CAND: Cross-Sign Ambiguity Inference for Early Detecting Nuanced Illness Deterioration

Lo Pang-Yun Ting[♠], Zhen Tan[♣], Hong-Pei Chen[♠], Cheng-Te Li[♠], Po-Lin Chen[♦], Kun-Ta Chuang[♠], Huan Liu[♣]

[♠]Dept. of Computer Science and Information Engineering, National Cheng Kung University
[♣]School of Computing and Augmented Intelligence, Arizona State University
[♦]Division of Infectious Diseases, National Cheng Kung University Hospital
{lpyting, hpchen, chengte}@netdb.csie.ncku.edu.tw,
{cplin, ktchuang}@mail.ncku.edu.tw, {ztan36, huanliu}@asu.edu

Abstract

Early detection of illness deterioration is essential for timely treatment, with vital signs like heart rates being key health indicators. Existing methods tend to solely analyze vital sign waveforms, ignoring transition relationships of waveforms within each vital sign and the correlation strengths among various vital signs. In this paper, we introduce *CAND*, a novel method that organizes the transition relationships and the correlations within and among vital signs as sign-specific and cross-sign knowledge. *CAND* jointly models these knowledge in a unified representation space and integrates a Bayes-based inference method that utilizes augmented knowledge from sign-specific and cross-sign knowledge to address the ambiguities in correlation strengths. Our experiments on a real-world ICU dataset demonstrate that *CAND* significantly outperforms existing methods in both effectiveness and earliness in detecting nuanced illness deterioration. We further conduct a detailed case study, showing the interpretable detection process and practicality of *CAND*.

1 Introduction

In intensive care units (ICUs), real-time monitoring and early detection of illness deterioration are crucial [50]. Previous research [38, 36] stresses the role of vital signs, such as heart rate or body temperature, as critical indicators for monitoring hospitalized patients [3]. Wearable devices offer a cost-effective solution for continuous monitoring of vital signs, allowing real-time and non-invasive data collection while reducing the risk of *nosocomial infections* [40].

Increases in illness severity scores [41] indicate **nuanced illness deterioration**, indicating a gradual decline in patient health that is not severe enough, like death or ICU admission, to trigger alarms. Detecting such subtle changes early allows physicians to intervene flexibly before conditions worsen.

Existing methods for vital sign analysis focus either on extracting critical concepts or patterns from a single vital-sign time series[1], such as discovering "shapelets" [55], which are informative waveforms or time-series subsequences [2, 23, 59], or analyze multiple vital signs [12, 26, 42]. However, these methods are generally designed to detect specific diseases or major deterioration events (e.g., ICU admissions and mortality) and may fail to capture *nuanced illness deterioration*, as they overlook two factors: ❶ transitions among key concepts (patterns) within a single vital sign, and ❷ correlations among key concepts across different vital signs. These factors become even more important in wearable device data, where measurement noise can obscure subtle changes in health status.

[1]A vital-sign time series refers to sequential measurements of patients' vital signs.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Learning from Time Series for Health.

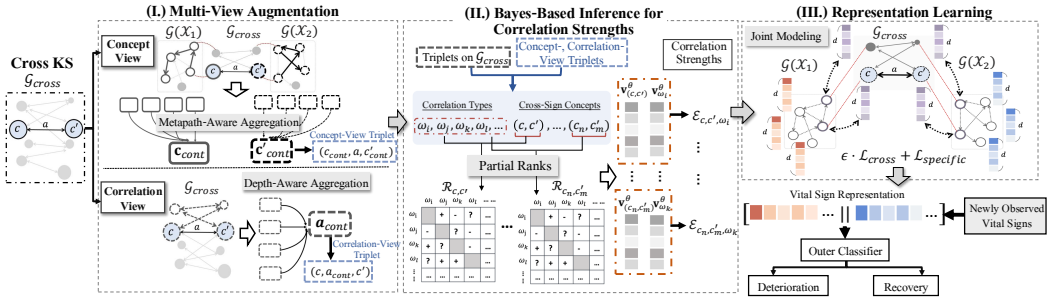


Figure 1: The overview of the *CAND* framework.

To enhance the early detection of nuanced illness deterioration from wearable vital signs, we model **sign-specific knowledge** as transition relationships within a single vital sign type and **cross-sign knowledge** as correlations between different vital sign types. Those knowledge can better capture subtle yet critical changes in vital signs and facilitate more effective information sharing among multiple sign types. We propose **C**ross-Sign **A**mbiguity Inference for Early Detecting **N**uanced Illness **D**eterioration, dubbed ***CAND***, which jointly models sign-specific and cross-sign knowledge in the representation space. These are then organized into distinct knowledge structures. A Bayes-based method is designed to infer correlation strengths to quantify cross-sign connections and guide their integration with sign-specific knowledge.

Our contributions are threefold: (**i**) We enable real-time patient monitoring with wearable devices by introducing a representation method for early detection of nuanced illness deterioration (illness severity scores increase) using low-cost vital sign data. (**ii**) We propose *CAND*, which formulates sign-specific and cross-sign knowledge structures and jointly models them, using a Bayes-based method to guide representation learning. (**iii**) Experiments on a real-world ICU dataset show that *CAND* outperforms baselines, and a case study shows its interpretable detection process.

2 Preliminary

We outline the key symbols and definitions. A knowledge graph (KG) is $\mathcal{G} = \{(h, r, t) \mid h, t \in \mathcal{E}, r \in \mathcal{R}\}$, where $\mathcal{E}$ and $\mathcal{R}$ are entity and relation sets, and each triplet (h, r, t) links head h to tail t via relation r. A scoring function $f(h, r, t)$ evaluates triplet plausibility. Given their ability to encode complex relations, KGs are ideal for modeling patient vital sign time series. For a vital sign type $\mathcal{X}$ (e.g., heart rate), the dataset $\mathcal{D}_{\mathcal{X}} = \{T^i_{\mathcal{X}}\}_{i=1}^{N}$ contains N patient vital-sign time series, each $T^i_{\mathcal{X}} \in \mathbb{R}^m$ over m timestamps. A **concept** is a shapelet [55], discovered via [16], forming a set $C_{\mathcal{X}} = \{c_1, c_2, \dots\}$ extracted from $\mathcal{D}_{\mathcal{X}}$. A Sign-Specific Knowledge Structure (**Specific KS**) is $\mathcal{G}(\mathcal{X}) = \{(c_i, \tau, c_j)\}$, where $c_i, c_j \in C_{\mathcal{X}}$ and τ encodes transition intervals between concepts from vital sign type $\mathcal{X}$. A Cross-Sign Knowledge Structure (**Cross KS**) is $\mathcal{G}_{cross} = \{(c, a, c')\}$, where c, c' are concepts from different specific KSs formulated from different vital sign types ($\mathcal{X}_1$, $\mathcal{X}_2$) and a denotes correlation strength. We employ the ComplEx model [49] for scoring. Finally, at each time slot p, given datasets $D_{\mathcal{X}_1}, D_{\mathcal{X}_2}$ and a patient's observed vital-sign time seires (from time 1 to p) $\{T'_{\mathcal{X}_1}, T'_{\mathcal{X}_2}\}$, our goal is to detect early patient deterioration before illness severity scores change.

3 The *CAND* Framework

Figure 4 presents an overview of our *CAND*, which consists of three main components, described in the following subsections.

3.1 Multi-View Augmentation

In our cross KS, stronger correlations between cross-sign concepts indicate closely related contexts with significant mutual influence. We employ multi-view augmentation to capture specific contexts of the cross KS from different views, serving as indicators of correlation strengths.

For each triplet (c, a, c') in cross KS $\mathcal{G}_{cross}$, we explore contexts from two perspectives: (1) contexts of cross-sign concepts (c, c') within their respective specific KSs, and (2) contexts of correlation a within the cross KS. These contexts are formulated into concept-view and correlation-view triplets.

Concept-View Triplet. We employ guided exploration inspired by node2vec [17] to focus on contexts most likely to be influenced by c and c' within their respective specific KSs. After N explorations of length L, the concept-view triplet is formulated as:

$$(c_{cont}, a, c'_{cont}) = \left(\{\psi_c^i\}_{i=1}^N \, , \; a, \; \{\psi_{c'}^i\}_{i=1}^N \right), \tag{1}$$

where ψ_c^i and $\psi_{c'}^i$ denote the sequences of explored concepts. We then employ metapath-aware aggregation to construct representation vectors $\mathbf{c}_{cont}$ and $\mathbf{c}'_{cont}$ (detailed in Appendix B.1).

Correlation-View Triplet. We explore contexts of correlation a within cross KS $\mathcal{G}_{cross}$ by considering path information between (c, c'). Let $\Pi_{\leq L'}$ be the set of all correlation paths from c to c' with length no longer than L'. The correlation-view triplet is:

$$(c, a_{cont}, c') = \left(c, \; \Pi_{\leq L'} \, , \; c' \right). \tag{2}$$

We employ depth-aware aggregation to construct the representation vector $\boldsymbol{a}_{cont}$ (detailed in Appendix B.2).

