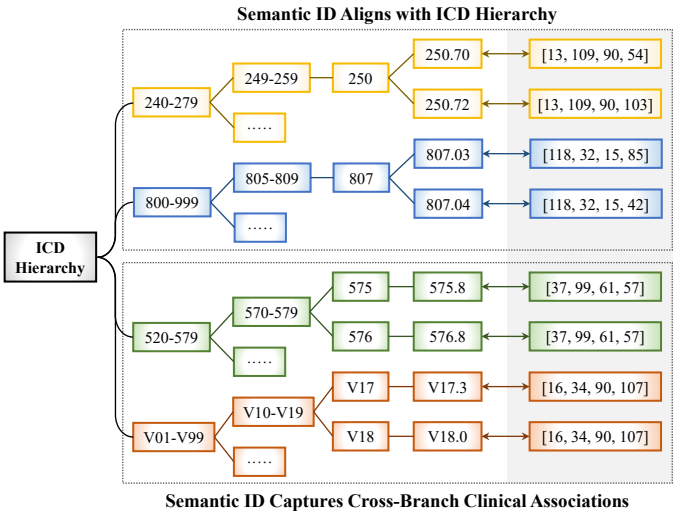


Fig. 4. An Analysis of diseases encoding between ICD hierarchy and semantic hierarchy.

C. Ablation Study (RQ2)

To quantify the contribution of each component in SHIELD, we evaluate three variants: (1) **w/o HSQ**, removing hierarchical quantization; (2) **w/o DHC**, excluding dual-view hypergraph convolutions; and (3) **w/o TAP**, replacing temporal attention with simple averaging.

As shown in Table III, all modules contribute to the final performance. Specifically: (1) Effect of HSQ. Removing HSQ leads to a consistent decline (e.g., P@10 drops from 54.72% to 53.82% on eICU), proving that Semantic IDs effectively capture fine-grained taxonomic structures. (2) **Effect of DHC.** SHIELD w/o DHC shows the most significant drop (Acc@10 decreases by 6.41% on eICU and 7.03% on MIMIC-III), underlining its critical role in modeling high-order comorbidity and patient-disease correlations. (3) **Effect of TAP.** The performance dip in Acc@20 without TAP suggests that the location-based attention is vital for identifying clinically significant visits in long-term sequences.

D. Hyperparameters Study (RQ3)

Our framework involves four key hyperparameters: K, μ, λ, and $\hat{L}$. Figure 3 shows the tuning results.

Codebook Size K. K determines the number of centroids in the codebook. The best performance is achieved at $K = 128$. Smaller values limit semantic representation capacity, while larger values introduce redundancy and quantization noise.

Loss Weights μ and λ. μ and λ control the quantization loss and diversity loss, respectively. The model performs best around $\mu = 1.0$ and $\lambda = 0.15$. A small μ weakens representation alignment, whereas an excessively large μ over-constrains quantization. Appropriate λ improves codebook utilization and alleviates code collapse.

Hypergraph Depth $\hat{L}$. $\hat{L}$ controls the propagation depth of hypergraph convolution. The optimal performance is obtained at $\hat{L} = 2$. Shallow layers fail to capture higher-order interactions, while deeper layers suffer from over-smoothing.

E. Case Study (RQ4)

We analyze the MIMIC-III dataset to examine the alignment between the learned semantic hierarchy and the ICD taxonomy.

Alignment with ICD Taxonomy The Semantic IDs exhibit strong structural consistency with the ICD tree. For example, codes 250.70 and 250.72 share identical Semantic ID prefixes and only diverge at the final level, confirming that RQ-VAE captures coarse-to-fine disease semantics that mirror established medical taxonomies.

Discovery of Cross-Branch Associations Beyond predefined structures, RQ-VAE uncovers latent clinical correlations. Codes 575.8 (*gallbladder disorders*) and 576.8 (*biliary tract disorders*) are assigned identical IDs despite their separation in the ICD system, reflecting their functional clinical proximity. Similarly, V18.0 (*diabetes history*) and V17.3 (*ischemic heart disease history*) share the same ID, effectively capturing real-world co-occurrence patterns.

