Predicting future disorders via temporal knowledge graphs and medical ontologies

Marco Postiglione, Daniel Bean, Zeljko Kraljevic, Richard JB Dobson, Vincenzo Moscato

***Abstract*—Despite the vast potential for insights and value present in Electronic Health Records (EHRs), it is challenging to fully leverage all the available information, particularly that contained in the free-text data written by clinicians describing the health status of patients. The utilization of Named Entity Recognition and Linking tools allows not only for the structuring of information contained within free-text data, but also for the integration with medical ontologies, which may prove highly beneficial for the analysis of patient medical histories with the aim of forecasting future medical outcomes, such as the diagnosis of a new disorder. In this paper, we propose MedTKG, a Temporal Knowledge Graph (TKG) framework that incorporates both the dynamic information of patient clinical histories and the static information of medical ontologies. The TKG is used to model a medical history as a series of snapshots at different points in time, effectively capturing the dynamic nature of the patient's health status, while a static graph is used to model the hierarchies of concepts extracted from domain ontologies. The proposed method aims to predict future disorders by identifying missing objects in the quadruple $\langle s, r, ?, t\rangle$, where s and r denote the patient and the *disorder* relation type, respectively, and t is the timestamp of the query. The method is evaluated on clinical notes extracted from MIMIC-III and demonstrates the effectiveness of the TKG framework in predicting future disorders and of medical ontologies in improving its performance.**

***Index Terms*—Temporal knowledge graph, evolutional representation learning, graph convolution network, electronic health records**

I. Introduction

ELECTRONIC Health Records (EHRs) are a vital tool for healthcare providers in today's digital age, as they provide a comprehensive record of a patient's health history, including demographics, medications, lab results, and treatment plans. Not only does this allow for improved continuity of care and better coordination among healthcare providers, but it also enables healthcare providers to identify trends and make data-driven decisions to improve patient care.

Many works have been proposed to analyze the structured information in EHRs to predict the risk for medical problems [1]–[3]. However, the vast majority of the data stored in EHRs is unstructured, posing a significant challenge for extracting relevant information and utilizing it effectively. To overcome this challenge, Natural Language Processing (NLP) techniques have been shown to be capable of extracting relevant information from unstructured data and linking it to medical ontologies [4], [5].

Knowledge Graphs (KGs) — which have recently revealed promising results in the fields of recommendation systems [6], information retrieval [7] and natural language processing [8] — offer the possibility to integrate a diverse array of patient data, spanning various types and originating from disparate sources. A traditional *static* KG is a structured representation of knowledge that utilizes a graph-based data topology to integrate factual information in the form of triples, $\langle s, r, o\rangle$, where s and o represent the subject and object entities, respectively, and r denotes the relation between them. Medical ontologies, such as SNOMED CT[1] and UMLS [9], are commonly structured as hierarchies of concepts, which allows for their representation in the form of KGs.

However, traditional KGs, while effective in representing structured relationships through triples of entities and relations, are inherently *static*. They can depict factual connections between entities, but they do not allow for the representation of temporal dependencies. This static nature becomes a significant limitation in healthcare contexts, where a patient's health status is not a fixed entity but a dynamic continuum, subject to change due to various factors such as treatment responses, disease progression, and lifestyle adjustments.

This highlights the need for alternative representations, such as Temporal Knowledge Graphs (TKGs) [10], which address this limitation by introducing a time dimension to the traditional KG framework. Specifically, TKGs can effectively capture the dynamic nature of the patient's health status by extending facts from a triple $\langle s, r, o\rangle$ to a quadruple $\langle s, r, o, t\rangle$, where a timestamp t is appended [11] [12]. Consequently, a medical history can be modeled as a TKG which consists in multiple snapshots that capture the health status of the patient at different points in time.

However, the integration of the dynamic information from medical histories and the static information from biomedical ontologies has not been explored yet. In our work, we propose a learning framework, named MedTKG, that aims to predict future disorders associated with a patient, i.e. the missing objects in the quadruple $\langle s, r, ?, t\rangle$, where s and r denote the patient and the *disorder* relation type, while t is the timestamp of the query. Figure 1 shows an example of a medical history and the related disorder diagnosis task. Each timestamp, denoted by t_i, represents a snapshot of the patient's health status at a particular point in time. These timestamps

M. Postiglione and V. Moscato are with the Department of Electrical and Information Technology of University of Naples Federico II (via Claudio 21, 80125, Naples, Italy). Email: {name.surname}@unina.it

V. Moscato is also with CINI - ITEM National Lab, Complesso Universitario Monte Sant'Angelo, 80145 Naples, Italy

D. Bean, Z. Kraljevic and R. Dobson are with the Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, U.K. London. D. Bean is currently employed at AstraZeneca. Email: {name.surname}@kcl.ac.uk

[1]https://www.snomed.org

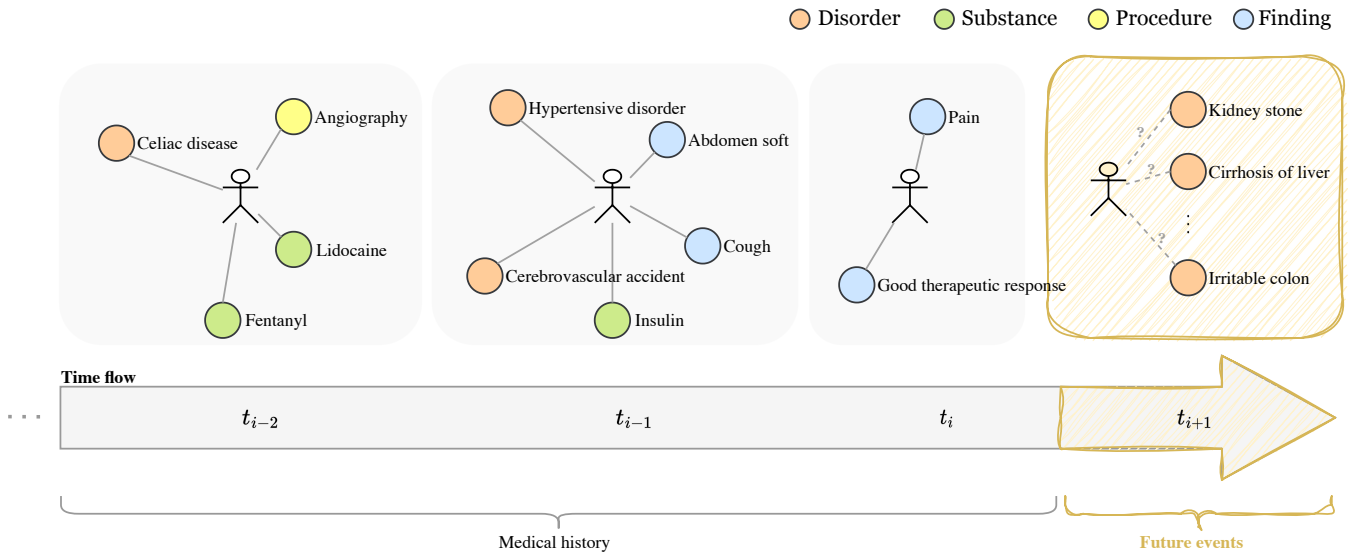


Fig. 1: Predicting future disorders with Temporal Knowledge Graphs. Letting t_i and t_{i-n} indicate the current and initial timestamps, respectively, with n being the length of the medical history seen so far, MedTKG aims to predict links to future disorders at a future timestamp t_{i+1}. Data in each timestamp is represented in the form of a knowledge graph, in which the patient is depicted as a central node connected to various medical concepts, such as disorders, substances, procedures, and findings.

capture all the events that took place during a single day of the patient's hospital stay. The patient is connected to all the concepts extracted from clinical notes that were recorded at that timestamp. To ensure that the model does not predict repetitive or periodic events that have already occurred in the patient's timeline, we only store new facts that have not been previously recorded.

Differently from the current literature on TKGs, our approach involves analyzing multiple TKGs, one for each individual patient within the dataset, and a query $\langle s, r, ?, t \rangle$ does not have a unique object as the correct answer, but rather a list of all possible disorders that may occur in the patient's future. This necessitates modifications to the training methodology, as well as a revision of the evaluation metrics. Specifically, we decided to leverage metrics commonly employed in the field of recommender systems.

The results of this study reveal that effectively managing the dynamic aspects of patients' health status enhances performance. Furthermore, integrating medical ontologies into the prediction model consistently yields better results. These findings provide valuable insights for future research in the field of healthcare and have the potential to enhance the decision-making process for clinicians, ultimately leading to improved patient outcomes.

II. Related Work

A. Deep learning for forecasting disorders

The majority of previous work for prediction or forecasting uses structured datasets or structured data in EHRs for the prediction of a limited number of possible future events. For example, Miotto et al. [13] employ a limited vocabulary comprising only 72 diseases. Similarly, Choi et al. [14] and Ma et al. [15] focus specifically on disorder categories, thereby constraining their predictions to a narrower range of about 200 ICD-9 codes.