3.2 Bayes-Based Inference for Correlation Strengths

Inspired by Bayesian Personalized Ranking (BPR) [34], we design a Bayes-based method that merges plausibilities of original triplets from KSs, concept-view triplets, and correlation-view triplets as implicit feedback to infer correlation strengths. For each cross-sign concept pair $(c, c') \in \mathcal{U} \subset \mathcal{C}_{cross} \times \mathcal{C}_{cross}$, we define correlation types $\Omega = \bigcup_{a \in \mathcal{A}_{cross}} \{a_{\text{origin}}, a_{\text{concept}}, a_{\text{correlation}}\}$ representing correlations between c and c' from the origin, concept, and correlation views. We then establish partial ranks $\mathcal{R}_{(c,c')}$ based on plausibility differences (see Appendix B.3 for details). Using matrix factorization with BPR, we minimize the negative log-likelihood:

$$\mathcal{L}_{\text{infer}} = \sum_{(c,c') \in \mathcal{U}} \sum_{(\omega_i, \omega_j) \in \mathcal{R}_{(c,c')}} -\ln\Big(\text{sigmoid}(v_{c,c',\omega_i} - v_{c,c',\omega_j})\Big) + \lambda_\theta \left\|\theta\right\|^2, \tag{3}$$

where $v_{c,c',\omega_i} := \mathbf{v}^\theta_{(c,c')} \cdot \mathbf{v}^\theta_{\omega_i}$ based on latent factors of (c, c') and ω_i, and λ_θ are regularization parameters. The correlation strength between c and c' under type (view) ω is computed as:

$$\mathcal{E}_{c,c',\omega} = \frac{\exp(\mathbf{v}^\theta_{(c,c')} \cdot \mathbf{v}^\theta_{\omega})}{\sum_{\omega_j \in \Omega} \exp(\mathbf{v}^\theta_{(c,c')} \cdot \mathbf{v}^\theta_{\omega_j})}. \tag{4}$$

3.3 Representation Learning

We model the cross KS with inferred correlation strengths, which refines representations and influences specific KS training.

Cross-Sign Modeling. We adopt sigmoid loss with dynamic margins. For each triplet ν corresponding to correlation type ω and cross-sign concepts (c, c'), the dynamic margin is:

$$\gamma_\nu = \gamma \cdot \exp\Big((\mathcal{E}_{c,c',\omega} - 1) \cdot \xi\Big), \tag{5}$$

where γ is a fixed default margin and ξ is a scale factor.

Let $\mathcal{V}$ be the set of triplets in $\mathcal{G}_{cross}$, and $\mathcal{V}_{cont}(\nu)$ contains concept- and correlation-view triplets derived from $\nu \in \mathcal{V}$. The loss of representing cross KS is formulated as:

$$\mathcal{L}_{\text{cross}} = \frac{1}{|\mathcal{V}|} \sum_{\nu \in \mathcal{V}} \left\{ \mathcal{L}(\nu) + \lambda_c \cdot \frac{1}{|\mathcal{V}_{cont}(\nu)|} \sum_{\nu' \in \mathcal{V}_{cont}(\nu)} \mathcal{L}(\nu') \right\} + \mathcal{L}_{\text{infer}}, \tag{6}$$

where $\mathcal{L}(\nu)$ is the knowledge graph embedding loss function (we use the RotatE model [43]), which embeds each triplet ν with our estimated margin γ_ν, thereby controlling the representation learning based on the correlations among cross-sign concepts (detailed in Appendix B.4).

Specific-Sign Modeling. We model each specific KS using self-adversarial negative sampling [43] in knowledge graph representation, with the loss $\mathcal{L}_{\text{specific}} = \mathcal{L}_{\mathcal{G}(\mathcal{X}_1)} + \mathcal{L}_{\mathcal{G}(\mathcal{X}_2)}$ (detailed in Appendix B.5).

Joint Optimization. The final objective function combines cross-sign and sign losses:

$$\mathcal{L}_{CAND} = \epsilon \cdot \mathcal{L}_{\text{cross}} + \mathcal{L}_{\text{specific}}, \tag{7}$$

where $\epsilon > 0$ balances the two components. We alternately optimize these losses using Adam optimizer [24]. The learned representations are used to represent vital sign time series for real-time illness deterioration detection (details in Appendix E.1.2).

4 Experiments

We conduct experiments on real-world ICU patients' heart rate and skin temperature data collected via wearable devices from National Cheng Kung University Hospital (NCKUH)[2]. We compare against three categories of baselines: feature-based (Naive Bayes (NB) [35] and Elastic Ensembles (EE) [29]), early classification (ECTS [54], MCFEC [21], EARLIEST [19], and RHC [20]), and graph embedding (Time2Graph [10], KE-GCN [57] and AliNet [46]) methods. Performance is measured by Accuracy (Acc.), Recall (Rec.), F1-score (F1.), AUC, and Earliness (Ear.), which are combined into an overall Composite score (Comp.). **Details on the dataset, preprocessing, compared methods, and evaluation metrics are in Appendix C.**

4.1 Main Results

Table 1 shows that *CAND* consistently outperforms all baselines across different data pruning levels (randomly deleting varying percentages of training data), achieving the highest Accuracy, Recall, F1-score, and AUC. This highlights its robustness. While Time2Graph sometimes achieves a higher earliness score, *CAND* strikes a much better balance between detection effectiveness and earliness, as reflected by its superior overall composite score (55.68%). Also, other graph-based methods like AliNet and KE-GCN, which do not infer correlation strengths between vital signs, fall short of *CAND*'s performance. The strong performance, especially the high recall and competitive earliness, demonstrates our model's effectiveness in identifying deteriorating patients early.

Table 1: Performance comparison (in %) of early deterioration detection.

Methods	Pruning = 0%					Pruning = 30%					Pruning = 50%					Overall
	Acc. (↑)	Rec. (↑)	F1. (↑)	AUC (↑)	Ear. (↑)	Acc. (↑)	Rec. (↑)	F1. (↑)	AUC (↑)	Ear. (↑)	Acc. (↑)	Rec. (↑)	F1. (↑)	AUC (↑)	Ear. (↑)	Comp. (↑)
NB [35]	54.70	23.21	32.49	59.62	18.30	51.30	16.07	23.93	60.10	12.16	53.85	14.28	22.53	60.47	10.49	19.98
EE [29]	50.43	51.79	49.27	52.97	35.38	52.15	57.14	53.30	52.92	38.05	50.43	35.71	40.72	47.95	24.33	40.17
ECTS [54]	52.11	30.35	37.80	54.62	1.92	51.24	23.22	30.91	52.66	1.22	51.26	17.86	25.71	48.70	2.43	16.66
MCFEC [21]	51.29	37.50	29.58	51.01	33.93	51.29	33.93	28.36	47.03	30.41	47.03	30.35	25.37	47.43	25.86	28.91
EARLIEST [19]	54.74	41.07	46.47	58.46	19.31	53.84	25.00	33.89	54.16	16.74	50.41	10.71	17.16	52.71	4.80	23.06
RHC [20]	45.29	66.07	53.62	40.67	32.58	48.67	66.07	53.65	44.88	40.51	43.68	67.85	53.37	38.00	44.03	46.29
Time2Graph [10]	49.51	67.85	56.15	52.42	55.14	54.60	66.07	56.47	60.83	52.12	52.11	66.07	56.19	54.27	50.44	54.41
AliNet [46]	49.54	87.50	62.44	57.37	42.19	49.53	85.71	62.00	57.20	45.76	50.42	71.43	57.40	54.11	38.17	51.32
KE-GCN [57]	50.42	83.93	61.78	54.94	40.73	52.98	80.36	61.90	54.57	41.74	47.89	67.86	55.52	50.17	37.50	49.86
***CAND* (ours)**	**58.11**	**92.86**	**68.39**	**62.25**	43.08	**54.71**	**91.20**	**65.65**	**62.67**	47.32	**59.85**	**73.22**	**63.33**	**61.20**	46.32	**55.68**

4.2 Ablation Analysis

To validate the contributions of main components of *CAND*, we conduct an ablation study by examining the performance after removing concept-view triplets (w/o conc.), correlation-view triplets (w/o corr.), and the inference of correlation strengths (w/o infer.). The results on two pruned datasets (pruning level={0%, 50%}) are presented in Table 2. We observe that each component contributes to the overall performance, highlighting the necessity of incorporating these components in *CAND*.

4.3 Case Study: Interpretable Detection

To showcase the interpretability of *CAND*, Figure 2 visualizes the detection process for a deteriorating patient. In the first 30 minutes, observed concepts (shapelets) (c_5, c_6') suggest a potential occurrence

[2]The study protocol was approved by the Institutional Review Board (IRB) of NCKUH (No. B-BR-106-044 & No. A-ER-109-027).

Table 2: Results (in %) on the ablation study.

Methods	Pruning = 0%				Pruning = 50%			
	Acc. (↑)	F1. (↑)	AUC (↑)	Ear. (↑)	Acc. (↑)	F1. (↑)	AUC (↑)	Ear. (↑)
Full Model	**58.11**	**68.39**	**62.25**	**43.08**	**59.85**	**63.33**	**61.20**	**46.32**
w/o conc.	52.11	64.13	51.70	41.41	55.58	58.67	57.86	39.40
w/o corr.	50.42	62.81	57.38	36.72	56.37	59.88	58.50	43.64
w/o infer.	51.27	63.68	58.01	40.52	57.29	60.18	59.32	39.28

of (c'_{11}) with high correlation (77%), but the collective evidence does not indicate a high probability of deterioration. From 30 to 60 minutes, however, the newly observed concept (c_5), its strong correlation with a likely but unobserved concept (c'_7), and the possible transition to c'_8 provide critical evidence. *CAND* identifies that the combination of these observed and potential information has a high probability of occurring in deteriorating patients. This leads to a correct and early detection 420 minutes before the increase in illness severity score. This analysis reveals how inferred correlations can bridge information across different types of vital signs to form a coherent clinical picture, providing physicians with transparent reasoning to support their diagnostic decisions. Additional discussion is provided in Appendix F.

