Knowledge-Enhanced Complementary Information Fusion with Temporal Heterogeneous Graph Learning for Disease Prediction

Zongbao Yang[1], Shihuan He[1], Zhichen Chen[1], Hao Zhang(✉)[1], and Ruxin Wang(✉)[1,2]

[1] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[2] University of Chinese Academy of Sciences, Beijing, China
{h.zhang10, rx.wang}@siat.ac.cn

Abstract. Disease prediction based on multimodal data is a critical yet challenging task in healthcare, especially in intensive care units (ICUs) where patients present complex clinical trajectories with multiple admissions and comorbidities. Current multimodal learning approaches lack effective modeling of cross-modal complementary information, which leads to suboptimal feature interactions. Besides, traditional methods that incorporate external knowledge graphs (KGs) often introduce noise and computational complexity, due to the use of all one-hop neighbors within the KGs. To address these challenges, we propose Knowledge-Enhanced Complementary Information Fusion with temporal heterogeneous graph learning (KCIF) for patient modeling. KCIF introduces a temporal heterogeneous admission graph (THAG) that integrates KGs to capture semantic and temporal dependencies across admissions. It further employs a complementary information fusion mechanism to leverage mutual enhancement between lab tests and medical events. Extensive experiments on the MIMIC-III/IV benchmarks demonstrate that KCIF consistently outperforms baselines, achieving improvements of over 2.5%–6.0% in w-F_1 score and 1.7%–4.5% in R@20 across multiple ICU disease prediction. The code is available at `https://github.com/Boaz-SCUT/KCIF`.

Keywords: Multimodal Disease prediction · Complementary Information · Temporal Admission Graph.

1 Introduction

The intensive care unit (ICU) is one of the most critical and complex environments in the modern healthcare system, where patients require continuous monitoring and timely intervention [14]. In this high-acuity environment, the ability to accurately predict a patient's clinical trajectory is extremely important as it can enable timely medical intervention, optimize the allocation of medical resources, and support effective clinical decision-making [5].

⋆ Corresponding authors: Ruxin Wang and Hao Zhang.

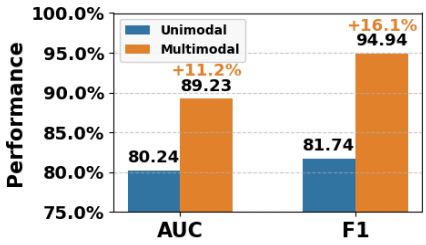

Fig. 1. Performance comparison between unimodal and multimodal approaches. Blue bars represent the unimodal (KGxDP [24]) method, while orange bars represent our multimodal method.

However, developing robust predictive models for ICU patients poses significant challenges, primarily due to the intrinsic complexity of clinical data. ICU data covers a variety of modalities, including laboratory tests and other medical events [14,21], from discrete diagnosis codes to time series information [19,24,26]. Crucially, a significant proportion of ICU patients (>68%) actually suffer from multiple comorbidities [16], and single-mode data are insufficient to fully reflect the patient's condition. This requires the integration of multimodal medical data to achieve more accurate predictions. Experimental results have consistently demonstrated the superior performance of incorporating multimodal knowledge as shown in Fig.1. Therefore, while significant progress has been made in unimodal medical data analysis [3,12,11,24], there is an urgent need to develop multimodal models that can effectively capture disease interactions and their temporal evolution patterns.

Existing multimodal methods in this field can be roughly divided into two categories: (1) alignment-based methods [21,22,18,9], which use contrastive learning and self-supervised pre-training techniques to align different modalities in a shared latent space; (2) Attention-based methods [15,25,26], which use feature weighted concatenation or cross-attention to fuse multimodal representations. However, these approaches cannot effectively model the inherent complementary relationships between different modalities (e.g., how lab tests provide physiological evidence for diagnosis and vice versa), resulting in suboptimal feature fusion and limited predictive performance. In addition, to enhance the modeling ability of medical events, researchers have introduced various external medical knowledge, such as medical ontologies[2,13], knowledge graphs[20,24,1] and even large language models (LLMs)[6,23,19]. However, these methods often incorporate all one-hop neighbors within KGs, introducing significant noise and computational complexity. Additionally, they focus mainly on static semantic connections while overlooking the important temporal dependencies between patient admissions that reflect the progression of patients' health status [24].