A significant body of research in the field utilizes transformer-based models to analyze electronic health records (EHRs). BEHRT [2] utilizes a subset of 301 disorders present in structured EHR data. However, this approach is limited to predicting disorders that occur during a specific, predefined time frame, as the information must be grouped by patient visits, and its multi-label approach can present challenges as the number of concepts to be predicted increases. G-BERT [3] utilizes single-visit samples from EHRs, limiting its ability to capture long-term contextual information. Similar to BEHRT, G-BERT only utilizes structured data. Med-BERT [1], is trained on structured diagnosis data coded using the International Classification of Diseases. However, the model is evaluated on a small subset of disorders, making it difficult to estimate overall performance. MedGPT [16] leverages unstructured data in clinical narratives by first performing a Named Entity Recognition and Linking (NER+L) task.

While transformer-based models have demonstrated capability in recognizing temporal patterns in data, they lack the capability to enhance their performance through integration with medical ontologies. These ontologies, which are rooted in scientific literature, have been demonstrated to be useful to aid the model in making more accurate predictions [17]. Graph data structures, when utilized for link-prediction tasks, are proven to be effective in predicting disorders [18]. Furthermore, they offer the advantage of integrating diverse data sources, including medical ontologies. In light of this, GRAM [14] exploits medical ontologies and the attention mechanism to learn robust medical code representations, KAME [15] pre-

dicts future visit information with medical ontologies, CompNet [19] learns correlative and adverse interactions between medicines by taking into consideration additional medical knowledge (e.g. drug-drug interactions).

B. Temporal Knowledge Graphs

Current research in the field of knowledge graphs predominantly addresses static knowledge graphs, wherein the represented facts are immutable over time. This approach overlooks the critical dimension of temporal dynamics in knowledge graphs, an aspect that is relatively underexplored [20]. The incorporation of temporal information is vital, especially in contexts such as patient medical histories. In such scenarios, the applicability and accuracy of structured knowledge are temporally constrained, necessitating an understanding that the evolution of medical facts (e.g., disorders, laboratory test results, and medical procedures) adheres to a chronological sequence.

Link prediction TKG-based techniques that have been proposed in current literature can be mainly divided in two categories [21]: (1) frameworks that adapt well-performing static KG-based methods and (2) applications of time series methods from other domains. Current approaches include TTransE [22], which augments the translation-based score function used in tradizional KG embedding methods [23] with an additional time embedding, HyTE [24], which projects entities and predicates into a specific time hyperplane, ToKEi [25], which deals with knowledge having varying time granularities, and supports multiple time points and non-contiguous validity intervals, DE-SimplE [26], which employs diachronic entity embeddings to represent entities at different timestamps, ATiSE [27] which learns time-aware embeddings of entities and predicates as a Gaussian distribution to represent time uncertainty, TeRo [28] which extends HyTE by learning time-sensitive entity and predicate embeddings via rotation operations specific to various timestamps, and TComplEx [29] which upgrades ComplEx by scoring each event via a fourth-order tensor decomposition that incorporates time information. Several models have been proposed that incorporate Graph Neural Networks (GNNs) or Recurrent Neural Networks (RNNs) to identify spatial-temporal patterns. Examples of these models include RE-NET [30], RE-GCN [31], HIP [32], and EvoKG [33].

III. Problem formulation

Definition 1 (Medical History)**.** *A medical history $\mathcal{M}_T$ is represented with a TKG, which can be formalized as a sequence of KGs, i.e. $\mathcal{M}_T = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_T\}$, where T is the length of the sequence and each KG $\mathcal{G}_t = \langle \mathcal{V}, \mathcal{R}, \mathcal{E}_t \rangle$ at timestamp t is a directed heterogeneous graph, $\mathcal{V}$, $\mathcal{R}$ and $\mathcal{E}_t$ being the sets of entities, relations and facts at timestamp t, respectively. While entities $\mathcal{V}$ and relations $\mathcal{R}$ are shared across timestamps and TKGs (i.e. the same medical concept can be present in different medical histories), facts $\mathcal{E}_t$ depend on the patient and the current timestamp. A fact $e \in \mathcal{E}_t$ can be formalized as a quadruple $e = \langle s, r, o, t \rangle$, where $s \in \mathcal{V}$, $o \in \mathcal{V}$, $r \in \mathcal{R}$ and t is the current timestamp. In this work, a fact $e = \langle s, r, o, t \rangle$ indicates that a patient s is related to the medical concept o of type r (e.g. disorder, medical procedure, medical substance) at timestamp t.*

Definition 2 (Medical Ontology Graph)**.** *The ontology graph $\mathcal{G}^s$ is a static knowledge graph modelling the knowledge embedded in a medical ontology. It can be formalized as a graph $\mathcal{G}^s = \langle \mathcal{V}^s, \mathcal{R}^s, \mathcal{E}^s \rangle$ where $\mathcal{V}^s \subset \mathcal{V}$ is the set of concepts included in the ontology, $\mathcal{R}^s$ is the set of possible relations between concepts and $\mathcal{E}^s$ are the edges.*

Definition 3 (Future disorder prediction)**.** *Given the patient $s \in \mathcal{V}$ and its medical history $\mathcal{M}_t$, the input to our model is a query $\langle s, r, ?, t+1 \rangle$, i.e. we want to predict the next concept associated to s. Despite our framework being easily generalisable to several medical concepts, we focus only on disorders, i.e. $r = disorder$. MedTKG models the conditional probability vector of all concept entities with patient s, relation r and history $\mathcal{M}_t$, $p(o|s, r, \mathcal{M}_t)$.*

IV. Methodology

In this section, we present our methodology for addressing the problem of future disorder prediction. Our approach builds upon the work of Li et al. [31], in which a TKG-based model is trained by considering the relationships among concurrent facts, the patterns of events over time and the inherent characteristics of entities. We adapt the approach to our clinical scenario (i.e. clinical notes as the input to model medical histories and medical ontologies as the source of static information for medical concepts) and extend it to deal with several independent TKGs, each of them representing the medical history of a patient.

The architecture of the proposed MedTKG model is illustrated in Figure 2. A detailed description of each individual component will be provided in the following sections. To facilitate the reader in the understanding of our methodology, we summarize the adopted notation in Table I.

A. Inputs

1) Medical History: Starting from the free text contained in clinical notes, the first step in our architecture is Named Entity Recognition and Linking (NER+L), which consists in extracting mentions of the clinical concepts we are interested to and linking them to a medical ontology. In our experiments, we extracted mentions of disorders, procedures, substances and findings, and linked them to the SNOMED-CT ontology by using the Medical Concept Annotation Toolkit (MedCAT) [34], a set of decoupled technologies exposing state-of-the-art models trained by self-supervised learning. Note that this module is easy to be replaced according to individual needs.

After extracting all the medical concepts pertaining to a patient's medical history, we represent this knowledge in the form of a TKG (see Section 1).

2) Medical ontology graph: To include the knowledge of the medical ontology in the learning framework, we need to represent it as an Ontology graph (see Section 2). In this study, we utilize the SNOMED-CT medical ontology and leverage the links between medical concepts. Specifically, we consider

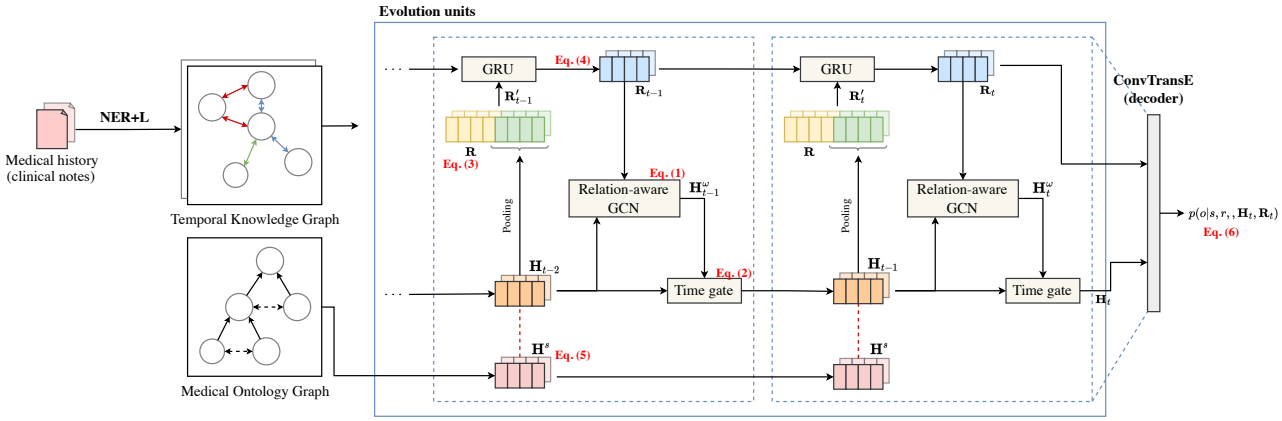


Fig. 2: A methodological flowchart illustrating the steps involved in processing a single patient's medical history.