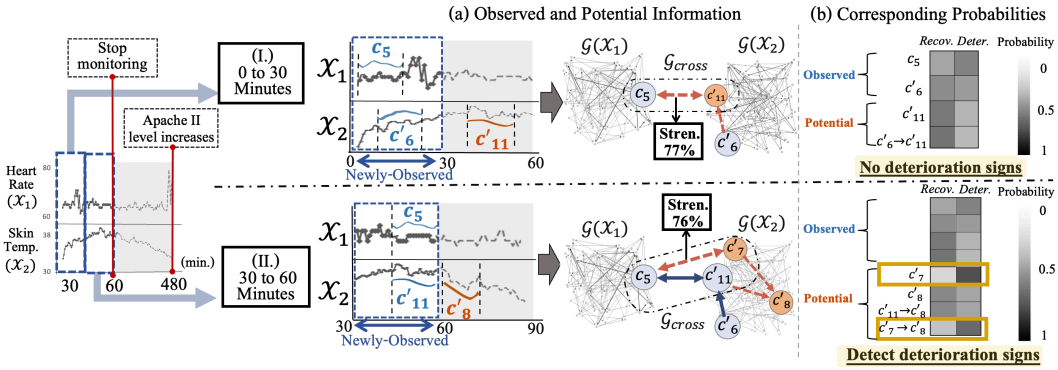


Figure 2: Visualization of the detection process for a deteriorating patient in two detection windows.

5 Conclusion

We present *CAND*, a framework that formulates sign-specific and cross-sign knowledge structures from wearable vital signs to detect nuanced illness deterioration early. It combines multi-view augmentation with Bayes-based inference to handle concept correlations and jointly model multiple knowledge structures. On real ICU data, *CAND* balances detection earliness and effectiveness while offering interpretable detection process, and its reliance on low-cost vital signs highlights its potential for remote healthcare.

Acknowledgments

We sincerely thank Prof. Nai-Ying Ko, Dr. Yi-Ting Chung, and Dr. Chang-Chun Chen from the Department of Nursing, as well as Prof. Chun-Yin Yeh from the Department of Nursing, College of Medicine, National Sun Yat-sen University, for their invaluable support in wearable data collection, patient coordination, and access to physiological data from NCKU Hospital.

In addition, this paper was supported in part by National Science and Technology Council (NSTC), R.O.C., under Contract 113-2221-E-006-203-MY2, 114-2622-8-006-002-TD1 and 114-2634-F-006-001-MBK.

References

[1] T. I. Alshwaheen, Y. W. Hau, N. Ass'Ad, and M. M. Abualsamen. A novel and reliable framework of patient deterioration prediction in intensive care unit based on long short-term memory-recurrent neural network. *IEEE Access*, 9:3894–3918, 2021.

[2] C. Bock, T. Gumbsch, M. Moor, B. Rieck, D. Roqueiro, and K. Borgwardt. Association mapping in biomedical time series via statistically significant shapelet mining. *Bioinformatics*, 34(13):i438–i446, 2018.

[3] I. J. Brekke, L. H. Puntervoll, P. B. Pedersen, J. Kellett, and M. Brabrand. The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review. *PloS one*, 14(1):e0210875, 2019.

[4] Y. Cao, Z. Liu, C. Li, Z. Liu, J.-Z. Li, and T.-S. Chua. Multi-channel graph neural network for entity alignment. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1452–1461, 2019.

[5] M. Chen, Y. Tian, M. Yang, and C. Zaniolo. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017.

[6] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[7] X. Chen, M. Boratko, M. Chen, S. S. Dasgupta, X. L. Li, and A. McCallum. Probabilistic box embeddings for uncertain knowledge graph reasoning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 882–893, 2021.

[8] X. Chen, M. Chen, W. Shi, Y. Sun, and C. Zaniolo. Embedding uncertain knowledge graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3363–3370, 2019.

[9] Z.-M. Chen, M.-Y. Yeh, and T.-W. Kuo. Passleaf: A pool-based semi-supervised learning framework for uncertain knowledge graph embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4019–4026, 2021.

[10] Z. Cheng, Y. Yang, W. Wang, W. Hu, Y. Zhuang, and G. Song. Time2graph: Revisiting time series modeling with dynamic shapelets. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3617–3624, 2020.

[11] N. El-Rashidy, S. El-Sappagh, T. Abuhmed, S. Abdelrazek, and H. M. El-Bakry. Intensive care unit mortality prediction: An improved patient-specific stacking ensemble model. *IEEE Access*, 8:133541–133564, 2020.

[12] K. D. Fairchild, D. E. Lake, J. Kattwinkel, J. R. Moorman, D. A. Bateman, P. G. Grieve, J. R. Isler, and R. Sahni. Vital signs and their cross-correlation in sepsis and nec: a study of 1,065 very-low-birth-weight infants in two nicus. *Pediatric research*, 81(2):315–321, 2017.

[13] M. F. Ghalwash, V. Radosavljevic, and Z. Obradovic. Extraction of interpretable multivariate patterns for early diagnostics. *2013 IEEE 13th International Conference on Data Mining*, pages 201–210, 2013.

[14] M. F. Ghalwash, V. Radosavljevic, and Z. Obradovic. Utilizing temporal patterns for estimating uncertainty in interpretable early decision making. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014.

[15] A. Godinjak, A. Iglica, A. Rama, I. Tančica, S. Jusufović, A. Ajanović, and A. Kukuljac. Predictive value of saps ii and apache ii scoring systems for patient outcome in a medical intensive care unit. *Acta medica academica*, 45 2:97–103, 2016.

[16] J. Grabocka, N. Schilling, M. Wistuba, and L. Schmidt-Thieme. Learning time-series shapelets. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 392–401, 2014.

[17] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

[18] A. Gupta, H. P. Gupta, B. Biswas, and T. Dutta. Approaches and applications of early classification of time series: A review. *IEEE Transactions on Artificial Intelligence*, 1(1):47–61, 2020.

[19] T. Hartvigsen, C. Sen, X. Kong, and E. Rundensteiner. Adaptive-halting policy network for early classification. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 101–110, 2019.

[20] T. Hartvigsen, C. Sen, X. Kong, and E. Rundensteiner. Recurrent halting chain for early multi-label classification. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1382–1392, 2020.

[21] G. He, Y. Duan, R. Peng, X. Jing, T. Qian, and L. Wang. Early classification on multivariate time series. *Neurocomputing*, 149:777–787, 2015.

[22] Y. Huang, G. G. Yen, and V. S. Tseng. Snippet policy network for multi-class varied-length ecg early classification. *IEEE Transactions on Knowledge and Data Engineering*, 35:6349–6361, 2021.

[23] S. L. Hyland, M. Faltys, M. Hüser, X. Lyu, T. Gumbsch, C. Esteban, C. Bock, M. Horn, M. Moor, B. Rieck, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature medicine*, 26(3):364–373, 2020.

[24] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In Y. Bengio and Y. LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015*, 2015.

[25] W. Knaus, E. Draper, D. Wagner, and J. Zimmerman. Apache ii: a severity of disease classification system. *Critical care medicine*, 13:818–29, 11 1985.

[26] N. Kumar, G. Akangire, B. A. Sullivan, K. D. Fairchild, and V. Sampath. Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront. *Pediatric Research*, 87:210 – 220, 2019.

[27] D. Li, P. G. Lyons, C. Lu, and M. H. Kollef. Deepalerts: Deep learning based multi-horizon alerts for clinical deterioration on oncology hospital wards. In *AAAI Conference on Artificial Intelligence*, 2020.

[28] S. Liang. Knowledge graph embedding based on graph neural network. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 3908–3912. IEEE, 2023.

[29] J. Lines and A. Bagnall. Time series classification with ensembles of elastic distance measures. *Data Mining and Knowledge Discovery*, 29:565–592, 2015.

[30] X. Mao, W. Wang, H. Xu, Y. Wu, and M. Lan. Relational reflection entity alignment. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 1095–1104, 2020.

[31] S. A. Naved, S. Siddiqui, and F. H. Khan. Apache-ii score correlation with mortality and length of stay in an intensive care unit. *Journal of the College of Physicians and Surgeons Pakistan*, 21(1):4, 2011.

[32] A. I. Paganelli, P. E. Velmovitsky, A. Branco, M. Endler, P. P. Morita, P. S. C. Alencar, and D. D. Cowan. A novel self-adaptive method for improving patient monitoring with composite early-warning scores. *2022 IEEE International Conference on Big Data (Big Data)*, pages 201–208, 2022.

[33] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710, 2014.

[34] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 452–461, 2009.

[35] I. Rish et al. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46, 2001.

[36] A. A. Romanovsky, M. C. Almeida, D. M. Aronoff, A. I. Ivanov, J. P. Konsman, A. A. Steiner, and V. F. Turek. Fever and hypothermia in systemic inflammation: recent discoveries and revisions. *Front Biosci*, 10(1-3):2193–2216, 2005.

[37] M. A. Rose, L. A. Hanna, S. A. Nur, and C. M. Johnson. Utilization of electronic modified early warning score to engage rapid response team early in clinical deterioration. *Journal for Nurses in Professional Development*, 31:E1–E7, 2015.

[38] F. Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5, 2017.

