Diagnosis Prediction over Patient Data using Hierarchical Medical Taxonomies*

Emil Riis Hansen[1], Tomer Sagi[1] and Katja Hose[1]

[1]Department of Computer Science, Aalborg University, Aalborg, Denmark

Abstract

A variety of hierarchical domain taxonomies exist in the medical domain for describing medical concepts such as laboratory tests, medications, and procedures. The structural information contained within domain taxonomies contains rich semantic information pertaining to the described concepts and their relationships to each other. As AI models are successfully applied in many medical areas, it is only natural to explore integrating AI models with medical domain taxonomies. However, only a few, nascent attempts have been made. In this work, we investigate how the structure of hierarchical medical taxonomies can be used to improve the performance of a diagnosis prediction task. Specifically, we suggest a method titled *TreeEmb* to pre-initialize the node embeddings of a patient graph derived from electronic health records using information from the taxonomy. We expect this method to improve the performance of graph convolution network models over the enriched patient graph. We evaluate our method over a patient graph created from the MIMIC-IV electronic health record dataset enriched by initializing node embeddings using hierarchical medical taxonomies. We use type-specific domain knowledge from hierarchical medical taxonomies such as the ICD-9 procedures, ATC medication, and LOINC laboratory test taxonomies. Experimental results from a multi-label diagnosis prediction task over this graph demonstrate the efficacy of our approach.

Keywords

Hierarchical Domain Knowledge, Embedding Initialization, Multi-Label Classification, Graph Convolution Networks, Patient Diagnosis Prediction, Inductive Artificial Intelligence

1. Introduction

The medical domain has accumulated an abundance of domain knowledge structured as hierarchical taxonomies. Integrating semantically rich domain knowledge such as hierarchical taxonomies into Artificial Intelligence (AI) technologies could improve their predictive capabilities in numerous medical applications such as patient diagnosis prediction and protein function prediction using end-to-end supervised learning [1].

Patients' Electronic Health Records (EHR) can be readily modeled as multi-relational graphs connect patients with their associated medical histories, such as prescriptions, laboratory tests, and procedures, as illustrated in Figure 1. We, henceforth, name such graphs *EHR graphs*. The AI technology of Graph Convolution Networks (GCNs) has recently become the *de facto* standard for solving many medical problems over EHR graphs due to their seamless ability to learn latent node embeddings for subsequent down-stream tasks, such as node classification, link prediction, and whole graph classification in an end-to-end manner [2].

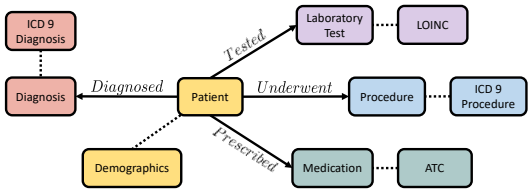

Figure 1: EHR graph representation relating patients to laboratory tests, procedure codes, and medication intake. Dashed lines represent related information such as patient demographics and hierarchical medical structures such as LOINC, ICD-9 Procedures, and ATC as described in Section 3.2.

Much work has recently been put into the model-centric development of novel GCN architectures, such as RelationalGCN [3] utilizing the multi-relational nature of graphs and GraphSAGE [4] with a scalable node sampling approach. However, although rich semantic information often exists alongside medical graphs, such as textual descriptions, hierarchical taxonomies, and uncertainty information [5], only a few works investigate incorporating such information in a data-centric way for improving classification and regression tasks [6].

As the structure of hierarchical medical domain taxonomies contains human-curated knowledge pertaining to the properties and similarity between taxonomic concepts, we surmise that such structural knowledge can

Published in the Workshop Proceedings of the EDBT/ICDT 2023 Joint Conference (March 28-March 31, 2023), Ioannina, Greece

* Research

emilrh@cs.aau.dk (E. R. Hansen); tsagi@cs.aau.dk (T. Sagi); khose@cs.aau.dk (K. Hose)

0000-0003-4103-1244 (E. R. Hansen); 0000-0002-8916-0128 (T. Sagi); 0000-0001-7025-8099 (K. Hose)

benefit downstream tasks if integrated into AI models. Hence, in this paper, we investigate a method termed *TreeEmb* for encoding the structure of hierarchical medical domain taxonomies to pre-initialize node embeddings in EHR graphs for improved classification performance in a patient diagnosis code prediction task.