To address these challenges, we propose Knowledge-Enhanced Complementary Information Fusion (KCIF) with temporal heterogeneous graph learning for

disease prediction. Specifically, KCIF leverages a time-enhanced Transformer to model discrete medical events, while constructing the THAG to capture both semantic and temporal dependencies among medical events. For lab test sequence modeling, we employ a hierarchical Transformer to effectively capture multi-scale temporal patterns at both intra-visit and inter-visit levels. In addition, KCIF extracts complementary information between lab tests and medical events to improve patient modeling ability.

The contributions of this work are: (1) We construct THAG through external KGs, which can effectively capture the semantic and temporal dependencies between admission histories. (2) We develop specialized constraints to extract complementary information from lab test and medical events, enabling more comprehensive patient representation learning. (3) We conducted experiments on the MIMIC-III/IV datasets and the results demonstrated the superiority of KCIF and the effectiveness of each module.

2 Methodology

2.1 Problem Formulation

Given a patient p with static demographic $x^p_{\text{static}} = [x^p_a, x^p_g, x^p_r]$, where x^p_a, x^p_g, and x^p_r represent the patient's age, gender, and race respectively. For notational simplicity, we will omit the superscript p in the subsequent formulations. The longitudinal medical history is formalized as a temporally-ordered sequence of hospital admissions $A = \{A_{t_1}, A_{t_2}, ..., A_{t_T}\}$, with corresponding timestamps $t = \{t_1, t_2, \cdots, t_T\}$. Each admission $A_{t_i} = [D_{t_i}, P_{t_i}, M_{t_i}; L_{t_i}]$ contains 1) medical events: diagnostic codes $D_{t_i} = \{d_c^{(t_i)}\}_{c=1}^{|D_{t_i}|}$, procedural codes $P_{t_i} = \{p_c^{(t_i)}\}_{c=1}^{|P_{t_i}|}$ and medicine codes $M_{t_i} = \{m_c^{(t_i)}\}_{c=1}^{|M_{t_i}|}$, where $|\cdot|$ represents the number of elements in the set. 2) lab tests L_{t_i}. The medical events are coded by the standardized ICD-9 coding system[1]. Therefore, the complete medical events of this patient can be denoted as $ME = \{[D; P; M]\}$, where $D, P, M = \{D_{t_i}\}_{t_i=1}^{t_T}, \{P_{t_i}\}_{t_i=1}^{t_T}, \{M_{t_i}\}_{t_i=1}^{t_T}$. The lab tests are represented as a temporal sequence $L = \{L_{t_1}, L_{t_2}, \ldots, L_{t_T}\}$, where $L_{t_i} = [l^1_{t_i}, l^2_{t_i}, \cdots, l^{w_{t_i}}_{t_i}]$, where w_{t_i} denotes the number of discrete time windows in t_i-th admission, $l^j_{t_i}$ stands for the lab test results in j-th time window.

To better capture the complex relationships between clinical events, we represent the admission history as a temporal heterogeneous admission graph (THAG) $G = (V, E)$. The node set V consists of admission event $V_a = \{a_{t_1}, a_{t_2}, \cdots, a_{t_T}\}$, diagnosis $\{v_d | v_d \in D\}$, procedure $\{v_p | v_p \in P\}$, and medication $\{v_m | v_m \in M\}$, where each a_{t_i} represents the admission event itself (rather than the medical records within that admission). The edge E contains three types of relationships: 1) E_{time} stands for the temporal sequence between adjacency admissions; 2) E_{has} stands for the association between admissions and their corresponding

