# BoxCare: A Box Embedding Model for Disease Representation and Diagnosis Prediction in Healthcare Data


Hang Lv
College of Computer and Data Science, Fuzhou University, China
lvhangkenn@gmail.com

Zehai Chen
College of Computer and Data Science, Fuzhou University, China
102101629@fzu.edu.cn

Yacong Yang
College of Computer and Data Science, Fuzhou University, China
102101304@fzu.edu.cn

Guofang Ma
School of Computer Science, Zhejiang Gongshang University, China
maguofang@zjgsu.edu.cn

Yanchao Tan*
College of Computer and Data Science, Fuzhou University, China
yctan@fzu.edu.cn

Carl Yang
Department of Computer Science, Emory University, USA
j.carlyang@emory.edu



## ABSTRACT

Diagnosis prediction is becoming crucial to develop healthcare plans for patients based on Electronic Health Records (EHRs). Existing works usually enhance diagnosis prediction via learning accurate disease representation, where many of them try to capture inclusive relations based on the hierarchical structures of existing disease ontologies such as those provided by ICD-9 codes. However, they overlook exclusive relations that can reflect different and complementary perspectives of the ICD-9 structures, and thus fail to accurately represent relations among diseases and ICD-9 codes. To this end, we propose to project disease embeddings and ICD-9 code embeddings into *boxes*, where a box is an axis-aligned hyperrectangle with a geometric region and two boxes can clearly "include" or "exclude" each other. Upon box embeddings, we further obtain patient embeddings via aggregating the disease representations for diagnosis prediction. Extensive experiments on two real-world EHR datasets show significant performance gains brought by our proposed framework, yielding average improvements of 6.04% for diagnosis prediction over state-of-the-art competitors.



## CCS CONCEPTS

• **Applied computing** → **Life and medical sciences**.

## KEYWORDS

Diagnosis prediction, Box embeddings, Representation learning, ICD-9 hierarchy

**ACM Reference Format:**
Hang Lv, Zehai Chen, Yacong Yang, Guofang Ma, Yanchao Tan, and Carl Yang. 2024. BoxCare: A Box Embedding Model for Disease Representation and Diagnosis Prediction in Healthcare Data. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3589335.3651448


*Yanchao Tan is the corresponding author.


Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.


*WWW '24 Companion, May 13–17, 2024, Singapore, Singapore*


© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.


ACM ISBN 979-8-4007-0172-6/24/05…$15.00
https://doi.org/10.1145/3589335.3651448

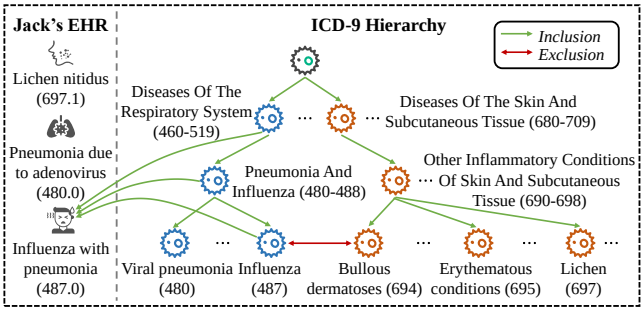




**Figure 1: An illustrative example of inclusive and exclusive relations among diseases and ICD-9 codes.**

## 1 INTRODUCTION

Diagnosis prediction, leveraging Electronic Health Records (EHRs) as valuable data sources, is becoming crucial to develop healthcare plans for patients [13, 16]. Nowadays, existing works usually enhance diagnosis prediction via learning accurate disease representation. A common approach among these works involves leveraging the hierarchical relations inherent in the Ninth Revision of International Classification of Diseases (ICD-9) codes[1] [4, 16].