TABLE I: Notations

Symbol	Description
$\mathcal{M}_T = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_T\}$	Medical history, represented as a TKG with size T.
$\mathcal{G}_t = \langle \mathcal{V}, \mathcal{R}, \mathcal{E}_t \rangle$	Directed heterogeneous graph, where $\mathcal{V}$, $\mathcal{R}$ and $\mathcal{E}_t$ are the sets of entities, relations and facts at timestamp t, respectively.
$e = \langle s, r, o, t \rangle$, $e \in \mathcal{E}_t$	Fact in a TKG. It indicates that a patient s is related to the medical concept o of type r (e.g. disorder, medical procedure, medical substance) at timestamp t.
$\mathcal{G}^s = \langle \mathcal{V}^s, \mathcal{R}^s, \mathcal{E}^s \rangle$	Medical Ontology (static) graph.
$\mathbf{H}_t$	Entity embedding matrix for the graph $\mathcal{G}_t$, output of the evolution unit.
$\mathbf{H}_t^\omega$	Entity embedding matrix for the graph $\mathcal{G}_t$, output of the final layer ω of the GCN.
$\mathbf{h}_{n,t}^l$	Embedding of the object (subject) n at timestamp t and GCN layer l.
$\mathbf{R}_t$	Relation embedding matrix.
$\mathbf{r}_t$	Embedding of the relation r at timestamp t.
$\mathbf{U}_t$	Time gate component.
$\mathcal{N}_{r,t}$	Entities related by r at timestamp t.
$\mathcal{L}_m^e$	Entity prediction loss on medical history m.
$\mathcal{L}_m^s$	Medical ontology constraint.
$\theta_t = \min(\gamma t, 90°)$	Threshold for the angle between static and evolutionary embeddings.
γ	Parameter that controls the "speed" of the θ_t threshold, i.e. how rapidly it increases over time.

two types of relations: (1) a *direct* relation, which is defined as an "is a" relationship between two concepts as represented in the ontology, and (2) an *indirect* relation, which is defined as a relationship between two concepts that are not directly linked in the ontology, but share a common parent, i.e. they are both related to the same medical concept, also with an "is a" relationship.

B. Evolution unit

The evolution unit, based on the work from Li et al. [31], comprises several elements that are utilized to model the temporal dynamics of the patient's health status and the static information from a medical ontology. A Relation-aware Graph Convolutional Network (GCN) is utilized to capture the structural dependencies within the knowledge graph (KG) at each timestamp. The temporal evolution of the KG is modeled through the integration of two gated recurrent components, namely, a time gated recurrent component and a Gated Recurrent Unit (GRU) component. These components enable the recurrent computation of the evolutional representations of entities and relations at each timestamp. Furthermore, to ensure the preservation of ontology-based properties within the KG, a static graph constraint component introduces constraints between the static embeddings and the evolutional embeddings of entities to integrate the static properties of the medical ontology. The aim of the evolution unit is to output an entity-embedding matrix $\mathbf{H}_i$ for each graph $\mathcal{G}_i$ in the medical history.

In the following, we will describe each module of the evolution unit in detail.

1) Structural dependencies: The structural dependencies among concurrent facts in a knowledge graph are captured in order to model the associations among the entities through the facts they participate in. Given their well-demonstrated ability to learn from multi-relational graph-structured data [35]–[37], we use a ω-layer relation-aware Graph Convolutional Network (GCN) to model structural dependencies. This approach allows for a comprehensive understanding of the relationships and dependencies within the knowledge graph, which can be used to improve performance on various knowledge-intensive tasks.

Specifically, given a KG $\mathcal{G}_t \in \mathcal{M}$ at timestamp t and the object entity $o \in \mathcal{V}$ at layer l, its embedding at the next layer $l+1$ is computed under a message-passing framework as shown as follows:

$$\mathbf{h}_{o,t}^{l+1} = \mathrm{RReLu}\Big(\frac{1}{c_0} \sum_{(s,r):\exists(s,r,o)\in\mathcal{E}_t} \mathbf{W}_1^l(\mathbf{h}_{s,t}^l + \mathbf{r}_t) + \mathbf{W}_2^l \mathbf{h}_{o,t}^l\Big), \tag{1}$$

where $\mathbf{h}_{o,t}^l$, $\mathbf{h}_{s,t}^l$ and $\mathbf{r}_t$ are the embeddings of the object o, subject s and relation r at layer l and timestamp t, respectively; $\mathbf{W}_1^l$ and $\mathbf{W}_2^l$ are the parameters for aggregating features and self-loop in the l-th layer; $\mathbf{h}_{s,t}^l + \mathbf{r}_t$ implies the translational property between s to o via the relation r; c_o denotes a normalization constant (in-degree of o).

2) Historical dynamics: Sequential patterns in the medical history are captured by stacking the ω-layer relation-aware GCN. However, to alleviate the over-smoothing problem, i.e. embeddings of different entities converging to the same values, and the vanishing gradient problem that may be caused by long medical histories — which translate in many stacked GCN

layers — a time gated recurrent component is applied as in [38]:

$$\mathbf{H}_t = \mathbf{U}_t \otimes \mathbf{H}_t^{\omega} + (1 - \mathbf{U}_t) \otimes \mathbf{H}_{t-1}, \tag{2}$$

where $\otimes$ indicates the dot-product operation, $\mathbf{H}_t^{\omega}$ is the output of the final layer of the relation-aware GCN and the time gate $\mathbf{U}_t$ performs the non-linear transformation $\mathbf{U}_t = \sigma(\mathbf{W}_4\mathbf{H}_{t-1} + \mathbf{b})$, σ and $\mathbf{W}_4$denoting the sigmoid function and the weight matrix of the time gate, respectively. By using the time gate component, the entity embedding matrix $\mathbf{H}_t$ is obtained by taking into consideration both the output of the final layer of the relation-aware GCN $\mathbf{H}_t^{\omega}$ and the embedding $\mathbf{H}_{t-1}$ from the previous timestamp.

To capture the sequential patterns of relations, a GRU component is adopted. In particular, given a relation r at timestamp t and its related entities $\mathcal{N}_{r,t} = \{i | (i, r, o, t) \text{ or } (s, r, i, t) \in \mathcal{E}_t\}$, we compute the input for the GRU at timestamp t as the concatenation of (1) the result of a mean pooling operation over the embedding matrix of entities in $\mathcal{N}_{r,t}$, i.e. pooling$(\mathbf{H}_{t-1,\mathcal{N}_{r,t}})$ and (2) the embedding $\mathbf{r} \in \mathbf{R}$ of the relation *r*:

$$\mathbf{r}'_t = [\text{pooling}(\mathbf{H}_{t-1,\mathcal{N}_{r,t}}); \mathbf{r}] \tag{3}$$

Then, the relation embedding matrix is updated via the GRU:

$$\mathbf{R}_t = \text{GRU}(\mathbf{R}_{t-1}, \mathbf{R}'_t), \tag{4}$$

where $\mathbf{R}'_t$ contains the $\mathbf{r}'_t$ values for all the relations.

3) Medical ontology static dependencies: The *static* embeddings (i.e. they do not change over time) of entities in the medical ontology are obtained through a 1-layer R-GCN [39] without self loops. The update rule is defined as follows:

$$\mathbf{h}_i^s = \text{ReLU}\Big(\frac{1}{c_i} \sum_{(r^s, j): \exists (i, r^s, j) \in \mathcal{E}^s} \mathbf{W}_{r^s} \mathbf{h}_j'^s\Big), \tag{5}$$

where $\mathbf{h}_i^s$ and $\mathbf{h}_j'^s$ denote the i-th and j-th lines of the output matrices $\mathbf{H}^s$ and $\mathbf{H}'^s$, respectively; $\mathbf{W}_{r^s}$ is the relation matrix of r^s and c_i is a normalization constant equal to the number of entities linked to i.

C. Scoring function

The scoring function aims to compute the conditional probability introduced in Section 3. In particular, given the medical history $\mathcal{M}_T$, we want the probability score of candidate triples (s, r, o), $p(o|s, r, \mathcal{M}_t) = p(o|s, r, \mathbf{H}_t, \mathbf{R}_t)$ since we are representing the medical histories with entity and relation embeddings. ConvTransE [36] is used as a decoder since it contains a one-dimensional convolution layer and a fully connected layer, which have been demonstrated to be well performing when combined with GCNs [40]. The scoring function is thus computed as follows:

$$p(o|s, r, \mathbf{H}_t, \mathbf{R}_t) = \sigma\Big(\mathbf{H}_t \text{ConvTransE}(\mathbf{s}_t, \mathbf{r}_t)\Big), \tag{6}$$

where $\sigma(\cdot)$ is the sigmoid function while $\mathbf{s}_t \in \mathbf{H}_t$ and $\mathbf{r}_t \in \mathbf{R}_t$ are the embeddings of the subject *s* and relation *r*, respectively.

D. Formulation of the loss function

The loss being optimized by our model combines two terms deriving from entity prediction task (i.e. $\mathcal{L}^e$) and the medical ontology constraint (i.e. $\mathcal{L}^s$). Medical histories related to several patients are independently processed by the model. Then, given M medical histories, the loss function is computed as follows:

$$\mathcal{L} = \sum_{m=0}^{M-1} \lambda_1 \mathcal{L}_m^e + \lambda_2 \mathcal{L}_m^s, \tag{7}$$

where λ_1 and λ_2 are parameters controlling the loss terms. We will describe the two loss terms in the following and omit the medical history identifier m for the sake of simplicity.