This paper is structured as follows; in Section 2, we present related work using domain hierarchies in the initialization of node embeddings and the task of patient diagnosis prediction using graph convolution networks. Section 3 presents the proposed method and theoretical concepts. In Section 4, we present the data used for experimentation, followed by Section 5, where the experimental setup and results are analyzed and explained. Lastly, in Section 6, we conclude and introduce future work.

2. Related Work

Embedding Initialization. Research into integrating domain information, such as textual descriptions, images, type-hierarchies, and uncertainty information into graph convolution models has lately shown promise [5]. Pre-initializing node embeddings is a central method for integrating auxiliary information with graph convolution networks. Hamilton et al. [4] use text attributes, node profile information, and node degrees to pre-initialize embeddings of three datasets. Zhao et al. [7] use TF/IDF and binary word presence vectors to pre-initialize node embeddings for citation graphs. Other works pre-initialize node embeddings by extracting graphlet features directly from the structure of the input graph [8]. Ali et al. [9] construct manual features such as age and follower count for each social network user. While individual or combinations of manually constructed features have shown promising results for the pre-initialization of node embeddings, none of these works have so far investigated integrating hierarchical domain taxonomies to pre-initialize node embeddings.

Patient Diagnosis Prediction. Diagnosis prediction is the vital medical application of finding patient comorbidities using the patient's medical history [10]. Hierarchical domain knowledge has recently been introduced into various AI models for diagnosis prediction. In [11], hierarchical medical taxonomies are used to embed medical concepts to leverage the general problem of data insufficiency and model interpretability by learning hierarchical medical concept embeddings, pre-initialized on co-occurrence information by a weighted sum of concept paths. Instead, in this work, we propose using the concept taxonomies for pre-initializing node embeddings of a medical patient graph for subsequent GCN-based diagnosis prediction. The approach by Sun et al. [12] utilizes GCNs on two bipartite graphs, e.g., symptom-relationship and patient-diagnosis, to learn an optimized space wherein patients will have a small distance to assigned diagnosis concepts. However, instead of dividing domain knowledge and patient information into separate bipartite graphs, we investigate the effect of integrating hierarchical auxiliary domain knowledge with a patient graph consisting of multiple patients and their related medical concepts, not limited to symptoms. The work closest to ours is that of [13], in which a knowledge graph is built using auxiliary domain knowledge from the MEDLINE medical corpus for multi-label prediction of patient diseases. Patients are associated with diagnosis codes related to laboratory tests, habits, and profiles in their work. However, different from our work, their method of diagnosis prediction is not related to graph convolutions, and patients are not associated with each other.

3. Initializing Graph Embeddings

In this section, we formalize our method *TreeEmb* of using hierarchical medical taxonomies to pre-initialize node embeddings for the medical application of multi-label diagnosis prediction. The overall approach is illustrated in Figure 2, with section references for further details.

An EHR graph is first created from an EHR dataset as detailed in Section 3.1. Concept embeddings are then created from the hierarchical medical taxonomies' structure to derive meaningful latent descriptions of medical concepts and used to pre-initialize node embeddings in the EHR graph as described in Section 3.2. Finally, multiple layers of graph convolutions, as described in Section 3.3, are trained for multi-label patient diagnosis prediction.

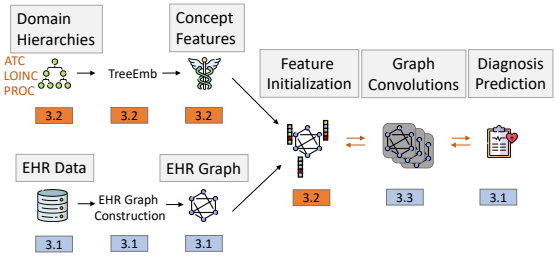

Figure 2: Illustration of the overall approach. Blue boxes reference sections with further details on the specific step. Arrows represent the directional flow of data. Orange boxes represent our primary contribution of pre-initializing graph node embeddings using concept features extracted from hierarchical medical domain taxonomies. Orange arrows describe the parts of the approach that are learned using backpropagation.