However, these approaches primarily focus on inclusive relations and tend to overlook the equally important exclusive relations between ICD-9 codes, which offer distinct and complementary insights into the structure of the ICD-9 structures. This neglect can lead to an inaccurate representation of disease and ICD-9 code relations. For instance, as depicted in Figure 1, if Influenza with pneumonia (487.0) is categorized under ICD-9 code 487 (demonstrating an **inclusion**), it logically follows that it falls within the broader category of ICD-9 code 480-488 (further **inclusion**), and conversely, it should not be associated with ICD-9 code 694 (**exclusion**). Therefore, it raises a crucial question: *How to model both inclusive and exclusive relations and enhance the accuracy of diagnosis prediction?*

In this work, we propose to leverage box embeddings for disease representation and diagnosis prediction in Healthcare Data (BoxCare). Specifically, we first propose to project disease embeddings and ICD-9 code embeddings into *boxes*, where a box is an axis-aligned hyperrectangle with a geometric region. Two boxes

[1] https://www.cdc.gov/nchs/icd/icd9cm.htm

can clearly "include" or "exclude" each other. In this way, we can utilize boxes to capture inclusions and exclusions among diseases and ICD-9 codes. Upon box embeddings, we further obtain patient embeddings via aggregating the disease representations for diagnosis prediction. Extensive experimental results demonstrate that BoxCare outperforms the state-of-the-art competitors, which constantly achieves an average of 6.04% improvement.

## 2 PRELIMINARY

**Box Definition.** A box (hyperrectangle) can be described by two vectors (points). Following [11], we use the minimum point $\boldsymbol{s} \in \mathbb{R}^m$ and the maximum point $\boldsymbol{t} \in \mathbb{R}^m$ to represent a box $\boldsymbol{b}$, which is a $m$-dimensional hyperrectangle. As shown in Figure 2(a), the vector $\boldsymbol{o} \in \mathbb{R}^m$ is an offset that is positive at all coordinates, we can derive the maximum corner point as $\boldsymbol{t} = \boldsymbol{s} + \boldsymbol{o}$.

**Box Volume.** The formula to calculate the volume of a box is given by $Vol(\boldsymbol{b}) = \prod_{k=1}^{m} (\boldsymbol{t}^k - \boldsymbol{s}^k)$, where $k$ is the indicator of dimension. If the entire box space's volume is 1, then the volume of a box can be modeled as its marginal probability [5, 11]. The intersection volume between two boxes $\boldsymbol{b}_i$ and $\boldsymbol{b}_j$ is denoted as $Vol(\boldsymbol{b}_i \bigcap \boldsymbol{b}_j) = \prod_{k=1}^{m} \max(z^k, 0)$, where $z^k = \min(\boldsymbol{t}^k_{\boldsymbol{b}_i}, \boldsymbol{t}^k_{\boldsymbol{b}_j}) - \max(\boldsymbol{s}^k_{\boldsymbol{b}_i}, \boldsymbol{s}^k_{\boldsymbol{b}_j})$. Note that, the $\bigcap$ operator enables the calculation of conditional probability between two boxes, i.e., $p(\boldsymbol{b}_i|\boldsymbol{b}_j) = Vol(\boldsymbol{b}_i \bigcap \boldsymbol{b}_j)/Vol(\boldsymbol{b}_j)$.

## 3 THE BOXCARE FRAMEWORK

### 3.1 Problem Statement

To explicitly model disease representation for diagnosis prediction, we first initialize disease embeddings as $\boldsymbol{D} = \{\boldsymbol{d}_1, \boldsymbol{d}_2, ..., \boldsymbol{d}_N\}$, and ICD-9 embeddings as $\boldsymbol{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, ..., \boldsymbol{c}_M\}$, where $N$ is the number of diseases and $M$ is the number of ICD-9 codes. Then, we propose to project disease embeddings $\boldsymbol{D}$ and ICD-9 code embeddings $\boldsymbol{C}$ into boxes, where a box is an axis-aligned hyperrectangle with a geometric region. We can effectively capture the inclusion and exclusion among diseases and ICD-9 codes based on EHRs and ICD-9 hierarchy via the insideness and disjointness between boxes. As shown in Figure 2, we transform the above two relations into geometric ones between their corresponding boxes, where the inclusive relation is modeled by the geometric insideness via $\mathcal{L}_{In-d}$ and $\mathcal{L}_{In-c}$, and the exclusive relation is modeled by the geometric disjointness via $\mathcal{L}_{Ex}$. Furthermore, we aggregate the learned disease embeddings into visit embeddings $\{\boldsymbol{v}_{i,k}\}_{k=0}^{T_i-1}$, to form the comprehensive patient embedding $\boldsymbol{p}_{i,T_i}$ for patient $p_i$. $T_i$ denotes the number of visits of patient $p_i$. Finally, we perform diagnosis prediction $\hat{\boldsymbol{v}}_{i,T_i}$ via the objective function $\mathcal{L}_{Pred}$.