1) Entity prediction loss: The entity prediction task is treated as a multi-label learning problem. Let $\mathbf{y}_{t+1} \in \mathbb{R}^{|\mathcal{V}|}$ denote the ground-truth label vector whose elements are 1 when the corresponding object occurs at timestamp $t + 1$, 0 otherwise. The loss function is computed as follows:

$$\mathcal{L}^e = \sum_{t=0}^{T-1} \sum_{(s,r,o,t+1) \in \mathcal{E}_{t+1}} \sum_{i=0}^{|\mathcal{V}|-1} y_{t+1,i} \log\, p_i(o|s, r, \mathbf{H}_t, \mathbf{R}_t), \tag{8}$$

where T is the length of the medical history, $y_{t+1,i}$ is the i-th element of $\mathbf{y}_{t+1}$ and p_i is the probability score of entity i.

2) Medical ontology constraint: The medical ontology constraint is designed to manage and guide the relationship between the evolutional embedding $\mathbf{h}_{i,t}$ and the static embedding $\mathbf{h}_i^s$ of the entity i at timestamp t . Specifically, the constraint controls the angle between these two embeddings, which is a measure of their similarity. The idea is that as time progresses and more facts are added, the model allows for greater divergence between the evolutional and static representations of the entity. In practical terms, this is achieved by making the angle not to exceed a threshold which increases over time as the permissible range of evolutionary embeddings values continually expands with the occurrence of additional facts, $\theta_t = \min(\gamma t, 90°)$, where γ defines the speed at which the threshold increases over time. As γ increases, the threshold angle θ_t also increases more rapidly over time, allowing for a faster adaptation of the evolutional embeddings in response to new information. The maximum angle between the static and evolutional embeddings is set to 90°. The loss of the medical ontology constraint component at timestamp t is thus defined as follows:

$$\mathcal{L}_t^s = \sum_{i=0}^{|\mathcal{V}^s|-1} \max\Big(\cos\theta_t - \cos(\mathbf{h}_i^s, \mathbf{h}_{t,i}), 0\Big) \tag{9}$$

It is important to note that the use of the max function in this formulation introduces a point of non-differentiability when $\cos\theta_t = \cos(\mathbf{h}_i^s, \mathbf{h}_{t,i})$. To effectively train our model using gradient-based optimization methods, we address this issue by employing subgradients [41] at these non-differentiable points.

Given a medical history of length T, the medical ontology constraint loss is $\mathcal{L}^s = \sum_{t=0}^{T} \mathcal{L}_t^s$.

V. Experiments

A. EHRs dataset

The MIMIC-III dataset [42], developed by the MIT Lab for Computational Physiology, was utilized as the source of data for training and evaluating our framework. This dataset contains information pertaining to patients who were admitted to the critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012, and is publicly accessible. The corpus for our analysis was derived from the complete set of unstructured clinical notes, consisting of 2083179 documents from 46520 patients.

After extracting the concepts with MedCAT, we implemented a preprocessing step that involved removing infrequent concepts, defined as those that occurred less than 100 times in the entire dataset, in order to eliminate rare diseases and that could potentially both be detrimental for the training and identify patients. Subsequently, the remaining concepts were grouped by individual patients and organized chronologically. To enhance the quality and completeness of the medical histories, several additional steps were taken: 1) A biomedical concept was retained in the patient's medical history if it appeared at least twice, which improves the precision of our Named Entity Recognition (NER) and Linking NER+L tool, but may result in a reduction in recall; 2) Concepts that were parent concepts of concepts already present in the timeline, based on the SNOMED ontology, were removed in order to reduce noise and eliminate redundant information; 3) The medical history was divided into segments with a duration of 1 day, and duplicate concepts within a segment were removed; 4) Medical histories containing less than 10 concepts were excluded from further analysis.

Finally, medical histories have been mapped to the TKG structure defined in Section 1. Research works are typically based on a single TKG, which is usually split based on the chronological order of events: for example, events related to the first 70% of the available timestamps are used to train the model, while the remaining 30% is used for testing. In our case, however, we have access to different medical histories, which translate into different TKGs. A portion of the TKGs is used to train the model, while the other is used for testing. Specifically, we divided patients in our dataset into a training, a validation and a test set, comprising 90%, 5% and 5% of the data, respectively. Note that each patient in the test set is then split into several test samples reflecting all the possible medical history lengths available. In each sample, the ground truth is given by all the concepts appearing in future time steps.

The details of the graph data generated in our study are presented in Table II, which provides statistical information on the dataset. Additionally, Figure 3 illustrates the trend of decreasing support in the test set as the length of the medical history, measured in terms of days, increases.

B. Medical Ontology

We utilized the SNOMED CT ontology to establish a mapping between all medical concepts and their corresponding codes. Through this approach, we were able to identify and analyze the various relationships that exist among the available concepts. Both direct relationships, represented by the *is a* relationship between the source and destination concepts, and indirect relationships, represented by the shared *is a* relationship between two concepts and a common ancestor, were considered. Heatmaps in Figure 4 illustrate the number of direct and indirect relationships retrieved from SNOMED CT.

TABLE II: Statistics of the dataset.

Nodes	train	dev	test
$\lvert\mathcal{V}_{\text{patient}}\rvert$	36,803	1,947	2,027
$\lvert\mathcal{V}_{\text{disorder}}\rvert$	1,376	1,330	1,322
$\lvert\mathcal{V}_{\text{procedure}}\rvert$	34	32	34
$\lvert\mathcal{V}_{\text{finding}}\rvert$	755	689	696
$\lvert\mathcal{V}_{\text{substance}}\rvert$	472	458	449

Facts	train	dev	test
$\lvert\mathcal{E}_{\text{disorder}}\rvert$	911,418	47,647	49,259
$\lvert\mathcal{E}_{\text{procedure}}\rvert$	72,511	3,747	3,896
$\lvert\mathcal{E}_{\text{finding}}\rvert$	596,900	31,844	32,470
$\lvert\mathcal{E}_{\text{substance}}\rvert$	421,551	22,151	22,738

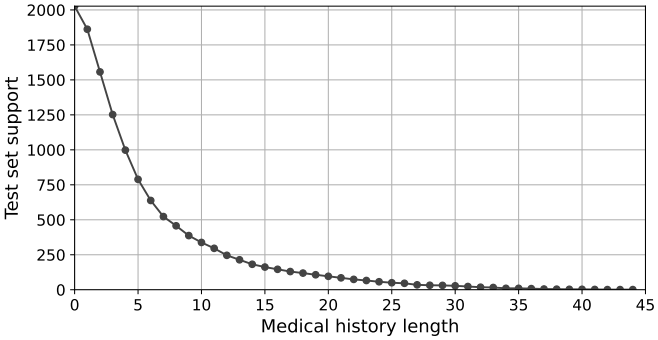
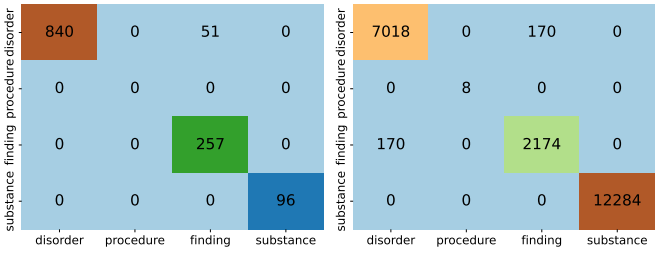


Fig. 3: Trend of test set support as the length of medical histories increases.

C. Metrics

While the literature in the field aims to test a model's ability to predict the occurrence of an event at a future timestamp, of which the actual occurrence is known, our ground truth is composed of a multiplicity of events, meaning that there is not a single disease that will be associated with the patient in the future, but a list of diseases that could potentially be associated with the patient. This required us to use a different evaluation protocol and set of metrics to test our models. To assess the effectiveness of our model, a set of metrics has been employed,



(a) Direct (is a) (b) Indirect (common parent)

Fig. 4: Relationships found in the medical ontology from source (rows) to destination (columns) concepts.

including Mean Reciprocal Rank (MRR), True Positive rate at k (TP_rate@k), Hits at k (Hits@k), Mean Recall at k (MR@k), Mean Averaged Precision at k (MAP@k), where k denotes the top-k ranked predictions made by the model. Metrics are detailed in the following.

Mean Reciprocal Rank: MRR measures the average of the reciprocal ranks of the correct results in a set of queries or predictions. The formula for MRR is:

$$MRR = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\text{rank}_i}, \tag{10}$$

where N is the number of correct concepts and rank_i denotes the ranked position of the i-th concept.

The utilization of MRR as an evaluation metric presents several advantages. Firstly, it is scale-independent, making it an appropriate metric for comparing the performance of different models or tasks, regardless of the number of items in the list, which can decrease as the medical history progresses. Additionally, MRR places a strong emphasis on the first correct result, which is particularly relevant in tasks such as information retrieval, where users are less likely to scroll through long lists of results. Furthermore, it takes into account the entire ranked list, as opposed to only the top k results, as seen in metrics such as precision and recall. Due to these advantages, we employ MRR as a method of selecting the optimal model among the results obtained at each epoch of training.