3.1. Multi-label Diagnosis Prediction over EHR Data

This section introduces how a multi-relational patient-centric graph can be constructed from an EHR dataset and the challenge of multi-label patient diagnosis prediction.

EHR data relate patients to medical concepts such as medications, laboratory tests, and procedures. Given a set of patients S and a set of medical concepts C, where $C^t \subset C$ is the subset of distinct medical concepts types, then an EHR dataset can formally be defined as the set H of tuples (s, c) relating a patient $s \in S$ with an associated medical concept $c \in C$.

Given the example EHR dataset H and a set of patients S as illustrated in Figure 3**a)**, we create an EHR graph as follows.

The set of graph nodes V is created as the union between the set of unique patients and the set of unique medical concepts from C as illustrated in Figure 3**b)**, and the graph edges are created as the set E of relations and reversed relations between concepts and patients from H. Furthermore, every edge in E is given an edge type as specified by the medical concept type involved in the relation. As an example, the edge (s_1, c_1^m) could have an edge type of $prescribed$ as illustrated in Figure 3**c)**, as the patient s_1 has been prescribed the medication c_1^m. The final patient graph created from H and S is illustrated in Figure 3**c)**. For brevity, reverse relations are not depicted in the graph. Over this graph, we define the mapping function $t_v : V \rightarrow T$ for getting the type of a node, the function $t_e : E \rightarrow R$ for getting the type r of an edge, and the function $f_v : V \rightarrow F$ for getting the embedding f_i^t of a node v_i of type $t = t_v(v_i)$.

Given an EHR dataset and a set of diagnosis concepts D, the challenge of patient diagnosis prediction is to find the subset $D' \subset D$ pertaining to a patient $s \in S$ s.t. D' matches the actual set of diagnosis concepts related to the patient. We model this challenge as a multi-label classification problem.

3.2. Pre-initialization Using Domain Hierarchies

Node features can be either pre-initialized using entity-specific information or random-initialized and learned as part of the model training process. Pre-initialization of node embeddings can be done by extracting type-specific entity information from the nodes or by extracting features from the graph structure. Examples of the former are pre-trained convolution neural networks for imaging information and natural language processing models for text data. An example of the use of graph structure is by counting sub-structures such as graphlets [14]. However, an overlooked source of rich semantic information can be found in type-specific domain hierarchies prevalent

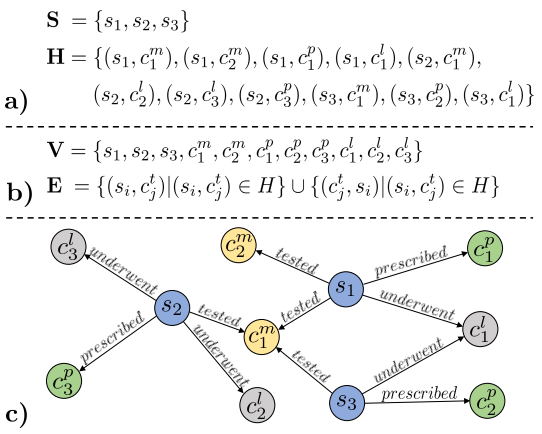

Figure 3: Illustration of the EHR graph creation process. **a)** set of patients S and an EHR datset H. **b)** the graph nodes V and graph edges E. **c)** the final graph represented by nodes and typed directed edges. For brevity, reverse relations are not depicted in the graph.

in many domains. Domain hierarchies are curated hierarchies of related concepts. Inherently, their structure contains knowledge regarding the relationship between concepts, and each hierarchical layer contains information about the properties of its concepts. Hence, we argue that the position of a concept within hierarchies contains rich semantic information.