[1] https://archive.cdc.gov/www_cdc_gov/nchs/icd/icd9cm.html

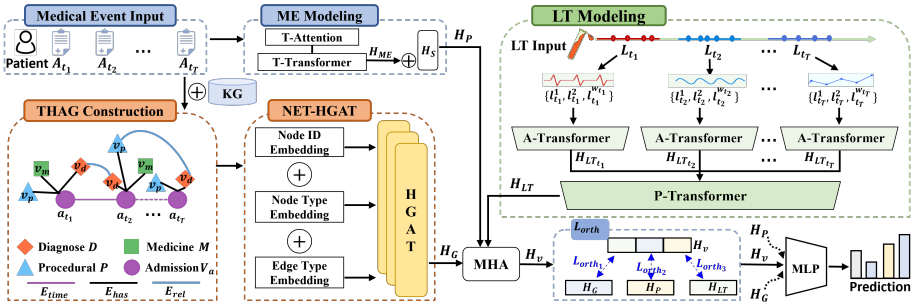

Fig. 2. The architecture of KCIF. (MHA: Multi-Head Attention, MLP: Multi-Layer Perceptron, HGAT: Heterogeneous Graph Attention Network)

medical events; 3) E_{rel} stands for the semantic relationships between medical events derived from external KGs. These edges are defined in Eq.(1):

$$\begin{aligned} E_{time} &= \{(a_{t_i}, \Delta_{t_i,t_{i+1}}, a_{t_{i+1}}) \mid a_{t_i}, a_{t_{i+1}} \in V_a\}, \\ E_{has} &= \{(a_{t_i}, r_{v_k,a_{t_i}}, v_k) \mid a_{t_i} \in V_a, v_k \in ME\}, \\ E_{rel} &= \{(v_{head}, r_{h,t}, v_{tail}) \mid v_{head}, v_{tail} \in ME, r_{h,t} \in KG\}, \end{aligned} \tag{1}$$

where $\Delta_{t_i,t_{i+1}}$ denotes the time interval between adjacency admissions, $r_{v_k,a_{t_i}}$ indicates t_i-th admission has medical event v_k. $r_{h,t}$ represents the semantic relationship between medical events as defined in the KGs.

2.2 Medical Event Modeling

The architecture of the proposed model is depicted in Fig.2. We model discrete medical events through time-aware attention (T-Attention), enabling the learning of admission-level representations. Specifically, the representation of t_i-th admission H_{t_i} is computed as Eq.(2):

$$H_{t_i} = \text{T-Attention}(ME, t_i) = \text{Self-Attention}(F_{me}(ME) \cdot F_T(\Delta t_{i,T})), \tag{2}$$

where $F_{me}(\cdot)$ represents the learnable medical events embedding mapping, and $F_T(\cdot)$ denotes a non-linear function encoding the time difference between current and last admission, with $\Delta t_{i,T} = t_T - t_i$. To capture long-term dependencies in the admission sequence, we develop a time-enhanced Transformer (T-Transformer) as Eq.(3):

$$H_{ME} = \text{T-Transformer}(H_t, t) = \text{Transformer}(H_t, F'_T(\Delta t_{i,i+1}) + \text{Pos}), \tag{3}$$

where $H_t = \{H_{t_1}, H_{t_2}, \cdots, H_{t_T}\}$ is the admission-level representations from Eq.(2). $F'_T(\cdot)$ is a non-linear function that encodes the temporal interval between adjacent admissions, $\Delta t_{i,i+1} = t_{i+1} - t_i$. Pos denotes the positional embedding of each admission. Finally, we use multihot encoding to process the patient's static demographic to obtain $H_S = \text{multi-hot}(x_{static})$, and then concatenate it with H_{ME} to obtain the medical event level representation $H_P = [H_{ME}; H_S]$.