True Positive rate: To evaluate the usability of our model in real-world scenarios, where it would serve as an assistant for risk prediction or diagnosis suggestion, we investigate how likely one of the top-k predictions is correct, i.e. it appears in future steps of the medical history. Thus, we consider a prediction as a *true positive* if at least one of the top-k scored disorders is correct. The TP rate is then defined as the ratio between the sum of true positives and the total number of test samples:

$$\text{TP\_rate@}k = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}(\mathbf{y}_i, \hat{\mathbf{y}}_i^k), \tag{11}$$

where N denotes the number of test samples, $\mathbf{y}_i$ is the list of correct future disorders, $\hat{\mathbf{y}}_i^k$ denotes the first k disorders scored by the model and $\mathbb{1}(\mathbf{y}_i, \hat{\mathbf{y}}_i^k)$ equals to 1 if at least one element of $\hat{\mathbf{y}}_i^k$ is also in $\mathbf{y}_i$, 0 otherwise.

Note that the number of correct concepts decreases as we move forward with the medical history, thus implying a possible decline of performance computed with this metric.

Hits: Despite physicians having the knowledge and expertise to distinguish useful predictions among all the others, we would like every disorder predicted by the model to be correct. Hence, we compute Hits@k as the percentage of correct top-k predictions, i.e. its averaged precision:

$$\text{Hits@}k = \frac{1}{N}\sum_{i=1}^{N}\frac{\text{TP}_i^k}{k}, \tag{12}$$

where TP_i^k denotes the number of true positives for the i-th test sample obtained by considering the top-k predictions.

As with TP_rate@k, performance could decrease as we proceed forward with the patient's timeline.

Mean Recall: While TP_rate@k and Hits@k give us an idea of the *precision* of the model, i.e. of its ability to correctly identify future disorders, we are interested also in its *recall*, i.e. its ability to find all the future disorders. Mean Recall is defined as the fraction of correct items found in the top-k predictions:

$$\text{MR@}k = \frac{1}{N}\sum_{i=1}^{N}\frac{\text{TP}_i^k}{\text{TP}_i^k + \text{FN}_i^k}, \tag{13}$$

where FN_i^k denotes the number of false negatives for the i-th test sample obtained by considering the top-k predictions.

Mean Averaged Precision: Ideally, we would like our model to score relevant future disorders at the top positions. MAP takes into account both the precision and recall of the recommendations made by the model and rewards first-loaded relevant recommendations, making it more informative than all the other metrics which does not consider the order in the rankings of predictions. It is computed as follows:

$$\text{MAP@}k = \frac{1}{N}\sum_{i=1}^{N}\text{AP@}k, \tag{14}$$

where AP@k is the average of the precision at each recall level for a particular test sample:

$$\text{AP@}k = \sum_{i=1}^{k}\text{Hits@}k \cdot \text{rel}(i), \tag{15}$$

where $\text{rel}(i)$ is an indicator function which is 1 if the i-th item is relevant, 0 otherwise.

D. Training parameters

In accordance with previous research studies [31] and empirical results, the training parameters for the evolution unit were selected as follows: the embedding dimension, d, was set to 200, the number of layers, ω, in the relation-aware GCN was set to 2, and a dropout rate of 0.2 was applied to each layer of the relation-aware GCN. The time window is set to a sufficiently large value, ensuring that it includes all relevant details from each medical history. Adam optimization [43] was utilized for parameter learning with a learning rate of 0.0001. Additionally, for the RGCN used in the *medical ontology constraint* component, the block dimension was set to 2×2 and a dropout rate of 0.2 was applied to each layer. Furthermore, for the ConvTransE model, the number of kernels was set to 50, the kernel size was set to 2×3 and the dropout rate was set to 0.2. All the models have been trained for a total of 10 epochs on a NVIDIA A100 gpu and the model with the best MRR score on validation data has been used for testing.

E. Results

1) Comparison with baselines: MedTKG is compared with 10 future-disorder prediction models, including traditional approaches and healthcare-specific methods. Traditional approaches include Convolutional Neural Networks

TABLE III: Comparison with baselines. The weight of the medical ontology graph is set to $w = 0.6$. Best results are reported in bold

Method	MRR	TP rate				Hits				MR				MAP			
		@1	*@3*	*@5*	*@10*	*@1*	*@3*	*@5*	*@10*	*@1*	*@3*	*@5*	*@10*	*@1*	*@3*	*@5*	*@10*
CNN	**10.29**	31.87	51.33	59.73	69.65	31.87	27.08	24.19	20.01	**4.32**	**10.32**	**14.58**	**22.51**	**4.32**	**7.99**	**9.92**	**12.60**
RNN	8.98	26.86	46.70	55.91	67.26	26.86	23.91	22.22	19.17	3.21	8.14	12.31	20.40	3.21	6.11	7.90	10.48
RETAIN	10.20	31.73	51.08	59.89	69.89	31.73	27.14	24.24	20.08	4.10	7.76	9.70	21.38	4.10	7.76	8.62	11.26
Transformer	8.93	26.96	46.81	56.40	67.09	26.95	23.84	21.91	18.77	3.23	8.37	13.50	20.15	3.23	6.15	7.88	10.36
TCN	9.41	28.37	48.13	57.11	68.41	28.37	24.91	22.75	19.36	3.68	8.89	13.20	21.38	3.68	6.79	8.62	11.26
AdaCare	7.23	24.15	41.58	49.80	60.14	24.15	20.21	18.93	16.67	2.60	5.94	9.08	15.12	2.60	4.48	5.79	7.74
ConCare	8.87	27.15	46.10	55.71	66.84	27.15	23.66	21.86	18.95	3.3	7.91	12.05	19.93	3.3	6.04	7.73	10.26
StageNet	8.46	26.58	45.84	55.41	65.84	26.58	23.06	21.32	17.99	3.12	7.64	11.7	18.95	3.12	5.70	7.33	9.59
Dr. Agent	8.68	26.87	45.28	54.61	65.90	26.87	23.16	21.47	18.63	3.22	7.70	11.78	19.47	3.22	5.88	7.54	10.00
GRASP	9.17	27.7	48.17	57.24	67.80	27.7	25.00	23.13	19.54	3.42	8.57	12.83	20.76	3.42	6.45	8.29	10.91
MedTKG	7.15	**43.4**	**66.61**	**75.05**	**82.44**	**43.4**	**37.85**	**34.23**	**28.52**	3.7	9.09	13.3	20.13	3.7	7.09	9.08	11.85

(CNNs) [44], Recurrent Neural Networks (RNNs) [45], Transformer architectures [46] and Temporal Convolution Networks (TCNs) [47]. Healthcare-specific models include RETAIN [48], AdaCare [49], ConCare [50], StageNet [51], Dr. Agent [52], GRASP [53]. We refer to PyHealth [54] for the implementations of baseline methods.

Table III presents the results of future disorders prediction of the different baselines. It can be observed that while the CNN approach achieves the best results in terms of MRR, MR and MAP scores, MedTKG demonstrates a substantial enhancement in the TP rate and Hits performance. This superior performance in precision metrics is particularly significant in the context of clinical applications, as it indicates that MedTKG can reliably identify relevant disorders among the top predictions. As a consequence, the reliability of top recommendations is higher, thus reducing the time and effort required to sift through potential predictions, which is essential in fast-paced medical environments.

It is important to note that the performance of of next-disorder prediction systems generally exhibits superior results for the TP rate and Hits@k metrics. The TP rate, a metric specifically designed to evaluate the presence of at least one accurate prediction among the top-k ones, understandably shows higher performance. However, the disparities between the Hits and other metrics such as MR and MAP are worthy of a further discussion. While MR theoretically lies in the $[0, 1]$ range (refer to Section V.C for the definition), values are usually small in practice because the ground-truth list of correct concepts is usually longer than k, making it impossible to reach high scores. On a different note, MAP does not only evaluate the ability of the model to make accurate predictions, but also its ability to provide predictions in the correct order. While our system proves high quality in making good predictions, and overcomes the selected baselines, it is not able to reach this level of precision.

2) Medical Ontology impact: Table IV compares the performance obtained with different medical ontology weights on the next-disorder prediction task. It can be observed that the utilization of the medical ontology yields consistent improvements across the various metrics that were investigated. Interestingly, the results also suggest that a weight of 0.6 generally produces the most favorable outcomes compared to all other weight values: this is attributed to the fact that increasing the weight can lead to the medical ontology constraint impeding the network's learning capability of evolving information. Furthermore, the high true positive rate and hits values suggest a high level of precision in the model, while the low recall is likely due to the extensive number of concepts considered. A potential solution to this issue involves predicting concepts located at higher levels in the SNOMED-CT hierarchy, although this would result in lower specificity in the model's responses.

In addition to the weight of the medical ontology constraint on the training loss, the model's ability to learn from static and dynamic information is controlled by the rate at which we allow the threshold angle between static and dynamic embeddings to increase with increasing timesteps. Results in Table V show that high pace threshold values ($\geq$ 15) are associated with superior performance in the model's top predictions due to its increased ability to learn from training data rather than being constrained by the medical ontology. Conversely, lower threshold values (≤ 10) lead to better results across a wider range of top predictions by enabling the model to retrieve a larger number of correct concepts.

3) Impact on concepts predicted: Table VI shows the number of concepts ever predicted by models with varying weights assigned to the medical ontology graph. Results demonstrate that utilizing the medical ontology with the appropriate weight during training improves the model's ability to identify a broader range of concepts. Furthermore, the validation curves depicted in Figure 5 indicate that unlike the model that does not employ the medical ontology, which initially identifies a large number of concepts but later begins to overfit on a smaller set, the medical ontology facilitates gradual incorporation of additional concepts into the model's understanding.