[39] F. E. Shamout, T. Zhu, P. Sharma, P. J. Watkinson, and D. A. Clifton. Deep interpretable early warning system for the detection of clinical deterioration. *IEEE Journal of Biomedical and Health Informatics*, 24:437–446, 2020.

[40] A. Sikora and F. Zahra. Nosocomial infections. 2020.

[41] K. Strand and H. Flaatten. Severity scoring in the icu: a review. *Acta Anaesthesiologica Scandinavica*, 52(4):467–478, 2008.

[42] B. Sullivan, V. Nagraj, K. Berry, N. Fleiss, A. Rambhia, R. Kumar, A. Wallman-Stokes, Z. Vesoulis, R. Sahni, S. Ratcliffe, et al. Clinical and vital sign changes associated with late-onset sepsis in very low birth weight infants at 3 nicus. *Journal of neonatal-perinatal medicine*, 14(4):553–561, 2021.

[43] Z. Sun, Z. Deng, J. Nie, and J. Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *7th International Conference on Learning Representations, ICLR 2019*, 2019.

[44] Z. Sun, W. Hu, and C. Li. Cross-lingual entity alignment via joint attribute-preserving embedding. In *The Semantic Web–ISWC 2017: 16th International Semantic Web Conference*, pages 628–644, 2017.

[45] Z. Sun, W. Hu, Q. Zhang, and Y. Qu. Bootstrapping entity alignment with knowledge graph embedding. In *International Joint Conference on Artificial Intelligence*, 2018.

[46] Z. Sun, C. Wang, W. Hu, M. Chen, J. Dai, W. Zhang, and Y. Qu. Knowledge graph alignment network with gated multi-hop neighborhood aggregation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 222–229, 2020.

[47] L. P. Ting, H. Chen, A. Liu, C. Yeh, P. Chen, and K. Chuang. Early detection of patient deterioration from real-time wearable monitoring system. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2025*, pages 9871–9879, 2025.

[48] B. D. Trisedya, J. Qi, and R. Zhang. Entity alignment between knowledge graphs using attribute embeddings. In *AAAI Conference on Artificial Intelligence*, 2019.

[49] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080, 2016.

[50] E. E. Vasilevskis, M. W. Kuzniewicz, M. L. Dean, T. Clay, E. Vittinghoff, D. J. Rennie, and R. A. Dudley. Relationship between discharge practices and intensive care unit in-hospital mortality performance: Evidence of a discharge bias. *Medical Care*, 47:803–812, 2009.

[51] Z. Wang, Q. Lv, X. Lan, and Y. Zhang. Cross-lingual knowledge graph alignment via graph convolutional networks. In *Conference on Empirical Methods in Natural Language Processing*, 2018.

[52] R. Wu, A. Der, and E. J. Keogh. When is early classification of time series meaningful? *IEEE Transactions on Knowledge and Data Engineering*, 35(3):3253–3260, 2021.

[53] K. Xin, Z. Sun, W. Hua, W. Hu, J. Qu, and X. Zhou. Large-scale entity alignment via knowledge graph merging, partitioning and embedding. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 2240–2249, 2022.

[54] Z. Xing, J. Pei, and P. S. Yu. Early classification on time series. *Knowledge and Information Systems*, 31:105–127, 2012.

[55] L. Ye and E. J. Keogh. Time series shapelets: a new primitive for data mining. In *Knowledge Discovery and Data Mining*, 2009.

[56] R. Ye, X. Li, Y. Fang, H. Zang, and M. Wang. A vectorized relational graph convolutional network for multi-relational network alignment. In *IJCAI*, pages 4135–4141, 2019.

[57] D. Yu, Y. Yang, R. Zhang, and Y. Wu. Knowledge embedding based graph convolutional network. In *Proceedings of the web conference 2021*, pages 1619–1628, 2021.

[58] Q. Zhang, Z. Sun, W. Hu, M. Chen, L. Guo, and Y. Qu. Multi-view knowledge graph embedding for entity alignment. In *International Joint Conference on Artificial Intelligence*, 2019.

[59] L. Zhao, H. Liang, D. Yu, X. Wang, and G. Zhao. Asynchronous multivariate time series early prediction for icu transfer. In *Proceedings of the 2019 International Conference on Intelligent Medicine and Health*, pages 17–22, 2019.

[60] H. Zhu, R. Xie, Z. Liu, and M. Sun. Iterative entity alignment via joint knowledge embeddings. In *International Joint Conference on Artificial Intelligence*, 2017.

A Related Work

Early Detection of Patient Deterioration. Previous studies focus on two main categories: detecting severe events (like death, ICU admissions, or cardiopulmonary arrest) and early diagnosis of specific diseases. For severe events, most of the research utilizes Electronic Health Records (EHR). Li *et al.* [27] propose a multi-task model for early detecting death and ICU admissions. El-Rashidy *et al.* [11] design stacking ensemble classifiers for mortality prediction. Alshwaheen *et al.* [1] design a deep genetic algorithm for similar purposes. Hartvigsen *et al.* [19] propose the EARLIEST method to analyze time series data in EHR. Shamout *et al.* [39] develop a neural network to predict deterioration from various data. Zhao *et al.* [59] and Hyland *et al.* [23] extract shapelets from vital signs to predict ICU admissions and circulatory failure. For specific diseases, Ghalwash *et al.* extract shapelets from ECG and blood gene expression for early diagnosis of specific diseases [13], and they also introduce a shapelet-based early classification method to measure temporal uncertainty [14]. Huang *et al.* [22] develop a snippet policy network for early cardiovascular disease classification. However, these studies require higher costs to access data since they analyze various types of patients' data rather than just vital signs, and face various issues [52]. Other methods [13, 14, 22, 59] analyze only ECG signals or vital signs but only focus on detecting specific diseases or severe events, or on detecting nuanced deterioration while neglecting the correlations among multiple signs [47].

Paganelli *et al.* [32] design a self-adaptive algorithm to build an early warning system for illness severity deterioration, but this work is limited to utilizing five specific vital signs. Therefore, there is a lack of a general method to early detect nuanced deterioration in various patients' conditions using low-cost vital sign data and analyze correlations among them to further improve performance.

Knowledge Graph Alignment. Fusing embeddings from different knowledge graphs (KGs) or structures through entity (node) alignments is crucial, with research emphasizing semantic matching and GNN-based models. Semantic models like MTransE [5], IPTransE [60], and BootEA [45] use TransE for relational learning, while others like JAPE [44], AttrE [48], and MultiKE [58] enhance entity semantics with diverse methods. GNN-based models, such as GCN-Align [51] and MuGNN [4], focus on parameter sharing and filling missing entities. AVG-GCN [56], AliNet [46], and RREA [30] advance embedding learning by considering entity structures and relational reflection. KE-GCN [57] and LargeGNN [53] update embeddings jointly and tackle scalability. Unlike these models that

treat aligned entities as identical, our approach recognizes aligned entities as distinct shapelets from different knowledge structures, emphasizing the inference of correlation strengths between them for enhanced representations. This introduces a novel method for modeling multiple knowledge structures.

Uncertain Knowledge Graph Embedding. Recent interest in uncertain KGs has led to the development of methods assigning confidence scores to triplets for more precise reasoning. Existing models like UKGE [8], PASSLEAF [9], BEUrRE [7], and BGNN [28], which focus on embedding within UKGs to predict triplet confidence. These methods, however, primarily concentrate on learning a single uncertain KG. Our approach differs by jointly modeling multiple sign-specific knowledge (deterministic) and cross-sign knowledge (ambiguous), without relying on pre-defined confidence scores. We focus on inferring connectivity strengths and guiding joint learning of knowledge structures with both deterministic and ambiguous information. This offers a unique perspective in knowledge graph modeling with uncertain or ambiguous information.

B Detailed Method Descriptions

B.1 Concept-View Triplet Details

B.1.1 Guided Exploration

Instead of simply considering n-hop neighbors, we employ a **guided exploration**, inspired by node2vec [17], to focus on contexts *most likely to be influenced by* c and c' *within their respective specific KSs*. Assuming c belongs to $\mathcal{G}_{cross}$ and specific KS $\mathcal{G}(\mathcal{X}_1)$, we demonstrate guided exploration for c in $\mathcal{G}(\mathcal{X}_1)$. Considering an exploration starting at c, proceeding through w to x, the unnormalized probability of moving to the next concept y is defined as follows:

$$\gamma^{x\to y} = \begin{cases} \frac{1}{\varphi}\cdot\mathcal{W}_{x,y} & , \text{ if } d^{w,y}\leq 1 \text{ or } y\in\mathcal{G}_{cross} \\ \varphi\cdot\mathcal{W}_{x,y} & , \text{ if } d^{w,y}=2 \text{ and} x\in\mathcal{G}_{cross},\ y\notin\mathcal{G}_{cross} \\ \mathcal{W}_{x,y} & , \text{ otherwise} \end{cases}, \tag{8}$$

where $\mathcal{W}_{x,y}$ is the probability of concept y occurred after x in the historical dataset $\mathcal{D}_{\mathcal{X}_1}$. $d^{w,y}$ denotes the shortest distance between w and y on $\mathcal{G}(\mathcal{X}_1)$. Here, $\varphi > 1$ biases the exploration towards a depth-first search, prioritizing concepts that are away from w and do not belong to $\mathcal{G}_{cross}$ while avoiding revisits.