### 3.2 Box Embedding for Disease Representation

In this part, we propose to capture inclusive and exclusive relations among diseases and ICD-9 codes and model disease representations by leveraging box embeddings from both the geometric view and the probabilistic view. Specifically, we first project disease embeddings and ICD-9 code embeddings into boxes. Then, we utilize the geometric insideness and disjointness between two boxes to model inclusion and exclusion, respectively. Finally, we optimize these two relations via the conditional probability between boxes. Figure 2 shows how inclusion and exclusion are transformed into geometric and probabilistic constraints and we describe each relation in detail.

*3.2.1* ***Inclusive Relation.*** Since ICD-9 codes can provide abstract concepts for multiple diseases (e.g., ICD-9 code 480-488 includes disease Pneumonia due to adenovirus (480.0)) and a parent ICD-9 code can include its children geometrically, we propose to leverage the geometric insideness between the corresponding box $\boldsymbol{b}(\boldsymbol{s}, \boldsymbol{t})$ for inclusive relations. As shown in Figure 2(b), we can illustrate the inclusion between boxes from both the geometric view and the probabilistic view as follows:

LEMMA 3.1. *(Inclusive property). A box $\boldsymbol{b}_{c_i}$ contains a $\boldsymbol{b}_{d_i}$ (or $\boldsymbol{b}_{c_j}$) if and only if $p(\boldsymbol{b}_{c_i}|\boldsymbol{b}_{d_i}) = Vol(\boldsymbol{b}_{c_i}, \boldsymbol{b}_{d_i})/Vol(\boldsymbol{b}_{d_i}) = 1$ (or $p(\boldsymbol{b}_{c_i}|\boldsymbol{b}_{c_j}) = Vol(\boldsymbol{b}_{c_i}, \boldsymbol{b}_{c_j})/Vol(\boldsymbol{b}_{c_j}) = 1$).*

In particular, diseases are modeled as boxes with small constant offsets $\boldsymbol{o}_d$ (i.e., $\boldsymbol{b}_{d_i} = (\boldsymbol{d}_i, \boldsymbol{d}_i + \boldsymbol{o}_d)$), and the offsets of ICD-9 code boxes are learnable (i.e., $\boldsymbol{b}_{c_i} = (\boldsymbol{c}_i, \boldsymbol{c}_i + \boldsymbol{o}_{c_i})$). Then, we define the inclusion between the disease and ICD-9 code objective function $\mathcal{L}_{In-d}$ by measuring the geometric inclusion (i.e., $p(\boldsymbol{b}_{c_j}|\boldsymbol{b}_{d_i}) = Vol(\boldsymbol{b}_{c_j} \bigcap \boldsymbol{b}_{d_i})/Vol(\boldsymbol{b}_{d_i}) = 1$ ):

$$\begin{aligned}\mathcal{L}_{In-d}(\boldsymbol{b}_{d_i}, \boldsymbol{b}_{c_j}) = &- \boldsymbol{b}_{c_j} \log(p(\boldsymbol{b}_{c_j}|\boldsymbol{b}_{d_i})) \\ &- (1 - \boldsymbol{b}_{c_j}) \log(1 - p(\boldsymbol{b}_{c_j}|\boldsymbol{b}_{d_i})),\end{aligned} \tag{1}$$

where $\boldsymbol{d}_i \in \boldsymbol{D}$ denotes the disease embedding and $\boldsymbol{c}_j \in \boldsymbol{C}$ denotes the ICD-9 code embedding.

Similarly, we transform the inclusive relation between ICD-9 codes into the above geometric constraint and propose the other inclusion loss $\mathcal{L}_{In-c}$ as follows:

$$\mathcal{L}_{In-c}(\boldsymbol{b}_{c_j}, \boldsymbol{b}_{c_i}) = 1 - \frac{Vol(\boldsymbol{b}_{c_j} \bigcap \boldsymbol{b}_{c_i})}{Vol(\boldsymbol{b}_{c_j})}, \tag{2}$$

where box $\boldsymbol{b}_{c_i}(\boldsymbol{s}_{c_i}, \boldsymbol{t}_{c_i})$ contains box $\boldsymbol{b}_{c_j}(\boldsymbol{s}_{c_j}, \boldsymbol{t}_{c_j})$.