In addition to assessing the number of concepts predicted by the model, we also evaluated its precision in predicting these concepts, taking into account the frequency of their occurrence among patients. The results, presented in Figure 6, demonstrate that the model's performance is strongly influenced by the

TABLE IV: Impact of the medical ontology weight. Best and second-best results are reported in bold and underlined, respectively. The last row (+%) reports the relative improvements obtained when using the medical ontology graph.

Weight	MRR	TP rate				Hits				MR				MAP			
		@1	*@3*	*@5*	*@10*	*@1*	*@3*	*@5*	*@10*	*@1*	*@3*	*@5*	*@10*	*@1*	*@3*	*@5*	*@10*
1.0	**7.25**	43.27	65.48	74.07	82.36	43.27	37.38	**34.27**	28.97	3.51	8.83	12.94	20.08	3.51	6.82	8.8	11.65
0.8	7.07	41.73	64.73	74.43	82.39	41.73	36.33	33.69	28.62	3.46	8.66	12.97	20.05	3.46	6.69	8.7	11.55
0.6	7.15	43.4	**66.61**	**75.05**	**82.44**	43.4	**37.85**	34.23	28.52	**3.7**	**9.09**	**13.3**	20.13	**3.7**	**7.09**	**9.08**	**11.85**
0.4	7.13	**43.54**	65.34	73.86	81.79	**43.54**	37.13	33.65	28.43	3.44	8.75	12.84	19.99	3.44	6.75	8.7	11.54
0.2	7.02	42.14	64.52	73.71	82.39	42.14	36.26	33.06	28.38	3.46	8.69	12.75	20.17	3.46	6.74	8.64	11.55
0.0	7.21	43.01	65.26	74.35	82.07	43.01	36.92	34.15	**28.98**	3.53	8.75	13.04	**20.23**	3.53	6.82	8.85	11.81
+%	↑ 0.5	↑ 1.2	↑ 2.1	↑ 0.9	↑ 0.4	↑ 1.2	↑ 2.5	↑ 0.3	↓ 0.03	↑ 4.8	↑ 3.9	↑ 2.0	↓ 0.3	↑ 4.8	↑ 4.0	↑ 2.6	↑ 0.3

TABLE V: Impact of the pace of the angle threshold between dynamic and static embeddings. The weight of the medical ontology graph is set to 1.0. Best and second-best results are reported in bold and underlined, respectively.

Angle	MRR	TP rate				Hits				MR				MAP			
		@1	*@3*	*@5*	*@10*	*@1*	*@3*	*@5*	*@10*	*@1*	*@3*	*@5*	*@10*	*@1*	*@3*	*@5*	*@10*
1	7.09	42.88	63.74	73.18	81.57	42.88	36.13	32.9	28.03	3.31	8.33	12.3	19.43	3.31	6.44	8.27	11.05
5	7.2	43.15	65.57	**74.98**	**82.94**	43.15	37.36	34.26	**29.16**	3.62	8.8	**13.13**	**20.6**	3.62	6.88	**8.9**	**11.92**
10	**7.25**	43.27	65.48	74.07	82.36	43.27	37.38	**34.27**	28.97	3.51	8.83	12.94	20.08	3.51	6.82	8.8	11.65
15	7.18	43.06	**66.66**	74.86	82.52	43.06	37.1	33.77	28.72	**3.64**	**8.98**	13.01	20.25	**3.64**	**6.93**	8.85	11.72
20	7.19	**43.49**	65.8	73.56	81.71	**43.49**	**38.14**	33.86	28.12	3.5	8.75	12.53	19.51	3.5	6.85	8.68	11.44

TABLE VI: Impact of the medical ontology weight on the number of concepts ever predicted ($CEP@k$, k being the number of top-ranked predictions considered).

Weight	CEP@1	CEP@3	CEP@5	CEP@10
1.0	75	152	208	366
0.8	81	146	222	383
0.6	**103**	**202**	**280**	**460**
0.4	70	146	217	396
0.2	74	144	223	393
0.0	79	165	251	439

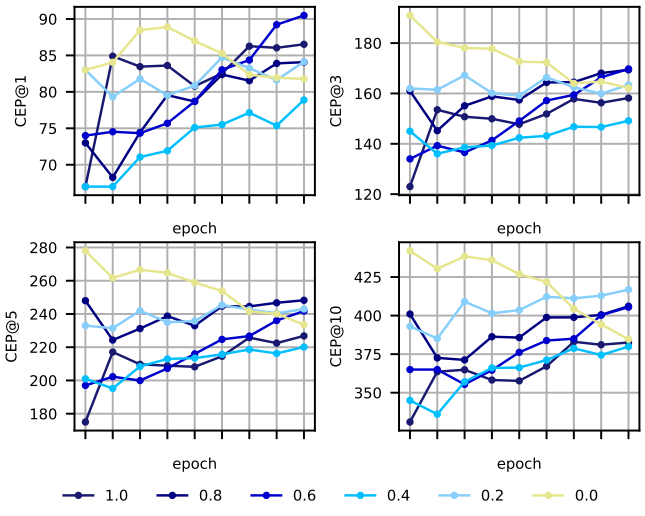


Fig. 5: Trends of concepts ever predicted (CEP) with different Medical Ontology weights on validation data over training epochs.

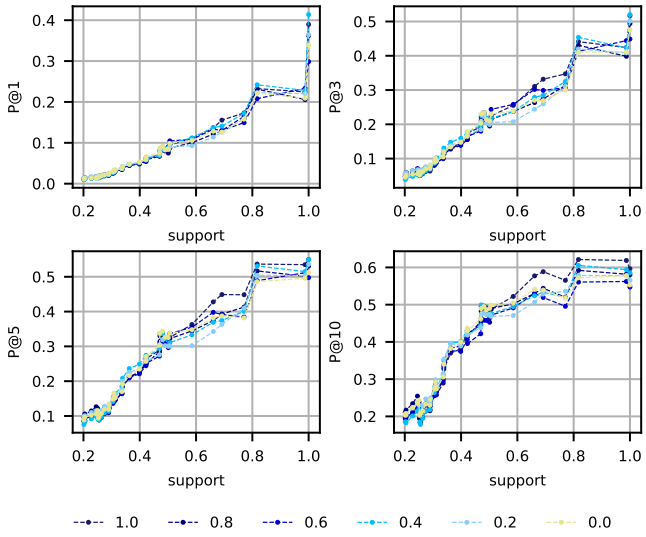


Fig. 6: Prevalence of concepts in test data (support) vs precision (P@k) of the model in identifying them. We plot only concepts appearing in more than the 20% of patients.

frequency of the predicted concepts in the patient population. As expected, the model performs better on concepts that are encountered frequently during its training, due to their higher support. However, we found that the utilization of biomedical ontologies provides superior results, particularly for concepts with lower support (see results between 0.6 and 0.8), where the use of medical ontologies proves to be more beneficial.

4) Performance vs medical history length: Figure 7 shows how model performance changes with the length of medical histories. Rather than a steady increase, there is an initial drop

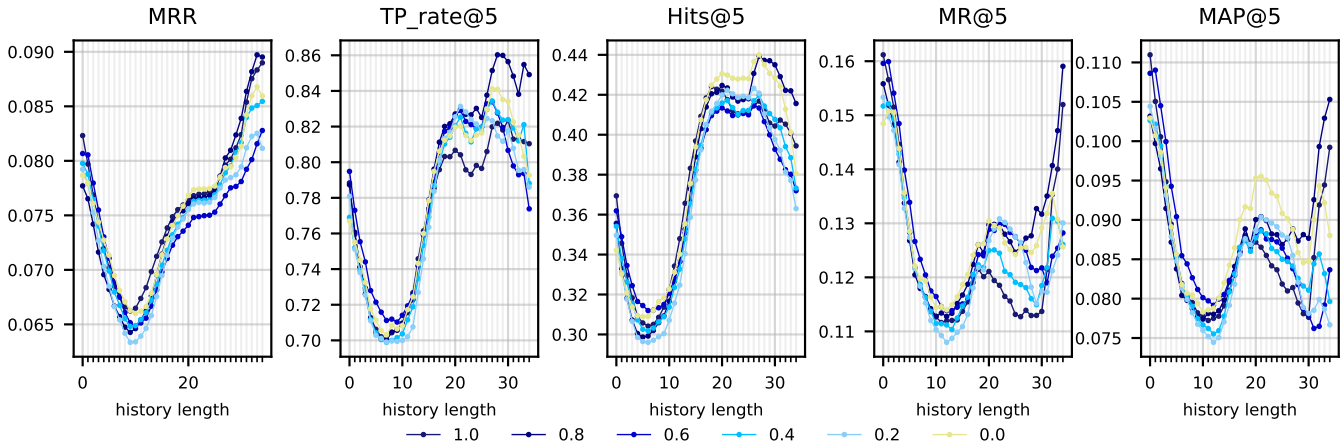


Fig. 7: Performance trends of models with different weights attributed to the medical ontology constraint as the length of medical histories increases.

in performance followed by an increase. This happens because initial disorders are easier to predict, due to the large dataset being used, the small number of concepts, and the high number of patients with just one or two events in their medical history, as depicted in Figure 3. In the middle part with the drop of performance, there is not enough data and a wide variety of concepts, while in the last part there is enough information in the timeline to infer the diagnosis. Comparing the different curves, we can observe that the use of the medical ontology can lead to improved results, especially for patients with shorter medical histories. This can be attributed to the model's inability to rely on sufficient temporal information, leading it to effectively leverage the medical ontology's knowledge.