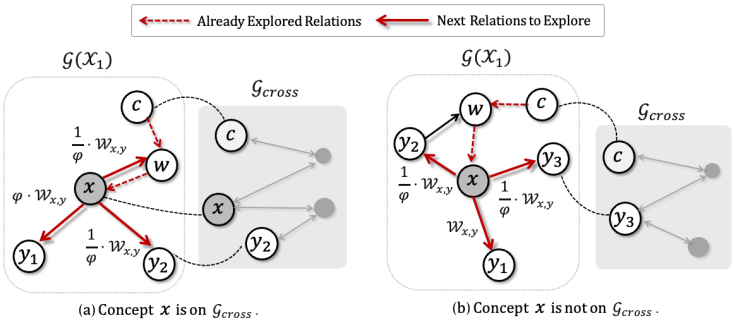


Figure 3: The example of the value of the unnormalized probability $\gamma^{x\to y}$ in the guided exploration.

To further how describe the designed unnormalized probability (Eq. 8) controls the guided exploration, we present two scenarios in Figure 3: when the current concept x is on $\mathcal{G}_{cross}$ (Figure 3a) and when it's not on $\mathcal{G}_{cross}$ (Figure 3b), by setting $\varphi > 1$ in Eq. 8.

For Figure 3a, $\gamma^{x\to y}$ is assigned a smaller value ($\frac{1}{\varphi}\cdot\mathcal{W}_{x,y}$) if:

1. The previous visited concept w is revisited (e.g., w in Figure 3a) or the next concept y is one step away from w ($d^{w,y}\leq 1$).
2. The next concept y belongs to $\mathcal{G}_{cross}$ (e.g., y_2 in 3a).

When $d^{w,y} = 2$ and $x \in \mathcal{G}_{cross},\ y \notin \mathcal{G}_{cross}$ (e.g., y_1 in Figure 3a), $\gamma^{x \to y}$ is assigned a larger value ($\varphi \cdot \mathcal{W}_{x,y}$). Similarly, in Figure 3b, $\gamma^{x \to y}$ is given a lower value ($\frac{1}{\varphi} \cdot \mathcal{W}_{x,y}$) either if $d^{w,y} \leq 1$ (e.g.,y_2 in Figure 3b) or if $y \in \mathcal{G}_{cross}$ (e.g.,y_3 in Figure 3b).

Formally, the guided exploration distribution for c is defined as $\mathcal{D}(y|x,c) = \gamma^{x \to y}/Z$ with Z as a normalizing constant as in [17].

B.1.2 Metapath-Aware Aggregation

For c_{cont}, we categorize all exploration sequences $\{\psi_c^i\}_{i=1}^N$ of c into different metapaths according to the types and order of relations encountered in $\mathcal{G}(\mathcal{X}_1)$. Taking ϕ as a metapath, we aggregate the representation vectors of concepts within each exploration ψ_c^i belonging to ϕ:

$$\mathbf{h}^{\phi} = \frac{1}{|\phi|} \sum_{\psi_c^i \in \phi} \sum_{e_j \in \psi_c^i} \alpha_j \mathbf{e}_j^t; \quad \alpha_j = \frac{\exp(p^{c \to e_j})}{\sum_{e_k \in \psi_c^i} \exp(p^{c \to e_k})}, \tag{9}$$

where e_j represents a concept within the exploration ψ_c^i, with its initial representation vector $\mathbf{e}_j^t \in \mathbb{R}^d$ at current epoch t. The weight α_j is determined by $p^{c \to e_j}$, which represents the probability of exploring from c to e_j via the guided exploration distribution $\mathcal{D}(y|x,c)$.

Let S_{meta} be the set of all metapaths, the final representation vector for c_{cont} is:

$$\mathbf{c}_{cont} = \sum_{\phi \in S_{meta}} \frac{\mathbf{h}^{\phi}}{|S_{meta}|}. \tag{10}$$

B.2 Correlation-View Triplet Details

B.2.1 Correlation Path Definition

Assuming a raw path from c to c' in cross KG $\mathcal{G}_{cross}$ is a sequence of concepts and correlations: $c(e_0) \xrightarrow{a_1} e_1 \xrightarrow{a_2} e_2 \ldots e_{n-1} \xrightarrow{a_l} c'(e_n)$. The corresponding *correlation path* $\pi = \{a_1, a_2, ..., a_n\}$ consists of all correlations in the raw path. $\Pi_{\leq L'}$ denotes the set of all correlation paths from c to c' in $\mathcal{G}_{cross}$, with each correlation path $\pi \in \Pi_{\leq L'}$ no longer than L'.

B.2.2 Depth-Aware Aggregation

Let $\Pi_{=l}$ represent the subset of correlation paths from c to c' with identical length l ($\Pi_{=l} \subset \Pi_{\leq L'}$). We aggregate the representation vector of each correlation in $\Pi_{=l}$:

$$\mathbf{h}^l = \frac{1}{|\Pi_{=l}|} \sum_{\pi \in \Pi_{=l}} \sum_{a_j \in \pi} \beta_j \boldsymbol{a}_j^t; \quad \beta_j = \frac{\exp\big(f(e_{j-1}, a_j, e_j)\big)}{\sum_{a_k \in \pi} \exp\big(f(e_{k-1}, a_k, e_k)\big)}, \tag{11}$$

where $\boldsymbol{a}_j^t \in \mathbb{R}^d$ denotes the representation vector of correlation a_j at current epoch t. The term β_j reflects the weight of a_j. Here, (e_{j-1}, a_j, e_j) represents a triplet in a raw path corresponding to correlation path π, and $f(\cdot)$ is the scoring function.

The final representation vector for a_{cont} is obtained by aggregating representations of correlation path subsets:

$$\boldsymbol{a}_{cont} = \sum_{\Pi_{=l} \in S_{path}} \frac{\mathbf{h}^l}{|S_{path}|}, \tag{12}$$

where S_{path} is the set of correlation path subsets, with each subset $\Pi_{=l}$ having $l \leq L'$.

B.3 Bayes-Based Inference Details

Definition 5 (Correlation Type): We consider each triplet $(c, a, c') \in \mathcal{G}_{cross}$, along with its contextual triplets, as embodying different correlation types: the origin view, the concept view, and the correlation view with correlation a in (c, c'). The set of correlation types is $\Omega = \bigcup_{a \in \mathcal{A}_{cross}} \{a_{\text{origin}}, a_{\text{concept}}, a_{\text{correlation}}\}$.

Definition 6 (Partial Rank): For each pair of cross-sign concepts $(c, c') \in \mathcal{U}$, if its plausibility with correlation type $\omega_i \in \Omega$ is higher than with $\omega_j \in \Omega$ and surpasses a predefined threshold, we denote a partial rank as $\omega_i >_{(c,c')} \omega_j$. The set with all partial ranks for (c, c') is $\mathcal{R}_{(c,c')} = \{(\omega_i, \omega_j) | \omega_i >_{(c,c')} \omega_j\}$. The plausibility estimation follows the scoring functions of knowledge graph embedding models, evaluating how likely a triplet is to represent a fact based on its current embeddings.

Our objective is to maximize the posterior probability $\mathcal{P}(\theta | >_{c,c'}) \propto \mathcal{P}(>_{c,c'} | \theta)\mathcal{P}(\theta)$, where $>_{c,c'}$ is the total ranking for (c, c'), and θ denotes the parameter vectors of matrix factorization.

B.4 Cross KS Modeling Details

In Eq. 20, we apply the RotatE model [43], a knowledge graph embedding model, to embed different triplets. Typically, knowledge graph embedding (KGE) aims to score positive triplets higher than negative ones. Let ν be a triplet on cross KS $\mathcal{G}_{\text{cross}}$ or a concept-view or correlation-view triplet. The loss for ν, used in Eq. 20, is defined as follows:

$$\mathcal{L}(\nu) = -\log \sigma\Big(\gamma_\nu - d(\nu)\Big) - \sum_{i=1}^{n}\Big(\frac{1}{n} \cdot \log \sigma\Big(d(\bar{\nu}_i) - \gamma_\nu\Big)\Big), \tag{13}$$

where $\sigma(\cdot)$ is the sigmoid function, and $\bar{\nu}_i$ represents the i^{th} negative triplet. The function $d(\cdot) = -f(\cdot)$, a negation of the scoring function $f(\cdot)$, serves as the distance function for a triplet. Each negative sample with equal importance $1/n$.

B.5 Specific KS Modeling Details

Using Eq. 20, we obtain the representation vector of each concept $c \in \mathcal{G}_{cross}$ at each epoch, which then serves as the default representation for training specific KSs. For modeling a specific KS like $\mathcal{G}(\mathcal{X}_1)$, with $\mathcal{S}$ denoting the set of triplets in $\mathcal{G}(\mathcal{X}_1)$, the objective function for modeling $\mathcal{G}(\mathcal{X}_1)$ is derived as follows:

$$\begin{aligned} \mathcal{L}_{\mathcal{G}(\mathcal{X}_1)} &= \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\Big(-\log\sigma\Big(\gamma - d(s)\Big) - \sum_{k=1}^{n} p(\bar{s}_k) \cdot \log\sigma\Big(d(\bar{s}_k) - \gamma\Big)\Big), \\ p(\bar{s}_k) &= \frac{\exp\ \alpha \cdot \big(f(\bar{s}_k)\big)}{\sum_{l=1}^{n} \exp\ \alpha \cdot \big(f(\bar{s}_l)\big)}, \end{aligned} \tag{14}$$

where $f(\cdot)$ is the scoring function of the ComplEx model [49]. $d(\cdot) = -f(\cdot)$ is a triplet's distance function. For a specific KS without ambiguities, we fix the margin value γ and define $p(\bar{s}_k)$ as the weight for the k^{th} negative sample, emphasizing the most informative ones [43]. The hyperparameter α is the temperature of self-adversarial weight [43].