2.3 Temporal Heterogeneous Admission Graph Modeling

We introduce the Node-Edge-Type aware heterogeneous graph attention network (NET-HGAT) to model THAG $G = (V, E)$. Node embedding of $v_k \in V$ is initialized by concatenating the entity embedding and type embedding, that is $h_{v_k}^{(0)} = [F_e(v_k); F_{type}(\Phi(v_k))]$, where $F_e(\cdot)$ and $F_{type}(\cdot)$ map nodes and their types to corresponding embedding spaces. $\Phi(\cdot)$ maps nodes to their types. To capture dynamic entity interactions, we design the multi-layer heterogeneous GAT (HGAT), where the embedding of v_k at layer $l+1$ is updated as Eq.(4):

$$h_{v_k}^{(l+1)} = h_{v_k}^{(l)} + \sigma\left(\sum_{u \in \mathcal{N}(v_k)} \alpha_{u,v_k} \mathbf{W} h_u^{(l)}\right), \tag{4}$$

where $\sigma(\cdot)$ is a non-linear activation function, α_{u,v_k} is calculated by integrating the node embeddings and edge semantics of nodes v_k and u, as shown in Eq.(5):

$$\alpha_{u,v_k} = \frac{\exp(F_v(\mathbf{W_1} h_{v_k}^{(l)}, \mathbf{W_2} h_u^{(l)}, F_{edge}(\Psi(u, v_k))))}{\sum_{u \in \mathcal{N}(v_k)} \exp(F_v(\mathbf{W_1} h_{v_k}^{(l)}, \mathbf{W_2} h_u^{(l)}, F_{edge}(\Psi(u, v_k))))}, \tag{5}$$

where $F_v(\cdot)$ is a learnable function that integrates node and edge information to calculate attention weights, $\mathbf{W_1}$ and $\mathbf{W_2}$ are the transformation matrix, $\Psi(\cdot,\cdot)$ maps edges to their types, and $F_{edge}(\cdot)$ embeds edge types. After applying L layers, we derive the patient graph level representation $H_G = \frac{1}{|V|} \sum_{v_k \in V} h_{v_k}^{(L)}$.

Moreover, to ensure semantic alignment between medical event level embeddings H_P and graph level embeddings H_G, we design a contrastive learning objective as Eq.(6):

$$\mathcal{L}_C = -\sum_m \left[\log \frac{\exp(s_{mm})}{\sum_n \exp(s_{mn})} + \lambda \log \frac{\exp(s_{mm})}{\sum_{n \neq n} \exp(s_{mn})}\right], \tag{6}$$

where $s_{mn} = \text{sim}(H_P^{(m)}, H_G^{(n)})$ measures the cosine similarity between H_P, H_G of patient m and n. λ controls the weight of negative sample pairs.

2.4 Lab Tests (LT) Modeling

Lab tests exhibit complex temporal dependencies at both intra-admission and inter-admission levels, which can provide quantitative indicators that complement discrete medical events by capturing clinical changes and early disease signals not yet recorded in medical codes. To effectively capture these multi-scale patterns, we propose a hierarchical Transformer (H-Transformer) for lab test modeling. Given the lab test sequence for t_i-th admission $L_{t_i} = [l_{t_i}^1, l_{t_i}^2, \cdots, l_{t_i}^{w_{t_i}}]$, we first employ a admission-level Transformer (A-Transformer) to model the temporal relationships between lab tests within the same admission, as Eq.(7):

$$H_{LT_{t_i}} = \text{A-Transformer}(L_{t_i}) = \text{Transformer}(l_{t_i}^1, l_{t_i}^2, \cdots, l_{t_i}^{w_{t_i}}). \tag{7}$$

We then aggregate the encoded admission-level representations of all admitted patients H_{LT_t} via a patient-level transformer (P-Transformer), as Eq.(8):

$$H_{LT} = \text{P-Transformer}(H_{LT_t}) = \text{Transformer}(H_{LT_{t_1}}, H_{LT_{t_2}}, \cdots, H_{LT_{t_T}}), \quad (8)$$

where H_{LT} is the final representation of lab tests. Transformer$(\cdot)$ represents the vanilla encoder layer of Transformer[17] with L_{t_i} and H_{LT_t} as input.