These results demonstrate that RQ-VAE not only preserves the essential medical hierarchy but also transcends static coding boundaries to reveal data-driven clinical associations.

V. Conclusion

We propose SHIELD, a novel framework for diagnosis prediction that integrates hierarchical semantic learning with hypergraph-based relational modeling. SHIELD constructs an interpretable, data-driven disease hierarchy to preserve both coarse- and fine-grained semantics. By employing dual-view hypergraph convolutions, the model effectively captures higher-order interactions between patients and diseases, while a temporal attention mechanism further leverages longitudinal EHR patterns. Extensive experiments demonstrate that SHIELD consistently outperforms state-of-the-art baselines, showcasing its practical utility in clinical environments.

References

[1] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor ai: Predicting clinical events via recurrent neural networks," in *Machine learning for healthcare conference*. PMLR, 2016, pp. 301–318.

[2] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart, "Retain: An interpretable predictive model for healthcare using reverse time attention mechanism," *Advances in neural information processing systems*, vol. 29, 2016.

[3] P. Nguyen, T. Tran, N. Wickramasinghe, and S. Venkatesh, "Deeper: a convolutional net for medical records," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 22–30, 2016.

[4] 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 *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2017, pp. 1903–1911.

[5] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "Biobert: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020.

[6] H. Lv, Z. Chen, Y. Yang, G. Ma, T. Yanchao, and C. Yang, "Boxcare: A box embedding model for disease representation and diagnosis prediction in healthcare data," in *Companion Proceedings of the ACM Web Conference 2024*, ser. WWW '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 1130–1133. [Online]. Available: https://doi.org/10.1145/3589335.3651448

[7] C. Mancusi, R. Izzo, G. di Gioia, M. A. Losi, E. Barbato, and C. Morisco, "Insulin resistance the hinge between hypertension and type 2 diabetes," *High blood pressure & cardiovascular prevention*, vol. 27, no. 6, pp. 515–526, 2020.

[8] Y. C. Lee, S.-H. Jung, A. Kumar, I. Shim, M. Song, M. S. Kim, K. Kim, W. Myung, W.-Y. Park, and H.-H. Won, "Icd2vec: Mathematical representation of diseases," *Journal of Biomedical Informatics*, vol. 141, p. 104361, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046423000825

[9] J. Yan, H. Gao, Z. Kai, W. Liu, D. Chen, J. Wu, and J. Chen, "Text2tree: Aligning text representation to the label tree hierarchy for imbalanced medical classification," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023, pp. 7705–7720.

[10] F. Ma, Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao, "Kame: Knowledge-based attention model for diagnosis prediction in healthcare," in *Proceedings of the 27th ACM international conference on information and knowledge management*, 2018, pp. 743–752.

[11] S. Rajput, N. Mehta, A. Singh, R. Hulikal Keshavan, T. Vu, L. Heldt, L. Hong, Y. Tay, V. Tran, J. Samost *et al.*, "Recommender systems with generative retrieval," *Advances in Neural Information Processing Systems*, vol. 36, pp. 10 299–10 315, 2023.

[12] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "Predicting healthcare trajectories from medical records: A deep learning approach," *Journal of biomedical informatics*, vol. 69, pp. 218–229, 2017.

[13] J. Chen, C. Yin, Y. Wang, and P. Zhang, "Predictive modeling with temporal graphical representation on electronic health records," in *IJCAI: proceedings of the conference*, vol. 2024, 2024, p. 5763.

[14] H. Lv, Z. Guo, Z. Wu, Y. Tan, G. Ma, Z. Lin, X. Chen, H. Cheng, and C. Yang, "Medalign: Enhancing combinatorial medication recommendation with multi-modality alignment," in *Proceedings of the 33rd ACM International Conference on Multimedia*, ser. MM '25. New York, NY, USA: Association for Computing Machinery, 2025, p. 6084–6092. [Online]. Available: https://doi.org/10.1145/3746027.3755265

[15] T. Wu, Y. Wang, Y. Wang, E. Zhao, and Y. Yuan, "Leveraging graph-based hierarchical medical entity embedding for healthcare applications," *Scientific reports*, vol. 11, no. 1, p. 5858, 2021.