C Experiment Setup

C.1 Dataset and Experimental Setup

C.1.1 Dataset and Preprocessing

We collect a real-world dataset from National Cheng Kung University Hospital (NCKUH)[3]. The dataset includes minute-by-minute vital signs from ICU patients, collected using wearable devices. It also features Apache II scores [25], a measure of illness severity, recorded at specific timestamps. We select 99 ICU patients from the dataset with the most vital signs records.

Patient Criteria. The study involves ICU patients between August 2020 and December 2021. The following conditions lead to the exclusion of certain patients: (i) Those who are physically or mentally disabled and unable to collaborate with researchers, (ii) Patients who may experience fever due to

[3]The study protocol was approved by the Institutional Review Board (IRB) of NCKUH (No. B-BR-106-044 & No. A-ER-109-027).

factors like rheumatic immunity, allergies, blood transfusions, malignant tumors, or other diseases, and (iii) Patients suffering from severe peripheral vascular disease.

Data Collection. Patients are required to wear devices to track heart rates and skin temperatures minutely. Additionally, patient data are collected through the hospital electronic system, including:

- Basic Information: age, occupation, medication usage, gender, BMI, and disease diagnosis.
- Vital Signs: body temperature, respiration rate, heart rate, blood pressure, and blood oxygen levels.
- Severity of Illness (Apache II scores [25]): the Apache II score (Acute Physiology and Chronic Health Evaluation II) is a widely used evaluation tool to assess the severity of disease in ICU patients, which can be treated as a useful indicator of illness deterioration or recovery over time [31]. Generally, a higher Apache II score corresponds to a more severe disease, and it is derived from a patient's age, medical history, and current physiological measurements, such as respiratory rate, serum sodium, blood pH, white blood cell count, Glasgow Coma Scale, among others. In our dataset, each patient has at least two Apache II score records, taken at different times during their ICU stay.

Each patient is required to wear the device for at least 14 days. This paper primarily focuses on data from wearable devices (heart rates, skin temperatures) and Apache II scores, with statistics for 99 selected ICU patients detailed in Table 3.

Table 3: Statistics (Statis.) of characteristic (Charac.) for selected 99 ICU patients.

	Gender		**Age**					**Death**	
Charac.	Female	Male	0-20	21-40	41-60	61-80	81-100	Yes	No
Statis. (%)	32.4	67.6	2.0	8.1	23.2	56.6	10.1	3.0	97.0

Charac.	**Avg. stay in ICUs / patient** (days)	**Avg. value of the Apache II score (std.)**
Statis.	29.63	16.39 (6.4)

Based on [15], we classify Apache II scores into **eight levels**: scores 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34 and above 35, corresponding to different potential mortality rates. In our experiments, an increase in the Apache II level signifies nuanced deterioration, while a decrease indicates recovery. For example, if a patient's Apache II score moves from the 0-4 range (level 1) to 5-9 (level 2), we annotate this patient as deteriorating. We select heart rate ($\mathcal{X}_1$) and skin temperature ($\mathcal{X}_2$) as the two primary vital signs. Based on [37], we analyze 8 hours of data prior to each Apache II level change, extracting 175 sets of vital sign measurements (**84 labeled as deterioration, 91 labeled as recovery**), each comprising 8-hour time series of heart rates and skin temperatures. Two-thirds of vital sign measurements are used for constructing knowledge structures and training the model, with the remaining third for testing. Each specific KS has 4 relation types for different time intervals between concepts. The cross KS includes 5 correlation types, each representing a range of likelihood for simultaneous occurrences of cross-sign concepts. To simulate real-world data scarcity, we randomly remove 30% and 50% of deteriorating data from the training dataset, creating pruned datasets with pruning levels of {0%, 30%, 50%}.

C.1.2 Compared Methods

Baselines are classified into three categories: feature-based (Naive Bayes (NB) [35] and Elastic Ensembles (EE) [29]), early classification (ECTS [54], MCFEC [21], EARLIEST [19], and RHC [20]), and graph embedding (Time2Graph [10], KE-GCN [57] and AliNet [46]). In our setup, Time2Graph constructs a directed graph for a vital sign type from extracted shapelets, training each graph independently. KE-GCN and AliNet, focusing on aligned knowledge graph embeddings, train our two specific KSs concurrently, treating cross-sign concepts as aligned and identical entities. Here, Time2Graph is considered to employ only sign-specific knowledge for training, while AliNet and KE-GCN utilize both sign-specific and cross-sign knowledge. Unlike our *CAND*, AliNet and KE-GCN do not infer the correlation strengths to influence the learning of each specific KS.

For feature-based methods, we select **Naive Bayes** [35] and Elastic Ensembles (**EE**) [29], using statistical features (mean, skew, standard deviation, kurtosis) from 30-minute intervals of patients' vital sign time series as input.

For early classification methods, we choose:

- **ECTS** [54]: ECTS is a prefix-based method aiming for early prediction on univariate time series. In our scenario, ECTS is applied individually to each vital sign time series. A patient is considered as deteriorating only when ECTS detects deterioration signs in time series from all types of vital signs. The earliest timestep at which deterioration is detected in any vital sign is then utilized for estimating the earliness score.
- **MCFEC** [21]: MCFEC is a shapelet-based method for the early classification of multivariate time series, utilizing distinctive and high-quality shapelets as core features. We employ its MCFEC-QBC (query by committee) classifier, which classifies time series based on the majority class among all matched shapelets.
- **EARLIEST** [19]: EARLIEST is a reinforcement learning-based method that supports early classification of both univariate and multivariate time series. It employs an RNN to model the transition dynamics of time series in conjunction with a policy network that decides at each timestep whether or not to halt the RNN and generate a prediction.
- **RHC** [20]: RHC extends EARLIEST [19] to support early multi-label classification for multivariate time series. It employs a transition model that jointly represents multivariate time series data and the conditional dependencies between labels. This approach potentially captures richer information, benefiting even single-label time series classification in our scenario.

For graph embedding methods, Time2Graph inherently incorporates shapelet transitions into a graph structure and is considered to employ only sign-specific knowledge for training. KE-GCN and AliNet, both aligned knowledge graph embedding models, simultaneously train our designed specific KSs, and the cross-sign concepts in our cross KS are regarded as aligned entities in KE-GCN and AliNet. Therefore, KE-GCN and AliNet are considered to utilize both sign-specific and cross-sign knowledge. The obtained embeddings are used to formulate time series representations in a similar way as our *CAND*. The following are detailed descriptions.

- **Time2Graph** [10]: Time2Graph learns *time-aware shapelets* for time series representations. A set of time series is converted into a directed graph, where nodes are shapelets and edges represent transitions between shapelets. DeepWalk [33] is used to derive shapelet representations, which are utilized to form the time series representation by considering the distance between shapelets and the time series. In our experiments, Time2Graph is employed and trained separately for the historical dataset of each vital sign. Representations of each vital sign time series are concatenated for classification as our method (Eq. 25). Note that Time2Graph only considers occurrence orders of shapelets, ignoring time intervals between them, and concentrates on the learning for a single graph.
- **AliNet** [46]: AliNet is an embedding model for entity-aligned knowledge graphs (KGs), uses an attention mechanism in GNNs (Graph Neural Networks) for multi-hop neighborhood aggregation to minimize representation differences of aligned entities across different KGs. In our experiments, AliNet trains our designed specific KSs simultaneously, treating cross-sign concepts in the cross KS as aligned entities.
- **KE-GCN** [57]: KE-GCN is an embedding model for entity-aligned or relation-aligned knowledge graphs (KGs), which combines GCNs in graph-based belief propagation and the knowledge graph embedding methods. In our experiments, KE-GCN is applied with the same settings as AliNet, training sign-specific knowledge structures (KSs) and treating cross-sign concepts as aligned entities.

C.1.3 Basic Setup

Except for early classification methods, to simulate real-time monitoring, other baselines and our *CAND* divide testing time series into 30-minute subsequences, each transmitted sequentially as observed data. Time2Graph, KE-GCN, AliNet, and *CAND* concatenate learned representation vectors to form observed time series representations, used as input for classification with the XGBoost model [6]. The monitoring is halted if nuanced illness deterioration is detected.

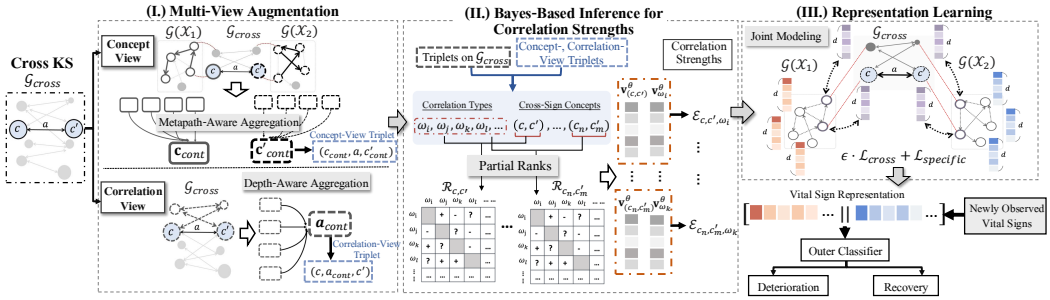


Figure 4: The overview of the *CAND* framework.