*3.2.2* ***Exclusive Relation.*** To properly model the exclusion between ICD-9 codes, we interpret the exclusion as geometric disjointness between boxes $\boldsymbol{b}(\boldsymbol{s}, \boldsymbol{t})$.

LEMMA 3.2. *(Exclusive property). A box $\boldsymbol{b}_{c_j}(\boldsymbol{s}_{c_j}, \boldsymbol{t}_{c_j})$ disconnects from a box $\boldsymbol{b}_{c_k}(\boldsymbol{s}_{c_k}, \boldsymbol{t}_{c_k})$ if and only if $Vol(\boldsymbol{b}_{c_j}, \boldsymbol{b}_{c_k})/Vol(\boldsymbol{b}_{c_k}) = 0$.*

Then, we propose an exclusion loss $\mathcal{L}_{Ex}$ for exclusive relation modeling as follows:

$$\mathcal{L}_{Ex}(\boldsymbol{b}_{c_k}, \boldsymbol{b}_{c_j}) = \frac{Vol(\boldsymbol{b}_{c_k} \bigcap \boldsymbol{b}_{c_j})}{Vol(\boldsymbol{b}_{c_k}) \times Vol(\boldsymbol{b}_{c_j})}, \tag{3}$$

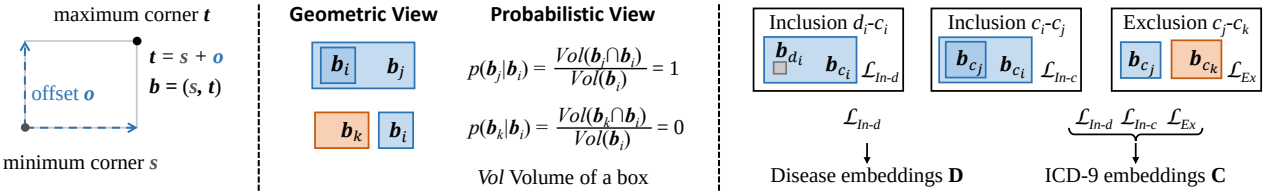
where box $\boldsymbol{b}_{c_j}(\boldsymbol{s}_{c_j}, \boldsymbol{t}_{c_j})$ disjoints from box $\boldsymbol{b}_{c_k}(\boldsymbol{s}_{c_k}, \boldsymbol{t}_{c_k})$.

### 3.3 Diagnosis Prediction

Based on the learned disease representations via box embeddings, we first adopt the self-attention mechanism [17] to obtain the historical visit embeddings of patient $p_i$:

$$\begin{aligned}&\{\boldsymbol{v}_{i,k}\}_{k=0}^{T_i-1} = \{\text{Self-Att}(\boldsymbol{D}_{i,k})\}_{k=0}^{T_i-1}, \\ &\text{Self-Att}(\boldsymbol{D}_{i,k}) = \text{softmax}((\boldsymbol{D}_{i,k}\boldsymbol{W}^Q)(\boldsymbol{D}_{i,k}\boldsymbol{W}^K)^T/\sqrt{dim})\boldsymbol{D}_{i,k}\boldsymbol{W}^V,\end{aligned} \tag{4}$$

where the matrices $\boldsymbol{W}^Q$, $\boldsymbol{W}^K$, and $\boldsymbol{W}^V$ are learnable. $\boldsymbol{D}_{i,k}$ is the set of disease embeddings diagnosed in the $k$-th visit of patient $p_i$. Then,




**(a) Box Definition** **(b) Box Relations from a Joint View** **(c) Box Embedding for Disease Representation**

**Figure 2: The overall framework of BoxCare.**

**Table 1: Statistics of the datasets used in our experiments.**

| Dataset | MIMIC-III | eICU |
|---|---|---|
| # of patients | 2,371 | 23,828 |
| # of visits | 7,279 | 59,908 |
| Avg. visits per patient | 3.07 | 2.51 |
| # of unique ICD-9 codes | 4,880 | 2,591 |
| Avg. # of diagnosis codes per visit | 13.39 | 4.22 |
| Max # of diagnosis codes per visit | 39.0 | 95.0 |

we generate patient embedding $\boldsymbol{p}_{i,T_i}$ via attention mechanism [8]:

$$\begin{aligned} \{w_{i,k}\}_{k=0}^{T_i-1} &= \text{softmax}(\text{MLP}(\{\boldsymbol{v}_{i,k}\}_{k=0}^{T_i-1})), \\ \boldsymbol{p}_{i,T_i} &= \sum_{k=0}^{T_i-1} w_{i,k}\boldsymbol{v}_{i,k}, \end{aligned} \tag{5}$$

where $\{\boldsymbol{v}_{i,k}\}_{k=0}^{T_i-1}$ denotes the historical visits of patient $p_i$, and $\{w_{i,k}\}_{k=0}^{T_i-1}$ denotes the weight assigned to each visit.