[16] C. A. Hidalgo, N. Blumm, A.-L. Barabási, and N. A. Christakis, "A dynamic network approach for the study of human phenotypes," *PLoS computational biology*, vol. 5, no. 4, p. e1000353, 2009.

[17] C. Lu, C. K. Reddy, P. Chakraborty, S. Kleinberg, and Y. Ning, "Collaborative graph learning with auxiliary text for temporal event prediction in healthcare," *arXiv preprint arXiv:2105.07542*, 2021.

[18] Y. Chen, T. Guan, and C. Wang, "Approximate nearest neighbor search by residual vector quantization," *Sensors*, vol. 10, no. 12, pp. 11 259–11 273, 2010.

[19] A. Van Den Oord, O. Vinyals *et al.*, "Neural discrete representation learning," *Advances in neural information processing systems*, vol. 30, 2017.

[20] D. Lee, C. Kim, S. Kim, M. Cho, and W.-S. Han, "Autoregressive image generation using residual quantization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 523–11 532.

[21] B. Zheng, Y. Hou, H. Lu, Y. Chen, W. X. Zhao, M. Chen, and J.-R. Wen, "Adapting large language models by integrating collaborative semantics for recommendation," in *2024 IEEE 40th International Conference on Data Engineering (ICDE)*. IEEE, 2024, pp. 1435–1448.

[22] N. Zeghidour, A. Luebs, A. Omran, J. Skoglund, and M. Tagliasacchi, "Soundstream: An end-to-end neural audio codec," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 495–507, 2021.

[23] D. Wang, Y. Huang, S. Gao, Y. Wang, C. Huang, and S. Shang, "Generative next poi recommendation with semantic id," *arXiv preprint arXiv:2506.01375*, 2025.

[24] T. J. Pollard, A. E. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, and O. Badawi, "The eicu collaborative research database, a freely available multi-center database for critical care research," *Scientific data*, vol. 5, no. 1, p. 180178, 2018.

[25] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[26] X. Zhang, B. Qian, S. Cao, Y. Li, H. Chen, Y. Zheng, and I. Davidson, "Inprem: An interpretable and trustworthy predictive model for healthcare," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 450–460.

[27] J. D. M.-W. C. Kenton, L. K. Toutanova *et al.*, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of naacL-HLT*, vol. 1, no. 2. Minneapolis, Minnesota, 2019.

[28] Y. Xu, K. Yang, C. Zhang, P. Zou, Z. Wang, H. Ding, J. Zhao, Y. Wang, and B. Xie, "Vecocare: Visit sequences-clinical notes joint learning for diagnosis prediction in healthcare data." in *IJCAI*, vol. 23, 2023, pp. 4921–4929.

[29] J. Luo, M. Ye, C. Xiao, and F. Ma, "Hitanet: Hierarchical time-aware attention networks for risk prediction on electronic health records," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 647–656.

[30] J. Gao, C. Xiao, Y. Wang, W. Tang, L. M. Glass, and J. Sun, "Stagenet: Stage-aware neural networks for health risk prediction," in *Proceedings of the web conference 2020*, 2020, pp. 530–540.

[31] C. Yin, S. Fu, B. Yao, T.-H. Pham, W. Cao, D. Wang, J. Caterino, and P. Zhang, "Sepsiscalc: Integrating clinical calculators into early sepsis prediction via dynamic temporal graph construction," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, 2025, p. 2779–2790.

[32] Y. Tan, H. Lv, Y. Zhan, G. Ma, B. Xiong, and C. Yang, "Boxlm: unifying structures and semantics of medical concepts for diagnosis prediction in healthcare," in *Proceedings of the 42nd International Conference on Machine Learning*, ser. ICML'25. JMLR.org, 2025.