In the medical domain, structured medical concepts such as medications, diagnoses, laboratory tests, and procedures are coded in hierarchical taxonomies. Medication can be coded using the world health organization's anatomical therapeutic classification system (ATC) [15] and classifies medication based on its active ingredients and organ or system. Hence, the location of medications within the hierarchy contains semantic information relevant to the task of diagnosis prediction. As an example, for the medication with code $A10BA02$, e.g., metformin, the first level of the ATC hierarchy specifies that the medication targets the alimentary tract and metabolism system. Level two specifies the therapeutic subgroup, e.g., the drug is used in diabetes. Level three defines the pharmacological subgroup, e.g., the drug lowers blood glucose. The fourth level indicates the chemical subgroup of the drug, in this case, biguanides, and the last level specifies the chemical substance, e.g., metformin. Given that a patient has received metformin, the patient likely suffered from type 2 diabetes. Explicitly integrating such hierarchical information into concept embeddings should enable the AI model to learn from the proximity of similar concepts.

Surgical procedures performed on patients can be coded using the ICD-9 Procedures (PROC) taxonomy [16]

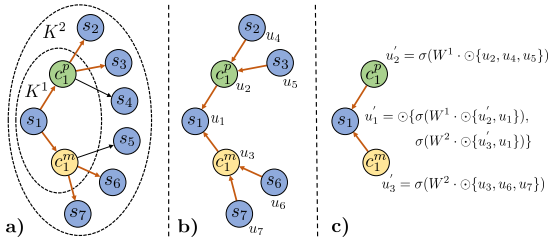

Figure 4: Example steps of our graph convolution. **a)** A 2-layer 2-node fanout sampling strategy finds the neighborhood K^1 of patient s_1. Each of the sampled nodes $\{c_1^p, c_1^m\}$ uses the same sampling strategy on their immediate neighborhood K^2 to further sample nodes $\{s_2, s_3, s_6, s_7\}$. **b)** The sampled subgraph with node features u_1 to u_7 as extracted through the node feature mapping f_v. **c)** Our combined graph convolution aggregate and update step. A relation-specific transformation matrix W^i is applied to the element-wise mean $\odot$ of similar typed entities as done in [4]. Finally, a non-linear activation function σ is applied to individual convolutions. If different typed features are to be combined as in the combination of $\{u_1, u_2^{'}, u_3^{'}\}$ the element-wise mean combines individual transformations.

grouping related procedures based on their site of operation. Given that a patient has received the surgical procedure with code 07.2, e.g., partial adrenalectomy, the patient likely suffered from a disease related to the endocrine glands.

Laboratory tests can be coded using the LOINC concept codes [17] over which a hierarchical taxonomy exists, grouping related laboratory tests by their class, component, and system, providing valuable information on the purpose of laboratory tests.

Using the aforementioned hierarchical medical taxonomies, and the example of the medical concept with code $B02AA02$, e.g., Tranexamic acid from the ATC hierarchy, we propose the *TreeEmb* method for pre-initializing node embeddings using type-specific hierarchical domain knowledge. Starting from 0, a unique index is assigned to each node in the tree as illustrated in Figure 5**b)**. Subsequently, a depth-first search is performed from the root of the hierarchical domain taxonomy to each leaf node for collecting the indexes along the shortest path to each leaf. Suppose the concept $B02AA02$ is given the initial index 3, then the indexes between $B02AA02$ and the root node is $[0, 1, 3]$ as illustrated in Figure 5**c)**. Eventually, leaf nodes are assigned an embedding as the one-hot encoded version of their shortest path indexes. As the concept $B02AA02$ has accumulated indexes $[0, 1, 3]$ and as the example tree has 11 nodes, $B02AA02$ is assigned an embedding vector of dimensionality 11 with 1 in the positions 0, 1, and 3 and 0 in every other position as illustrated in Figure 5**d)**. The computed features of tree leaf concepts can then be used to pre-initialize node embeddings. Furthermore, using this embedding technique ensures that concepts closely related in the tree will have similar embeddings compared to concepts far away. Hence, we conjecture that GCNs will be more easily able to learn that groups of closely related concepts are used in treating the same disease, thus decreasing the epistemic uncertainty by adding domain knowledge.