2.5 Cross-Modal Complementary Information Learning

The basic principle of complementary information extraction is exploiting the deep interactions between modalities. Given H_P, H_G, and H_{LT}, the inter-modal interactions are captured using multi-head attention (MHA) followed by mean pooling, as Eq.(9):

$$H_v = \text{Mean}(\text{Multi-HeadAttention}(H_P, H_G, H_{LT})). \quad (9)$$

To ensure that the model captures truly valuable complementary information. We impose a soft orthogonality constraint, as defined in Eq.(10):

$$\mathcal{L}_{orth} = \underbrace{\cos(H_v, H_P)}_{\mathcal{L}_{orth_1}} + \underbrace{\cos(H_v, H_G)}_{\mathcal{L}_{orth_2}} + \underbrace{\cos(H_v, H_{LT})}_{\mathcal{L}_{orth_3}}, \quad (10)$$

where $\cos(\cdot, \cdot)$ is the cosine similarity between H_v and all modal representations.

Moreover, the loss $\mathcal{L}_{pred}$ based on the fused representation $[H_P; H_G; H_v]$ is minimized to ensure the best performance of the model. $\mathcal{L}_{pred}$ represents the binary cross entropy or multi-class cross-entropy loss, depending on the specific clinical prediction task. The final total loss function is given by Eq.(11):

$$\mathcal{L}_{total} = \mathcal{L}_{pred} + \lambda_1 \mathcal{L}_{orth} + \lambda_2 \mathcal{L}_C. \quad (11)$$

3 Experimental Setup

3.1 Datasets and Baselines

Following prior studies [3,11,12,24], we evaluate KCIF on two widely used benchmarks: MIMIC-III [8] and MIMIC-IV [7], ensuring fair and consistent comparisons. Furthermore, we strictly adhere to the patient cohort selection criteria established in [24,10]. To accommodate the proposed approach, we further filter the data by lab tests. For lab test extraction, we followed [4] and selected 17 clinically significant lab tests. We then segmented them into 12-hour time windows within each admission to effectively capture the temporal patterns. After preprocessing, the MIMIC-III contains 7,073 patients, with 5,617/463/993 patients for train/dev/test, and MIMIC-IV contains 3,556 patients, with 2,876/335/345 patients for train/dev/test. For quantitative evaluation, we adopt the same metrics as in paper[24,10]: w-F_1, $R@k$ ($k = 10, 20$) for multi-disease prediction, and F_1, AUC for binary disease prediction.

Table 1. Performance comparison on MIMIC-III dataset. (%)

Models	Multi-Disease			Cardiovascular			Hypertension		Heart Failure	
	w-F_1	R@10	R@20	w-F_1	R@10	R@20	AUC	F_1	AUC	F_1
DIPOL	19.35	24.98	34.02	48.37	66.69	77.18	74.90	78.26	82.08	70.35
RETAIN	20.69	26.13	35.08	52.10	72.89	81.97	80.24	77.08	83.21	71.32
HiTA	21.15	26.02	35.97	52.74	73.15	82.54	78.20	81.74	82.77	71.93
KGxDP	31.35	30.98	41.29	53.55	70.99	80.54	79.83	81.70	86.57	74.74
MMUGL	26.40	-	41.54	-	-	-	-	-	-	-
GraphCare	14.80	-	29.97	-	-	-	-	-	-	-
KCIF	**32.41**	**34.63**	**46.10**	**54.05**	**74.57**	**83.69**	**89.23**	**94.94**	**87.52**	**75.74**

Table 2. Performance comparison on MIMIC-IV dataset. (%)