VI. Conclusion & Future Work

In this study, we proposed a Temporal Knowledge Graph (TKG) framework, called MedTKG, for predicting future disorders by integrating both dynamic and static information from Electronic Health Records (EHRs) and medical ontologies, respectively. Our findings suggest that incorporating medical ontologies shows promise in improving the model's performance in predicting future disorders. Furthermore, we investigated the impact of different parameters controlling the weight assigned to the medical ontology during the training process. The inherent structure of knowledge graphs, like MedTKG, facilitates a transparent and interpretable modeling of patient medical histories, thereby addressing the growing need for explainability in biomedical applications. A key area of our future investigations will be to further enhance the explainability aspect of our model. We plan to delve deeper into how MedTKG can provide clear and understandable rationales for its predictions, making it a valuable tool for clinicians and healthcare providers. In our future work, we also aim to broaden the scope of our study by incorporating new datasets, including those in different languages, and expanding our predictions to cover additional medical events, such as medications and procedures. Furthermore, while we have shown that the utilization of static relationships between medical concepts produces an enhancement of model performance in predicting future disorders, the influence of such static information has the potential to be even greater. Notably, association networks between genes, disorders, symptoms, treatments, and other related factors are being developed by many researchers [55], [56]. Integrating these networks into the system could further amplify its performance. Additionally, we plan to evaluate the clinical utility of our framework by conducting a clinical trial with healthcare providers to assess its effectiveness in improving patient outcomes.

Acknowledgements

We acknowledge financial support from the PNRR MUR project PE0000013-FAIR.

References

[1] L. Rasmy, Y. Xiang, Z. Xie, C. Tao, and D. Zhi, "Med-bert: pre-trained contextualized embeddings on large-scale structured electronic health records for disease prediction," *CoRR*, vol. abs/2005.12833, 2020. [Online]. Available: https://arxiv.org/abs/2005.12833

[2] Y. Li, S. Rao, J. R. A. Solares, A. Hassaïne, D. Canoy, Y. Zhu, K. Rahimi, and G. S. Khorshidi, "BEHRT: transformer for electronic health records," *CoRR*, vol. abs/1907.09538, 2019. [Online]. Available: http://arxiv.org/abs/1907.09538

[3] J. Shang, T. Ma, C. Xiao, and J. Sun, "Pre-training of graph augmented transformers for medication recommendation," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, S. Kraus, Ed. ijcai.org, 2019, pp. 5953–5959. [Online]. Available: https://doi.org/10.24963/ijcai.2019/825

[4] V. Yadav and S. Bethard, "A survey on recent advances in named entity recognition from deep learning models," in *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, Aug. 2018, pp. 2145–2158. [Online]. Available: https://aclanthology.org/C18-1182

[5] E. French and B. T. McInnes, "An overview of biomedical entity linking throughout the years," *Journal of Biomedical Informatics*, vol. 137, p. 104252, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S153204642200257X

[6] X. Wang, X. He, Y. Cao, M. Liu, and T. Chua, "KGAT: knowledge graph attention network for recommendation," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, A. Teredesai, V. Kumar, Y. Li, R. Rosales, E. Terzi, and G. Karypis, Eds. ACM, 2019, pp. 950–958. [Online]. Available: https://doi.org/10.1145/3292500.3330989

[7] Z. Liu, C. Xiong, M. Sun, and Z. Liu, "Entity-duet neural ranking: Understanding the role of knowledge graph semantics in neural information retrieval," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 2395–2405. [Online]. Available: https://aclanthology.org/P18-1223

[8] T. Sun, Y. Shao, X. Qiu, Q. Guo, Y. Hu, X. Huang, and Z. Zhang, "Colake: Contextualized language and knowledge embedding," in *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, D. Scott, N. Bel, and C. Zong, Eds. International Committee on Computational Linguistics, 2020, pp. 3660–3670. [Online]. Available: https://doi.org/10.18653/v1/2020.coling-main.327

[9] O. Bodenreider, "The unified medical language system (UMLS): integrating biomedical terminology," *Nucleic Acids Res.*, vol. 32, no. Database-Issue, pp. 267–270, 2004. [Online]. Available: https://doi.org/10.1093/nar/gkh061

[10] B. Cai, Y. Xiang, L. Gao, H. Zhang, Y. Li, and J. Li, "Temporal knowledge graph completion: A survey," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China*. ijcai.org, 2023, pp. 6545–6553. [Online]. Available: https://doi.org/10.24963/ijcai.2023/734

[11] J. Wang, B. Wang, M. Qiu, S. Pan, B. Xiong, H. Liu, L. Luo, T. Liu, Y. Hu, B. Yin, and W. Gao, "A survey on temporal knowledge graph completion: Taxonomy, progress, and prospects," *CoRR*, vol. abs/2308.02457, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2308.02457

[12] C. Mo, Y. Wang, Y. Jia, and Q. Liao, "Survey on temporal knowledge graph," in *Sixth IEEE International Conference on Data Science in Cyberspace, DSC 2021, Shenzhen, China, October 9-11, 2021*. IEEE, 2021, pp. 294–300. [Online]. Available: https://doi.org/10.1109/DSC53577.2021.00047

[13] R. Miotto, L. Li, and J. T. Dudley, "Deep learning to predict patient future diseases from the electronic health records," in *Advances in Information Retrieval*, N. Ferro, F. Crestani, M.-F. Moens, J. Mothe, F. Silvestri, G. M. Di Nunzio, C. Hauff, and G. Silvello, Eds. Cham: Springer International Publishing, 2016, pp. 768–774.

[14] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun, "GRAM: graph-based attention model for healthcare representation learning," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017*. ACM, 2017, pp. 787–795. [Online]. Available: https://doi.org/10.1145/3097983.3098126

[15] F. Ma, Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao, "KAME: knowledge-based attention model for diagnosis prediction in healthcare," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018*, A. Cuzzocrea, J. Allan, N. W. Paton, D. Srivastava, R. Agrawal, A. Z. Broder, M. J. Zaki, K. S. Candan, A. Labrinidis, A. Schuster, and H. Wang, Eds. ACM, 2018, pp. 743–752. [Online]. Available: https://doi.org/10.1145/3269206.3271701

[16] Z. Kraljevic, A. Shek, D. Bean, R. Bendayan, J. T. Teo, and R. J. B. Dobson, "Medgpt: Medical concept prediction from clinical narratives," *CoRR*, vol. abs/2107.03134, 2021. [Online]. Available: https://arxiv.org/abs/2107.03134

[17] K. Taha, R. Davuluri, P. Yoo, and J. Spencer, "Personizing the prediction of future susceptibility to a specific disease," *PLOS ONE*, vol. 16, no. 1, pp. 1–26, 01 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0243127

[18] M. Shabaz and U. Garg, "Predicting future diseases based on existing health status using link prediction," *World Journal of Engineering*, 2021. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/WJE-10-2020-0533/full/html

[19] S. Wang, P. Ren, Z. Chen, Z. Ren, J. Ma, and M. de Rijke, "Order-free medicine combination prediction with graph convolutional reinforcement learning," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019*, W. Zhu, D. Tao, X. Cheng, P. Cui, E. A. Rundensteiner, D. Carmel, Q. He, and J. X. Yu, Eds. ACM, 2019, pp. 1623–1632. [Online]. Available: https://doi.org/10.1145/3357384.3357965

[20] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 2, pp. 494–514, 2022. [Online]. Available: https://doi.org/10.1109/TNNLS.2021.3070843

[21] T. Le, H. Nguyen, and B. Le, "A survey of the link prediction on static and temporal knowledge graph," *Journal of Research and Development on Information and Communication Technology*, 2021. [Online]. Available: https://ictmag.vn/ict/article/view/972

[22] J. Leblay and M. W. Chekol, "Deriving validity time in knowledge graph," in *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*, P. Champin, F. Gandon, M. Lalmas, and P. G. Ipeirotis, Eds. ACM, 2018, pp. 1771–1776. [Online]. Available: https://doi.org/10.1145/3184558.3191639

[23] A. Bordes, N. Usunier, A. García-Durán, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, C. J. C. Burges, L. Bottou, Z. Ghahramani, and K. Q. Weinberger, Eds., 2013, pp. 2787–2795. [Online]. Available: https://proceedings.neurips.cc/paper/2013/hash/1cecc7a77928ca8133fa24680a88d2f9-Abstract.html

[24] S. S. Dasgupta, S. N. Ray, and P. P. Talukdar, "Hyte: Hyperplane-based temporally aware knowledge graph embedding," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii, Eds. Association for Computational Linguistics, 2018, pp. 2001–2011. [Online]. Available: https://doi.org/10.18653/v1/d18-1225