Since the diagnosis prediction is a multi-label classification task, we use a dense layer with a softmax function to calculate the predicted probability as follows:

$$\hat{\boldsymbol{v}}_{i,T_i} = \text{softmax}(\text{MLP}(\boldsymbol{p}_{i,T_i})), \tag{6}$$

where $\hat{\boldsymbol{v}}_{i,T_i}$ is the prediction of $p_i$'s $T_i$-th diagnosis. The objective function $\mathcal{L}_{Pred}$ for diagnosis prediction is listed:

$$\mathcal{L}_{Pred} = \sum_{i=1}^{K} -(\boldsymbol{v}_{i,T_i} \log(\hat{\boldsymbol{v}}_{i,T_i})) - (1 - \boldsymbol{v}_{i,T_i}) \log(1 - \hat{\boldsymbol{v}}_{i,T_i}), \tag{7}$$

where $\boldsymbol{v}_{i,T_i}$ is the ground-truth of $p_i$'s $T_i$-th diagnosis and $K$ is the number of patients.

### 3.4 Overall Objectives

Finally, the overall objective function of BoxCare is:

$$\mathcal{L} = \mathcal{L}_{Pred} + \lambda(\mathcal{L}_{In-d} + \mathcal{L}_{In-c} + \mathcal{L}_{Ex}), \tag{8}$$

where $\lambda$ is a weight hyperparameter to control the regularization for box relation modeling.

## 4 EXPERIMENT

### 4.1 Experimental Settings

**Datasets and Evaluation Protocols.** To verify the effectiveness of compared methods, we use two real-world EHR datasets MIMIC-III [6] and eICU [12] that have been fully anonymized and carefully sanitized before our access. For both datasets, we chose the patients who had at least two visits and then summarized the statistics in Table 1. ICD-9 [7, 16] is the official code for the disease. We evaluate the prediction performance using two metrics: Recall@$k$ and NDCG@$k$ that are consistent with [7].

**Methods for Comparison.** We compare our proposed BoxCare with the following baselines from two perspectives: (1) interaction modeling methods: GRAM [3], KAME [10], MHM [13], TAdaNet [15], CGL [7], and EHR2HG [14]; (2) dynamic modeling methods: RETAIN [2], Dipole [9], Timeline [1], and HiTANet [8].

**Implementation Details.** We split the dataset randomly according to patients into training/validation/test sets (i.e., 1660/237/474 on MIMIC-III and 16679/2383/4766 on eICU), which is consistent with [7]. We employ the standard Adam optimizer to optimize the compared baselines. For all compared methods on MIMIC-III and eICU, we set the embedding dimension *dim* as 128 and the batch size as 128. In particular, we set the dimension of box embeddings in BoxCare as 50, and observe that the model already outperforms other methods. The hyperparameters of all baselines are carefully tuned as suggested in the original papers to achieve their best performance. After tuning all hyperparameters on training sets, we set the weight hyperparameter $\lambda$ to 3.0.

### 4.2 Overall Performance Comparison

We compare the prediction performance of the proposed BoxCare framework with those of the baseline models. Table 2 shows the Recall@$k$ and NDCG@$k$ scores on MIMIC-III and eICU datasets with $k$={5, 10}. We have the following observations.

BoxCare consistently outperforms all baselines across all metrics on both datasets. It demonstrates the capability of our proposed BoxCare to predict diagnoses by effectively and efficiently capturing the inclusive and exclusive relations among diseases and ICD-9 codes from ICD-9 hierarchy. Compared with the second-best performance, the performance gains of BoxCare range from 2.47% with NDCG@5 on eICU to 8.98% achieved with Recall@5 on MIMIC-III.