Models	Multi-Disease			Cardiovascular			Hypertension		Heart Failure	
	w-F_1	R@10	R@20	w-F_1	R@10	R@20	AUC	F_1	AUC	F_1
DIPOL	24.34	22.66	30.43	55.64	56.90	63.11	76.42	79.81	86.70	76.67
RETAIN	27.45	26.60	35.60	57.09	61.09	66.79	77.13	80.41	90.11	79.37
HiTA	26.41	29.51	39.30	56.58	72.19	80.86	84.84	88.33	85.97	73.29
KGxDP	29.45	30.64	39.81	57.24	61.43	66.83	87.04	89.40	93.00	78.49
KCIF	**31.99**	**31.20**	**41.59**	**57.90**	**72.52**	**81.64**	**88.22**	**90.99**	**94.13**	**84.04**

We evaluate KCIF against several baselines, including sequence-based models such as **RETAIN**[3], **Dipole**[12], and **HiTANet**[11], which employ sequence models and attention mechanisms for admission history modeling but do not incorporate medical knowledge. Additionally, we compare with graph-based models like **KGxDP**[24], **MMUGL**[1], and **GraphCare**[6], which integrate external KGs or large language models (LLMs) to enhance medical event representation. Due to reproducibility constraints, we report the results of GraphCare and MMUGL as presented in paper[1].

The layer number of NET-HGAT is set to 2, the learning rate is set to 1e-4, and the batch size is set to 16. The output dimension of all Transformer is 768. We conducted grid search for loss function weights, and finding optimal performance with $\lambda_1 = 0.01$, $\lambda_2 = 0.05$. The KG used in this paper is SNOMED CT [2], which is an authoritative, widely-adopted open-access medical KG.

3.2 Main Results

Tables 1 and 2 present the evaluation results of KCIF against baselines across three clinical prediction-level tasks: **1) Multi-disease prediction,** which is the most challenging scenario involving classification among thousands of potential diseases (4,880/5,890 for MIMIC III/IV), KCIF demonstrates substantial improvements over all baselines. It achieves w-F_1 of 32.41% on MIMIC-III and 31.99% on MIMIC-IV, which outperforms the strongest baseline KGxDP by 1.06% and 2.54%. The R@20 shows even more significant improvements of 4.81%

[2] https://www.nlm.nih.gov/healthit/snomedct/index.html

Table 3. Ablation study results on MIMIC-III and MIMIC-IV (%).

Ablation Module	MIMIC-III			MIMIC-IV		
	w-F_1	R@10	R@20	w-F_1	R@10	R@20
w/o THAG	27.96	30.19	40.67	30.07	30.75	40.46
w/o $\mathcal{L}_{\mathcal{C}}$	31.47	33.93	44.77	30.77	30.55	39.46
w/o $\mathcal{L}_{orth}$	30.70	33.51	44.35	30.91	29.69	38.50
KCIF	**32.41**	**34.63**	**46.10**	**31.99**	**31.20**	**41.59**

and 1.78% on the respective datasets compared to the best-performing baselines. **2) Cardiovascular disease prediction,** which focuses on 58 cardiovascular-related diseases, KCIF maintains its superiority with w-F_1 of 54.05% on MIMIC-III and 57.90% on MIMIC-IV. The R@20 reaches 83.69% and 81.64% on the respective datasets, with improvements of 1.15% and 0.78% over the best baselines. **3) Binary classification tasks,** which predicts whether a patient has a certain disease. KCIF demonstrates remarkable performance gains. On hypertension prediction, KCIF achieves AUC improvements of 9.0% (MIMIC-III) and 1.18% (MIMIC-IV) over the best baselines, with even more pronounced F_1 score improvements of 13.24% and 1.59%. For heart failure prediction, KCIF reaches AUC values of 87.52% and 94.13% on the respective datasets, outperforming the best baselines by 0.95% and 1.13%. Notably, KCIF also outperforms recent knowledge-enhanced methods such as KGxDP[24], MMUGL[1], and GraphCare[6], indicating that simply incorporating external knowledge is not enough, even if some of this knowledge is provided by LLM (GraphCare). In contrast, KCIF's dual focus on both temporal graph modeling and complementary information fusion enables consistent performance improvements across all experimental settings. The consistent performance improvement in all tasks demonstrates the effectiveness of KCIF, which can effectively model complex disease interactions and capture important semantic relationships between medical events, and extract complementary information from different modalities through THAG.