C.1.4 Evaluation Metrics

We present results averaged over 3-fold cross-validation on different pruned datasets, covering accuracy **(Acc.)**, recall **(Rec.)**, F1-score **(F1.)**, AUC value **(AUC)**, and earliness score **(Ear.)**. The earliness score is defined as $\frac{\mathrm{T}-t}{\mathrm{T}}$, following the setting in [18]. Here, T is the length (each unit represents one minute) of the complete time series, defined as 480 minutes (8 hours) in Sec. C.1.1, and t is the length (in minutes) of observed time series for detecting nuanced deterioration. A higher earliness score indicates earlier deterioration detection with fewer observations for vital signs. Additionally, we report overall composite score **(Comp.)** to summarize average performance across pruned datasets. The composite score is defined as $\frac{\text{F1-score}+\text{earliness score}}{2}$, revealing the model's capability in balancing detection effectiveness with earliness. Note that all evaluation metrics range from 0% to 100%, with higher values indicating better performance.

D The *CAND* Framework

Figure 4 presents an overview of our *CAND*, which consists of three main components, described in the following subsections.

D.1 Multi-View Augmentation

In our cross KS, stronger correlations between cross-sign concepts indicate closely related contexts with significant mutual influence. We employ multi-view augmentation to capture specific contexts of the cross KS from different views, serving as indicators of correlation strengths.

For each triplet (c, a, c') in cross KS $\mathcal{G}_{cross}$, we explore contexts from two perspectives: (1) contexts of cross-sign concepts (c, c') within their respective specific KSs, and (2) contexts of correlation a within the cross KS. These contexts are formulated into concept-view and correlation-view triplets.

Concept-View Triplet. We employ guided exploration inspired by node2vec [17] to focus on contexts most likely to be influenced by c and c' within their respective specific KSs. After N explorations of length L, the concept-view triplet is formulated as:

$$(c_{cont}, a, c'_{cont}) = \left(\{\psi_c^i\}_{i=1}^N \,,\; a,\; \{\psi_{c'}^i\}_{i=1}^N \right), \tag{15}$$

where ψ_c^i and $\psi_{c'}^i$ denote the sequences of explored concepts. We then employ metapath-aware aggregation to construct representation vectors $\mathbf{c}_{cont}$ and $\mathbf{c}'_{cont}$ (detailed in Appendix B.1).

Correlation-View Triplet. We explore contexts of correlation a within cross KS $\mathcal{G}_{cross}$ by considering path information between (c, c'). Let $\Pi_{\leq L'}$ be the set of all correlation paths from c to c' with length no longer than L'. The correlation-view triplet is:

$$(c, a_{cont}, c') = \left(c,\; \Pi_{\leq L'} \,,\; c' \right). \tag{16}$$

We employ depth-aware aggregation to construct the representation vector $\boldsymbol{a}_{cont}$ (detailed in Appendix B.2).

D.2 Bayes-Based Inference for Correlation Strengths

Inspired by Bayesian Personalized Ranking (BPR) [34], we design a Bayes-based method that merges plausibilities of original triplets from KSs, concept-view triplets, and correlation-view triplets as implicit feedback to infer correlation strengths. For each cross-sign concept pair $(c, c') \in \mathcal{U} \subset \mathcal{C}_{cross} \times \mathcal{C}_{cross}$, we define correlation types $\Omega = \bigcup_{a \in \mathcal{A}_{cross}} \{a_{\text{origin}}, a_{\text{concept}}, a_{\text{correlation}}\}$ representing correlations between c and c' from the origin, concept, and correlation views. We then establish partial ranks $\mathcal{R}_{(c,c')}$ based on plausibility differences (see Appendix B.3 for details). Using matrix factorization with BPR, we minimize the negative log-likelihood:

$$\mathcal{L}_{\text{infer}} = \sum_{(c,c') \in \mathcal{U}} \sum_{(\omega_i, \omega_j) \in \mathcal{R}_{(c,c')}} -\ln\Big(\text{sigmoid}(v_{c,c',\omega_i} - v_{c,c',\omega_j})\Big) + \lambda_\theta \left\|\theta\right\|^2, \tag{17}$$

where $v_{c,c',\omega_i} := \mathbf{v}^\theta_{(c,c')} \cdot \mathbf{v}^\theta_{\omega_i}$ based on latent factors of (c, c') and ω_i, and λ_θ are regularization parameters. The correlation strength between c and c' under type (view) ω is computed as:

$$\mathcal{E}_{c,c',\omega} = \frac{\exp(\mathbf{v}^\theta_{(c,c')} \cdot \mathbf{v}^\theta_{\omega})}{\sum_{\omega_j \in \Omega} \exp(\mathbf{v}^\theta_{(c,c')} \cdot \mathbf{v}^\theta_{\omega_j})}. \tag{18}$$

D.3 Representation Learning

We model the cross KS with inferred correlation strengths, which refines representations and influences specific KS training.

Cross-Sign Modeling. We adopt sigmoid loss with dynamic margins. For each triplet ν corresponding to correlation type ω and cross-sign concepts (c, c'), the dynamic margin is:

$$\gamma_\nu = \gamma \cdot \exp\Big((\mathcal{E}_{c,c',\omega} - 1) \cdot \xi\Big), \tag{19}$$

where γ is a fixed default margin and ξ is a scale factor.

Let $\mathcal{V}$ be the set of triplets in $\mathcal{G}_{cross}$, and $\mathcal{V}_{cont}(\nu)$ contains concept- and correlation-view triplets derived from $\nu \in \mathcal{V}$. The loss of representing cross KS is formulated as:

$$\mathcal{L}_{\text{cross}} = \frac{1}{|\mathcal{V}|} \sum_{\nu \in \mathcal{V}} \left\{ \mathcal{L}(\nu) + \lambda_c \cdot \frac{1}{|\mathcal{V}_{cont}(\nu)|} \sum_{\nu' \in \mathcal{V}_{cont}(\nu)} \mathcal{L}(\nu') \right\} + \mathcal{L}_{\text{infer}}, \tag{20}$$

where $\mathcal{L}(\nu)$ is the knowledge graph embedding loss function (we use the RotatE model [43]), which embeds each triplet ν with our estimated margin γ_ν, thereby controlling the representation learning based on the correlations among cross-sign concepts (detailed in Appendix B.4).

Specific-Sign Modeling. We model each specific KS using self-adversarial negative sampling [43] in knowledge graph representation, with the loss $\mathcal{L}_{\text{specific}} = \mathcal{L}_{\mathcal{G}(\mathcal{X}_1)} + \mathcal{L}_{\mathcal{G}(\mathcal{X}_2)}$ (detailed in Appendix B.5).

Joint Optimization. The final objective function combines cross-sign and sign losses:

$$\mathcal{L}_{CAND} = \epsilon \cdot \mathcal{L}_{\text{cross}} + \mathcal{L}_{\text{specific}}, \tag{21}$$

where $\epsilon > 0$ balances the two components. We alternately optimize these losses using Adam optimizer [24]. The learned representations are used to represent vital sign time series for real-time illness deterioration detection (details in Appendix E.1.2).

E Reproducibility

E.1 Implementation Details of *CAND*

E.1.1 Knowledge Structure Construction.

Specific KS. An example of constructing specific KS $\mathcal{G}(\mathcal{X}_1)$ is shown in Figure 5. We first extract concepts (shapelets) from historical dataset $\mathcal{D}_{\mathcal{X}_1} = \{T^i_{\mathcal{X}_1}\}_{i=1}^N$ (Def.1) using the existing method [16].

Setting the shapelet length (L) to 15 (15 minutes) and the number of shapelets (K) to 60, we can obtain a set of concepts (shapelets) $\mathcal{C}_{\mathcal{X}_1} \in \mathbb{R}^{K \times L}$, as shown in Figure 5 (a)(b). Subsequently, for each time series $T^i_{\mathcal{X}_1} \in \mathcal{D}_{\mathcal{X}_1}$, we estimate matching scores to determine which concepts in $\mathcal{C}_{\mathcal{X}_1}$ are matched to $T^i_{\mathcal{X}_1}$ (as in Figure 5(c)). Firstly, we divide $T^i_{\mathcal{X}_1}$ into n subsequences of length L, represented as $T^i_{\mathcal{X}_1} = \{v_1, ..., v_n\}$, with each subsequence $v_k = \{x_{l*(k-1)+1}, ..., x_{l*(k-1)+L}\}$, where each $x \in \mathbb{R}$ is a real-number reading of vital sign $\mathcal{X}_1$. The distance $\hat{d}(c, v)$ between a subsequence $v \in T^i_{\mathcal{X}_1}$ and a concept $c \in \mathcal{C}_{\mathcal{X}_1}$ is calculated as the average of their Euclidean and Dynamic Time Warping (DTW) distances. Therefore, the matching score δ_c of concept c to subsequence v is designed as follows:

$$\delta_c = \frac{\max_{c' \in \mathcal{C}_{\mathcal{X}_1}, v' \in T^i_{\mathcal{X}_1}} \hat{d}(c', v') - \hat{d}(c, v)}{\max_{c' \in \mathcal{C}_{\mathcal{X}_1}, v' \in T^i_{\mathcal{X}_1}} \hat{d}(c', v') - \min_{c' \in \mathcal{C}_{\mathcal{X}_1}, v' \in T^i_{\mathcal{X}_1}} \hat{d}(c', v')}. \tag{22}$$

The concept c is assigned to subsequence v if $\delta_c \geq 0.7$ and c is the top 3 similiest concepts to v among all concepts in $\mathcal{C}_{\mathcal{X}_1}$.