Specifically, BoxCare outperforms CGL and EHR2HG, which use transition functions on disease graphs to predict health events. It demonstrates the effectiveness of fully modeling the structural information of ICD-9 hierarchy. Although HiTANet can leverage time information for risk prediction, it does not integrate the significant relations among the diseases and ICD-9 codes. It further affirms the significance of leveraging abundant relations extracted from ICD-9 hierarchy on EHRs.

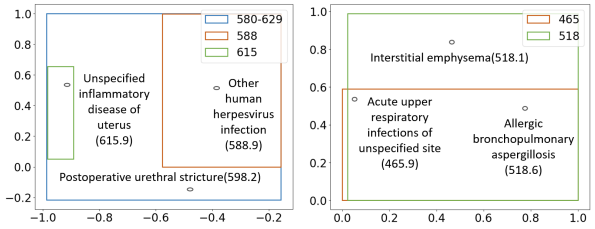
### 4.3 Case Studies

In this section, we provide several examples on MIMIC-III, and visualize the corresponding box embeddings learned by BoxCare.

**Modeling Inclusive and Exclusive Relations via Boxes**. As shown in Figure 3(a), Unspecified Inflammatory Disease of Uterus (615.9) and Other Human Herpesvirus Infection (588.9) belong to the ICD-9 code 615 and 588, respectively, indicating inclusive relations between diseases and ICD-9 codes. Furthermore, since ICD-9 code

**Table 2: Experimental results on two benchmark EHR datasets with Recall and NDCG. The best performances are highlighted in boldface and the second runners are underlined.**

| Method | Recall@5 | NDCG@5 | Recall@10 | NDCG@10 | Recall@5 | NDCG@5 | Recall@10 | NDCG@10 |
|---|---|---|---|---|---|---|---|---|
| | MIMIC-III | | | | eICU | | | |
| RETAIN | 0.1510 | 0.4188 | 0.2134 | 0.3537 | 0.3213 | 0.3428 | 0.3901 | 0.3605 |
| Dipole | 0.1442 | 0.3999 | 0.2038 | 0.3378 | 0.3071 | 0.3274 | 0.3727 | 0.3452 |
| GRAM | 0.1429 | 0.4059 | 0.2112 | 0.3510 | 0.3049 | 0.3318 | 0.3862 | 0.3576 |
| Timeline | 0.1487 | 0.4123 | 0.2100 | 0.3482 | 0.3175 | 0.3376 | 0.3840 | 0.3548 |
| KAME | 0.1353 | 0.3992 | 0.2055 | 0.3070 | 0.2887 | 0.3268 | 0.3759 | 0.3126 |
| MHM | 0.1383 | 0.4080 | 0.2128 | 0.3481 | 0.2954 | 0.3340 | 0.3893 | 0.3547 |
| TAdaNet | 0.1433 | 0.4114 | 0.2172 | 0.3568 | 0.3056 | 0.3371 | 0.3972 | 0.3642 |
| HiTANet | 0.1502 | 0.4166 | 0.2122 | 0.3518 | 0.3204 | 0.3413 | 0.3881 | 0.3584 |
| CGL | 0.1538 | 0.4265 | 0.2173 | 0.3602 | 0.3379 | 0.3624 | 0.4298 | 0.3872 |
| EHR2HG | 0.1560 | 0.4284 | 0.2206 | 0.3619 | 0.3442 | 0.3632 | 0.4383 | 0.3893 |
| BoxCare | **0.1714** | **0.4418** | **0.2385** | **0.3791** | **0.3754** | **0.3724** | **0.4710** | **0.4025** |




**(a) Box Relations** **(b) Syndromic Diseases**

**Figure 3: Visualizations of disease and ICD-9 embeddings.**

615 and 588 are exclusive, the two diseases are well separated from each other. Meanwhile, they share the same parent 580-629, which shows that the ICD-9 code 580-629 is a more abstract concept. It clearly illustrates that our proposed BoxCare accurately models the inclusive and exclusive relations via box embeddings, which is consistent with the ICD-9 hierarchy.