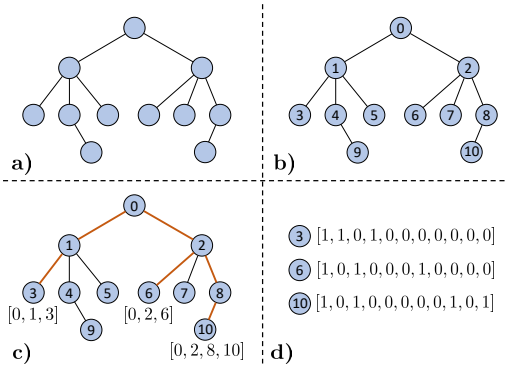

Figure 5: *TreeEmb* method for constructing concept embeddings from hierarchical taxonomies. **a** A tree-structured hierarchical taxonomy. **b)** breath first search indexes every node from 0. **c)** Depth-first search from the root to each leaf collects shortest path indexes. **d** One-hot encoding generates embeddings for leaf node concepts.

3.3. Graph Convolution Networks

Graph convolutions can learn from the structure of graphs by propagating node features between neighboring nodes using learnable *aggregation* and *update* functions as illustrated in Figure 4. Aggregation functions combine neighborhood information by imposing transformation matrices on the output of the neighborhood aggregation. Update functions, then learn how to integrate information from the current node embedding and the features of the neighborhood aggregation function. We employ a multi-relational variant of the GraphSAGE [4] algorithm for learning latent node embeddings for graphs with multiple relation types between concept nodes by exploiting not only the structure but also the multi-relational nature of EHR graphs.

4. Data

We perform experiments on the MIMIC-IV [18] EHR dataset from PhysioNet [19] consisting of 382, 278 intensive care unit patients from the Beth Israel Deaconess Medical Center from the period 2008 to 2019. MIMIC-IV

encompasses laboratory results, vital signs, diagnoses ascertained, administered medications, and demographics. The data is structured as a relational database.

To disambiguate medical concepts, we transform the dataset into the observational medical outcomes partnership (OMOP) common data model (CDM) [20] using an extract-transform-load (ETL) conversion flow.[1] The CDM format disambiguates and standardizes medical concepts and thus provides a means of interoperability for subsequent AI models to operate on disparate medical datasets converted into the CDM. In the CDM format laboratory tests are coded using the LOINC taxonomy, procedures are coded using the ICD-9 procedures taxonomy, and laboratory tests are coded using the RxNorm taxonomy [21]. Since RxNorm is a flat taxonomy, we map each medication concept through its active ingredients to the hierarchical ATC medication taxonomy.

Table 1
Number of distinct concepts for EHR data types.

Data Type	Distinct Concepts
Medication	1,749
Diagnosis	537
Laboratory Test	1,328
Procedure	1,228

For patient multi-label diagnosis prediction, we build the EHR graph based on patient diagnostic EHR concept types used in related work in EHR-based diagnosis prediction [22, 23, 10] and end up with demographic information, prescriptions, procedures, laboratory tests, and the task labels as patient diagnosis codes.

Patient diagnosis codes are coded using the 9th version of the International Classification of Diseases (ICD-9) and consist of approximately $13,000$ diagnosis codes [24]. We omit codes related to the ICD-9 E and V hierarchies as these are related to external causes of injury and are generally not discernible by EHR data. We further omit hierarchies of codes as summarized in Table 2. Omitting these hierarchies, we are left with $8,681$ disease codes. Since it is usually not possible to generalize from a low number of cases, we omit codes for which less than 500 patient cases exist. We are ultimately left with $128,605$ patients diagnosed with a total of $1,054,670$ diagnoses from 537 distinct diagnosis codes. The full list of 537 diagnosis codes are available online[2]. Table 1 summarizes the number of distinct concepts for each medical EHR concept type.

[1] https://github.com/OHDSI/MIMIC
[2] https://github.com/dkw-aau/graph_embedding_initialization

Table 2
Summarizing disease codes omitted from further analysis.