3.3 Ablation Studies

We conduct comprehensive ablation experiments to evaluate the contribution of each component in the KCIF, and the results are shown in Table 3. The results show that the removal of THAG leads to the most substantial performance degradation across all metrics (w-F_1 drops by 4.45% on MIMIC-III, 1.92% on MIMIC-IV), demonstrating its pivotal role in encoding external medical knowledge and capturing temporal dependencies. Ablating the contrastive loss $\mathcal{L}_C$ results in moderate performance decreases (0.94% and 1.22% in w-F_1), validating its effectiveness. Similarly, removing the orthogonality constraint $\mathcal{L}_{orth}$ causes performance to drop by 1.71% and 1.08% on respective datasets, confirming its importance in extracting complementary information that cannot be derived from any single modality.

4 Conclusion

In this paper, we propose KCIF, a novel framework for ICU disease prediction that effectively integrates multimodal clinical data and external medical KGs. By constructing a THAG, our approach successfully captures both semantic and temporal dependencies across admission history. The proposed complementary information fusion mechanism enables lab tests and medical events to reinforce each other, leading to better patient modeling. Extensive experiments on the MIMIC-III/IV benchmarks demonstrate the effectiveness of KCIF.

Acknowledgments. This study was funded National Key R&D Program of China (2022YFA1008300), National Natural Science Foundation of China (12471308, 62472415), Excellent Young Scholars of Shenzhen (RCYX20231211090247060).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Burger, M., Rätsch, G., Kuznetsova, R.: Multi-modal graph learning over umls knowledge graphs. In: Proceedings of Machine Learning for Health Symposium. vol. 225, pp. 52–81 (2023)
2. Choi, E., Bahadori, M.T., Song, L., Stewart, W.F., Sun, J.: Gram: graph-based attention model for healthcare representation learning. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 787–795 (2017)
3. Choi, E., Bahadori, M.T., Sun, J., Kulas, J., Schuetz, A., Stewart, W.: Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In: Proceedings of the Advances in Neural Information Processing Systems. vol. 29 (2016)
4. Harutyunyan, H., Khachatrian, H., Kale, D.C., Ver Steeg, G., Galstyan, A.: Multitask learning and benchmarking with clinical time series data. Scientific Data **6**(1), 96 (2019)
5. Huang, Y., Zhang, R., Deng, Y., Meng, M.: Accuracy of physician and nurse predictions for 28-day prognosis in icu: a single center prospective study. Scientific Reports **13**(1), 22023 (2023)
6. Jiang, P., Xiao, C., Cross, A.R., Sun, J.: Graphcare: Enhancing healthcare predictions with personalized knowledge graphs. In: Proceedings of the International Conference on Learning Representations (2024)
7. Johnson, A.E., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., Pollard, T.J., Moody, B., Gow, B., Lehman, L.w.H., et al.: Mimic-iv, a freely accessible electronic health record dataset. Scientific Data **10**(1), 1 (2023)
8. Johnson, A.E., Pollard, T.J., Shen, L., Lehman, L.w.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., Mark, R.G.: Mimic-iii, a freely accessible critical care database. Scientific Data **3**(1), 1–9 (2016)
9. King, R., Yang, T., Mortazavi, B.J.: Multimodal pretraining of medical time series and notes. In: Proceedings of Machine Learning for Health. pp. 244–255 (2023)