[25] J. Leblay, M. W. Chekol, and X. Liu, "Towards temporal knowledge graph embeddings with arbitrary time precision," in *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, M. d'Aquin, S. Dietze, C. Hauff, E. Curry, and P. Cudré-Mauroux, Eds. ACM, 2020, pp. 685–694. [Online]. Available: https://doi.org/10.1145/3340531.3412028

[26] R. Goel, S. M. Kazemi, M. A. Brubaker, and P. Poupart, "Diachronic embedding for temporal knowledge graph completion," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, pp. 3988–3995. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/5815

[27] C. Xu, M. Nayyeri, F. Alkhoury, H. S. Yazdi, and J. Lehmann, "Temporal knowledge graph completion based on time series gaussian embedding," in *The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part I*, ser. Lecture Notes in Computer Science, J. Z. Pan, V. A. M. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, and L. Kagal, Eds., vol. 12506. Springer, 2020, pp. 654–671. [Online]. Available: https://doi.org/10.1007/978-3-030-62419-4_37

[28] C. Xu, M. Nayyeri, F. Alkhoury, H. Shariat Yazdi, and J. Lehmann, "TeRo: A time-aware knowledge graph embedding via temporal rotation," in *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 1583–1593. [Online]. Available: https://aclanthology.org/2020.coling-main.139

[29] T. Lacroix, G. Obozinski, and N. Usunier, "Tensor decompositions for temporal knowledge base completion," in *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. [Online]. Available: https://openreview.net/forum?id=rke2P1BFwS

[30] W. Jin, M. Qu, X. Jin, and X. Ren, "Recurrent event network: Autoregressive structure inferenceover temporal knowledge graphs," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds. Association for Computational Linguistics, 2020, pp. 6669–6683. [Online]. Available: https://doi.org/10.18653/v1/2020.emnlp-main.541

[31] Z. Li, X. Jin, W. Li, S. Guan, J. Guo, H. Shen, Y. Wang, and X. Cheng, "Temporal knowledge graph reasoning based on evolutional representation learning," in *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, and T. Sakai, Eds. ACM, 2021, pp. 408–417. [Online]. Available: https://doi.org/10.1145/3404835.3462963

[32] Y. He, P. Zhang, L. Liu, Q. Liang, W. Zhang, and C. Zhang, "HIP network: Historical information passing network for extrapolation reasoning on temporal knowledge graph," in *Proceedings of the*

Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021, Z. Zhou, Ed. ijcai.org, 2021, pp. 1915–1921. [Online]. Available: https://doi.org/10.24963/ijcai.2021/264

[33] N. Park, F. Liu, P. Mehta, D. Cristofor, C. Faloutsos, and Y. Dong, "Evokg: Jointly modeling event time and network structure for reasoning over temporal knowledge graphs," in *WSDM '22: The Fifteenth ACM International Conference on Web Search and Data Mining, Virtual Event / Tempe, AZ, USA, February 21 - 25, 2022*, K. S. Candan, H. Liu, L. Akoglu, X. L. Dong, and J. Tang, Eds. ACM, 2022, pp. 794–803. [Online]. Available: https://doi.org/10.1145/3488560.3498451

[34] Z. Kraljevic, T. Searle, A. Shek, L. Roguski, K. Noor, D. Bean, A. Mascio, L. Zhu, A. A. Folarin, A. Roberts, R. Bendayan, M. P. Richardson, R. Stewart, A. D. Shah, W. K. Wong, Z. M. Ibrahim, J. T. Teo, and R. J. B. Dobson, "Multi-domain clinical natural language processing with medcat: The medical concept annotation toolkit," *Artif. Intell. Medicine*, vol. 117, p. 102083, 2021. [Online]. Available: https://doi.org/10.1016/j.artmed.2021.102083

[35] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, 2022. [Online]. Available: https://doi.org/10.1109/TNNLS.2021.3084827

[36] C. Shang, Y. Tang, J. Huang, J. Bi, X. He, and B. Zhou, "End-to-end structure-aware convolutional networks for knowledge base completion," in *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019, pp. 3060–3067. [Online]. Available: https://doi.org/10.1609/aaai.v33i01.33013060

[37] R. Ye, X. Li, Y. Fang, H. Zang, and M. Wang, "A vectorized relational graph convolutional network for multi-relational network alignment," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, S. Kraus, Ed. ijcai.org, 2019, pp. 4135–4141. [Online]. Available: https://doi.org/10.24963/ijcai.2019/574

[38] G. Li, M. Müller, A. K. Thabet, and B. Ghanem, "Deepgcns: Can gcns go as deep as cnns?" in *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*. IEEE, 2019, pp. 9266–9275. [Online]. Available: https://doi.org/10.1109/ICCV.2019.00936

[39] M. S. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, ser. Lecture Notes in Computer Science, A. Gangemi, R. Navigli, M. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai, and M. Alam, Eds., vol. 10843. Springer, 2018, pp. 593–607. [Online]. Available: https://doi.org/10.1007/978-3-319-93417-4_38

[40] S. Vashishth, S. Sanyal, V. Nitin, and P. P. Talukdar, "Composition-based multi-relational graph convolutional networks," in *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. [Online]. Available: https://openreview.net/forum?id=BylA_C4tPr

[41] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, ser. Springer Series in Computational Mathematics. Springer, 1985, vol. 3. [Online]. Available: https://doi.org/10.1007/978-3-642-82118-9

[42] A. E. W. Johnson, T. J. Pollard, L. Shen, L. wei H. Lehman, M. Feng, M. M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific Data*, vol. 3, 2016.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[44] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in Neural Information Processing Systems 2, [NIPS Conference, Denver, Colorado, USA, November 27-30, 1989]*, D. S. Touretzky, Ed. Morgan Kaufmann, 1989, pp. 396–404. [Online]. Available: http://papers.nips.cc/paper/293-handwritten-digit-recognition-with-a-back-propagation-network

[45] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, and S. Khudanpur, "Recurrent neural network based language model," in *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010*, T. Kobayashi, K. Hirose, and S. Nakamura, Eds. ISCA, 2010, pp. 1045–1048. [Online]. Available: https://doi.org/10.21437/Interspeech.2010-343

[46] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, Eds., 2017, pp. 5998–6008. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[47] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *CoRR*, vol. abs/1803.01271, 2018. [Online]. Available: http://arxiv.org/abs/1803.01271

[48] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. F. Stewart, "RETAIN: an interpretable predictive model for healthcare using reverse time attention mechanism," in *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, Eds., 2016, pp. 3504–3512. [Online]. Available: https://proceedings.neurips.cc/paper/2016/hash/231141b34c82aa95e48810a9d1b33a79-Abstract.html

[49] L. Ma, J. Gao, Y. Wang, C. Zhang, J. Wang, W. Ruan, W. Tang, X. Gao, and X. Ma, "Adacare: Explainable clinical health status representation learning via scale-adaptive feature extraction and recalibration," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, pp. 825–832. [Online]. Available: https://doi.org/10.1609/aaai.v34i01.5427

[50] L. Ma, C. Zhang, Y. Wang, W. Ruan, J. Wang, W. Tang, X. Ma, X. Gao, and J. Gao, "Concare: Personalized clinical feature embedding via capturing the healthcare context," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, pp. 833–840. [Online]. Available: https://doi.org/10.1609/aaai.v34i01.5428

[51] J. Gao, C. Xiao, Y. Wang, W. Tang, L. M. Glass, and J. Sun, "Stagenet: Stage-aware neural networks for health risk prediction," in *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, Y. Huang, I. King, T. Liu, and M. van Steen, Eds. ACM / IW3C2, 2020, pp. 530–540. [Online]. Available: https://doi.org/10.1145/3366423.3380136

[52] J. Gao, C. Xiao, L. M. Glass, and J. Sun, "Dr. agent: Clinical predictive model via mimicked second opinions," *J. Am. Medical Informatics Assoc.*, vol. 27, no. 7, pp. 1084–1091, 2020. [Online]. Available: https://doi.org/10.1093/jamia/ocaa074

[53] C. Zhang, X. Gao, L. Ma, Y. Wang, J. Wang, and W. Tang, "GRASP: generic framework for health status representation learning based on incorporating knowledge from similar patients," in *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 2021, pp. 715–723. [Online]. Available: https://doi.org/10.1609/aaai.v35i1.16152

[54] C. Yang, Z. Wu, P. Jiang, Z. Lin, J. Gao, B. Danek, and J. Sun, "PyHealth: A deep learning toolkit for healthcare predictive modeling," in *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) 2023*, 2023. [Online]. Available: https://github.com/sunlabuiuc/PyHealth

[55] K. Lu, K. Yang, H. Sun, Q. Zhang, Q. Zheng, K. Xu, J. Chen, and X. Zhou, "Sympgan: A systematic knowledge integration system for symptom-gene associations network," *Knowl. Based Syst.*, vol. 276, p. 110752, 2023. [Online]. Available: https://doi.org/10.1016/j.knosys.2023.110752

[56] J. P. González, J. M. Ramírez-Anguita, J. Saüch-Pitarch, F. Ronzano, E. Centeno, F. Sanz, and L. I. Furlong, "The disgenet knowledge platform for disease genomics: 2019 update," *Nucleic Acids Res.*, vol. 48, no. Database-Issue, pp. D845–D855, 2020. [Online]. Available: https://doi.org/10.1093/nar/gkz1021