Next, we define the relation set $\mathcal{T}_{\mathcal{X}_1}$, with each relation $\tau \in \mathcal{T}_{\mathcal{X}_1}$ representing a specific time interval range between two concepts in each time series $T^i_{\mathcal{X}_1} \in \mathcal{D}_{\mathcal{X}_1}$. Specifically, we categorize these intervals into four ranges: τ_1 (0-30 minutes), τ_2 (30-60 minutes), τ_3 (60-90 minutes), and τ_4 (up to 90 minutes). Each $\tau_i \in \mathcal{T}_{\mathcal{X}_1}$. A triplet (c_i, τ, c_j) is formed if both c_i and c_j from $\mathcal{C}_{\mathcal{X}_1}$ are assigned to the same time series, with c_i occurring before c_j within the interval τ. As revealed in Figure 5 (d), this forms specific KS $\mathcal{G}(\mathcal{X}_1) = \{(c_i, \tau, c_j) | c_i, c_j \in \mathcal{C}_{\mathcal{X}_1}, \tau \in \mathcal{T}_{\mathcal{X}_1}\}$. A similar approach is used to construct specific KS $\mathcal{G}(\mathcal{X}_2)$.

Cross KS. Upon constructing the specific KSs $\mathcal{G}(\mathcal{X}_1)$ and $\mathcal{G}(\mathcal{X}_2)$, we estimate the likelihood of simultaneous occurrences for each concept pair (c, c'), where $c \in \mathcal{G}(\mathcal{X}_1)$ and $c' \in \mathcal{G}(\mathcal{X}_2)$. These likelihoods (excluding zero values) are divided into five ranges of equal intervals, forming the set of ambiguous correlations $\mathcal{A}_{cross} = \{a_1, a_2, ..., a_5\}$. Consequently, cross KS can be formed as $\mathcal{G}_{cross} = \{(c, a, c') | c, c' \in \mathcal{C}_{cross}, a \in \mathcal{A}_{cross}\}$, where $\mathcal{C}_{cross} \subset \mathcal{C}_{\mathcal{X}_1} \cup \mathcal{C}_{\mathcal{X}_2}$.

The illustration of specific KSs and the cross KS are shown in Figure 6

E.1.2 Representation Formulation for Time Series

After finishing the training of specific KSs and the cross KS. For the i^{th} set of vital sign time series $\{T^i_{\mathcal{X}_1}, T^i_{\mathcal{X}_2}\}$ in training data, where $T^i_{\mathcal{X}_1} \in \mathcal{D}_{\mathcal{X}_1}$ and $T^i_{\mathcal{X}_2} \in \mathcal{D}_{\mathcal{X}_2}$, we formulate the representation of the i^{th} measurement based on the occurred triplets in $T^i_{\mathcal{X}_1}$ and $T^i_{\mathcal{X}_2}$. Let $V(T^i_{\mathcal{X}_1})$ be triplet sets including triplets occurred in $T^i_{\mathcal{X}_1}$, where each triplet $\nu_n \in V(T^i_{\mathcal{X}_1})$ represents the n^{th} occurred triplet in $T^i_{\mathcal{X}_1}$. The important of triplet ν_n is designed as:

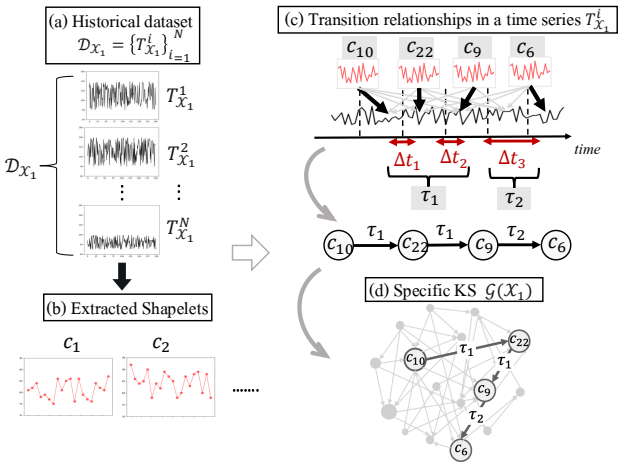


Figure 5: An example of constructing specific KS $\mathcal{G}(\mathcal{X}_1)$.

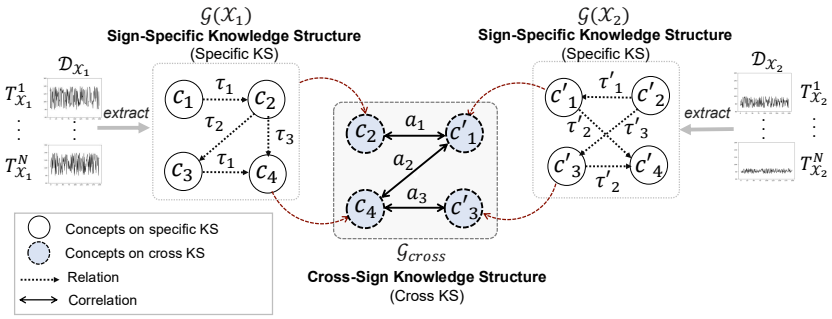


Figure 6: Illustration of sign-specific and cross-sign knowledge structures.

$$\mu(\nu_n) = \rho^{|V(T^i_{\mathcal{X}_1})|-n} \tag{23}$$

where ρ, set at 0.8, assigns higher importance to triplets that occur later. Then, the representation of $T^i_{\mathcal{X}_1}$ can be formulated as follows:

$$\mathbf{\Psi}^i_{\mathcal{X}_1} = \sum_{\nu_n=(c_i,\tau,c_j)\in V(T^i_{\mathcal{X}_1})} \frac{\mu(\nu_n)\cdot\left[\mathbf{c}_i||\boldsymbol{\tau}||\mathbf{c}_j\right]}{|V(T^i_{\mathcal{X}_1})|}, \tag{24}$$

where $\mathbf{c}_i$, $\boldsymbol{\tau}$ and $\mathbf{c}_j$ are representation vectors of concept c_i, relation τ and concept c_j, respectively.

The representation $\mathbf{\Psi}^i_{\mathcal{X}_2}$ of vital sign time series $T^i_{\mathcal{X}_2}$ is formulated similarly. Therefore, the final representation of the i^{th} set of vital sign time series $\{T^i_{\mathcal{X}_1}, T^i_{\mathcal{X}_2}\}$ is formed by concatenating $\mathbf{\Psi}^i_{\mathcal{X}_1}$ and $\mathbf{\Psi}^i_{\mathcal{X}_2}$ as follows:

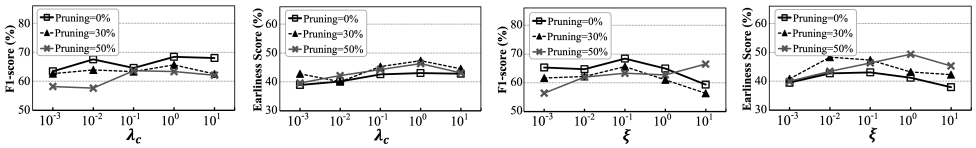
$$\mathbf{\Psi}^i_{\text{vital signs}} = [\mathbf{\Psi}^i_{\mathcal{X}_1}||\mathbf{\Psi}^i_{\mathcal{X}_2}]. \tag{25}$$

In the testing phase, representations for newly observed vital signs are constructed similarly.

F Additional Experiments

F.1 Parameter Sensitivity Analysis

To evaluate the effects of contextual loss weight (λ_c in Eq.20) and the scale factor for dynamic margin adjustment (ξ in Eq.19) on *CAND*, we vary λ_c and ξ within $\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1\}$. as shown in Figures 7. F1-score and earliness score of *CAND* improve as λ_c increases from 10^{-2} to 10^0 in the 30% and 50% pruned datasets, but decrease at 10^1. This suggests that excessively high weights on contextual triplets diminish the significance of original triplets in cross KS, degrading representation quality. For ξ, optimal results in 0% and 30% pruned datasets occur at 10^{-1} or 10^{-2}, highlighting the inefficacy of extreme margin adjustments.



(a) F1-score. (b) Earliness Score. (c) F1-score. (d) Earliness Score.

Figure 7: *CAND*'s performance with different weights of contextual loss (λ_c) and with different values of the scale factor (ξ).

F.2 Further Discussion

While our *CAND* model effectively balances detection earliness and effectiveness as shown in Table 1, enhancing its accuracy remains a challenge due to its sole reliance on vital sign data. Vital signs are significantly influenced by individual patient characteristics, including gender, age, and BMI. Future enhancements could focus on integrating these factors and exploring their causal relationships with vital sign dynamics. This could lead to a more nuanced understanding of illness deterioration signs and potentially improve the accuracy of CAND across diverse patient populations.

Additionally, while in ICUs, the earliness of detection is as crucial as the reliability of the results (e.g., accuracy, recall, F1-score), the importance placed on detection earliness versus reliability may vary in different scenarios. For example, in non-acute wards, where patient conditions are generally less urgent, earliness might be less critical. Investigating how the model could automatically adjust its decision-making process for patients with varying degrees of urgency could be a promising direction for future research. Moreover, the dataset currently in use involves a limited number of patients. One of our future objectives is to actively acquire more diverse patient data to enhance the generalizability of our model.