10. Lu, C., Han, T., Ning, Y.: Context-aware health event prediction via transition functions on dynamic disease graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 4567–4574 (2022)
11. Luo, J., Ye, M., Xiao, C., Ma, F.: Hitanet: Hierarchical time-aware attention networks for risk prediction on electronic health records. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 647–656 (2020)
12. Ma, F., Chitta, R., Zhou, J., You, Q., Sun, T., Gao, J.: Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1903–1911 (2017)
13. Ma, F., You, Q., Xiao, H., Chitta, R., Zhou, J., Gao, J.: Kame: Knowledge-based attention model for diagnosis prediction in healthcare. In: Proceedings of the ACM International Conference on Cnformation and Knowledge Management. pp. 743–752 (2018)
14. Olang, O., Mohseni, S., Shahabinezhad, A., Hamidianshirazi, Y., Goli, A., Abolghasemian, M., Shafiee, M.A., Aarabi, M., Alavinia, M., Shaker, P.: Artificial intelligence-based models for prediction of mortality in icu patients: a scoping review. Journal of Intensive Care Medicine (2024)
15. Park, S., Bae, S., Kim, J., Kim, T., Choi, E.: Graph-text multi-modal pre-training for medical representation learning. In: Proceedings of Health, Inference, and Learning. pp. 261–281 (2022)
16. Taha, A., Jacquier, M., Meunier-Beillard, N., Ecarnot, F., Andreu, P., Roudaut, J.B., Labruyère, M., Rigaud, J.P., Quenot, J.P.: Anticipating need for intensive care in the healthcare trajectory of patients with chronic disease: A qualitative study among specialists. Plos One **17**(9), e0274936 (2022)
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I.: Attention is all you need. In: Proceedings of Advances in Neural Information Processing Systems. vol. 30 (2017)
18. Wang, Z., Wu, Z., Agarwal, D., Sun, J.: Medclip: Contrastive learning from unpaired medical images and text. In: Proceedings of Empirical Methods in Natural Language Processing. pp. 3876–3887 (2022)
19. Xu, R., Shi, W., Yu, Y., Zhuang, Y., Jin, B., Wang, M.D., Ho, J., Yang, C.: RAM-EHR: Retrieval augmentation meets clinical predictions on electronic health records. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics. pp. 754–765 (2024)
20. Xu, Y., Chu, X., Yang, K., Wang, Z., Zou, P., Ding, H., Zhao, J., Wang, Y., Xie, B.: Seqcare: Sequential training with external medical knowledge graph for diagnosis prediction in healthcare data. In: Proceedings of the ACM Web Conference. pp. 2819–2830 (2023)
21. Xu, Y., Yang, K., Zhang, C., Zou, P., Wang, Z., Ding, H., Zhao, J., Wang, Y., Xie, B.: Vecocare: Visit sequences-clinical notes joint learning for diagnosis prediction in healthcare data. In: Proceedings of the International Joint Conference on Artificial Intelligence. vol. 23, pp. 4921–4929 (2023)
22. Xu, Z., So, D.R., Dai, A.M.: Mufasa: Multimodal fusion architecture search for electronic health records. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 10532–10540 (2021)
23. Yang, K., Xu, Y., Zou, P., Ding, H., Zhao, J., Wang, Y., Xie, B.: Kerprint: local-global knowledge graph enhanced diagnosis prediction for retrospective and prospective interpretations. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 5357–5365 (2023)

24. Yang, Z., Lin, Y., Xu, Y., Hu, J., Dong, S.: Interpretable disease prediction via path reasoning over medical knowledge graphs and admission history. Knowledge-Based Systems **281**, 111082 (2023)
25. Zhang, X., Li, S., Chen, Z., Yan, X., Petzold, L.R.: Improving medical predictions by irregular multimodal electronic health records modeling. In: Proceedings of International Conference on Machine Learning. pp. 41300–41313 (2023)
26. Zhu, Y., Ren, C., Wang, Z., Zheng, X., Xie, S., Feng, J., Zhu, X., Li, Z., Ma, L., Pan, C.: Emerge: Integrating rag for improved multimodal ehr predictive modeling. arXiv preprint arXiv:2406.00036 (2024)