**Modeling Complicated Relations among Syndromic Diseases.** We observe that BoxCare can effectively capture the complicated relations among syndromic diseases. As shown in Figure 3(b), although ICD-9 code 465 and 518 are exclusive, there is a partial overlap between their corresponding box embeddings. This is mainly because Allergic Bronchopulmonary Aspergillosis (518.6) and Acute Upper Respiratory Infections of Unspecified Site (465.9) are syndromic diseases, and patients often concurrently suffer from them on EHRs, which leads to the partial overlap between the boxes of ICD-9 codes. Figure 3(b) intuitively reflects that BoxCare can not only model inclusion and exclusion from the ICD-9 hierarchy, but also capture the complicated relations among syndromic diseases from patient-disease interactions on EHRs.

## 5 CONCLUSION

In this paper, we propose to model disease representation and achieve diagnosis prediction based on the EHR data. Specifically, we propose a novel diagnosis model (BoxCare), which effectively captures inclusive and exclusive relations among diseases and ICD-9 codes via box embeddings. Extensive quantitative experiments demonstrate the clear advantages of our BoxCare over state-of-the-art baselines in diagnosis prediction, and insightful case studies show the accuracy and interpretability of our box relation modeling.

## 6 ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grants (No.6230071268), the Natural Science Foundation of Zhejiang Province (LQ23F020007), Zhejiang Gongshang University "Digital+" Disciplinary Construction Management Project (SZJ2022B001). Carl Yang was not supported by any fund from China.

## REFERENCES

[1] T. Bai, S. Zhang, B. L. Egleston, and S. Vucetic. Interpretable representation learning for healthcare via capturing disease progression through time. In *SIGKDD*, pages 43–51, 2018.
[2] E. Choi, M. T. Bahadori, J. A. Kulas, A. Schuetz, W. F. Stewart, and J. Sun. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In *NeurIPS*, 2016.
[3] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun. Gram: graph-based attention model for healthcare representation learning. In *SIGKDD*, 2017.
[4] E. R. Hansen, T. Sagi, and K. Hose. Diagnosis prediction over patient data using hierarchical medical taxonomies. *Workshop Proceedings of the EDBT/ICDT*, 2023.
[5] S. Jiang, Q. Yao, Q. Wang, and Y. Sun. A single vector is not enough: Taxonomy expansion via box embeddings. In *WWW*, pages 2467–2476, 2023.
[6] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-Wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 2016.
[7] C. Lu, C. K. Reddy, P. Chakraborty, S. Kleinberg, and Y. Ning. Collaborative graph learning with auxiliary text for temporal event prediction in healthcare. In *IJCAI*, 2021.
[8] J. Luo, M. Ye, C. Xiao, and F. Ma. Hitanet: Hierarchical time-aware attention networks for risk prediction on electronic health records. In *SIGKDD*, pages 647–656, 2020.
[9] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In *SIGKDD*, pages 1903–1911, 2017.
[10] F. Ma, Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao. Kame: Knowledge-based attention model for diagnosis prediction in healthcare. In *CIKM*, 2018.
[11] Y. Onoe, M. Boratko, A. McCallum, and G. Durrett. Modeling fine-grained entity types with box embeddings. In *ACL*, 2021.
[12] T. J. Pollard, A. E. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, and O. Badawi. The eicu collaborative research database, a freely available multi-center database for critical care research. *Scientific data*, 2018.
[13] Z. Qiao, Z. Zhang, X. Wu, S. Ge, and W. Fan. Mhm: Multi-modal clinical data based hierarchical multi-label diagnosis prediction. In *SIGIR*, 2020.
[14] Z. Sun, X. Yang, Z. Feng, T. Xu, X. Fan, and J. Tian. Ehr2hg: Modeling of ehrs data based on hypergraphs for disease prediction. In *BIBM*, 2022.
[15] Q. Suo, J. Chou, W. Zhong, and A. Zhang. Tadanet: Task-adaptive network for graph-enriched meta-learning. In *SIGKDD*, pages 1789–1799, 2020.
[16] Y. Tan, C. J. Yang, X. Wei, C. Chen, W. Liu, L. Li, J. Zhou, and X. Zheng. Metacare++: Meta-learning with hierarchical subtyping for cold-start diagnosis prediction in healthcare data. In *SIGIR*, pages 449–459, 2022.
[17] M. Usama, B. Ahmad, W. Xiao, M. S. Hossain, and G. Muhammad. Self-attention based recurrent convolutional neural network for disease prediction using healthcare data. *Comput Methods Programs Biomed*, 190:105191, 2020.