Codes	Count	Description
$290-319$	375	Mental Disorders
$630-679$	530	Comp. of Pregnancy
$780-799$	330	Injuries and Poison
$800-999$	$1,617$	Ill-Defined Conditions
E and V	$1,467$	Ext. Causes of Injury

5. Experiments and Results

To investigate the effect of pre-initializing node embeddings using domain hierarchies, we conduct several empirical experiments as summarized in Table 3 using the model pipeline as illustrated in Figure 2. Each experiment is trained on the problem of multi-label patient diagnosis prediction using a multi-relational version of the GraphSAGE algorithm as described in Section 3 with the input EHR dataset described in Section 4. In the **Rand** experimental setting, initial graph node embeddings are random-initialized using Xavier initialization [25] and made trainable as part of the supervised model training phase [26]. Hence, **Rand** serves as a transductive baseline experiment. Transductive methods generally perform better on subsequent downstream prediction tasks, however, with the cost of not being able to extrapolate to unseen examples [4].

Table 3
Overview of experimental settings.

Experiment	Learning	Embedding Data
FeatInit	Inductive	Hierarchical Taxonomies
Rand	Transductive	Xavier Initialization
Graphlet	Inductive	Graph Structure

In the **Graphlet** experimental setting, features are pre-initialized using state-of-the-art graphlet and edge count features [14] as in [8]. **Graphlet** serves as an inductive baseline experiment, as trained models can extrapolate to unseen examples.

The **FeatInit** experimental setting investigates the effect of pre-initializing node concept embeddings using the latent information contained within hierarchical medical taxonomies using the *TreeEmb* method as described in Section 3. In **FeatInit**, node embeddings should already contain domain information relevant to the task of diagnosis prediction; hence embeddings are kept constant during training. Furthermore, in the **FeatInit** experimental setting, patient features are pre-initialized using categorical values for sex, race, and ethnicity and a continuous variable for the patient's age. Moreover, as **FeatInit** does not train node embeddings, trained models can extrapolate to unseen examples.

Table 4
Parameter settings for hyperparameter optimization using tree-based Parzen estimation. U means uniform distribution.

Parameter	Values
Model Depth	{2, 3}
Learning Rate	{1e-3, 5e-3, 1e-2}
Dropout	$U(0.0..0.5)$
Hidden Dim	{32, 64, 128, 256}

5.1. Experimental Details

For each experiment, we perform 100 iterations of tree-based Parzen estimation (TPE) [27] for hyperparameter optimization over the set of parameters as summarized in Table 4. Each iteration is trained using the Adam [28] variation of stochastic gradient descent with binary cross-entropy as the loss function. Each experimental setting is investigated on the prediction of five sets of diagnosis codes as in [29, 30], with each set relating to a level of aggregation on the hierarchical ICD-9 diagnosis taxonomy. In the first setting, named *L5*, the task is to predict the raw comorbidities of patients from the entirety of the 537 diagnosis codes as described in Section 4. The remaining settings investigate diagnosis code prediction on aggregated levels of the ICD-9 diagnosis taxonomy named *L4* through *L1* with 427 disparate diagnosis codes for *L4* to 13 disparate diagnosis codes for *L1*. Aggregating diagnosis codes enables us to investigate the effect of pre-initializing graph concept embeddings from hierarchical medical taxonomies extracted through *TreeEmb* on classification problems of varying complexities.

As graph convolutions require the same dimensionality for each node type, we do an initial transformation on node input features using type-specific non-linear transformations into the feature dimensionality required by the graph convolution layers. Thus, the transformation is learned end-to-end with the task of diagnosis prediction. Additionally, we transform the output node embeddings as computed by the final convolution layer using

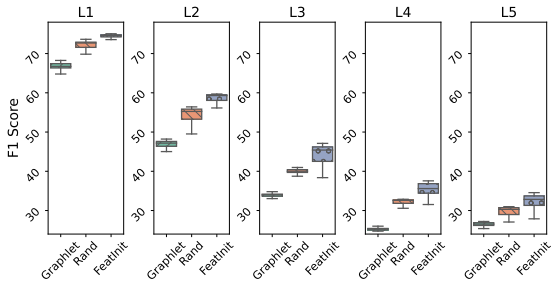

Figure 6: Experimental results of diagnosis code prediction on five sets of diagnosis codes for the experimental settings **Rand**, **Graphlet**, and **FeatInit**.

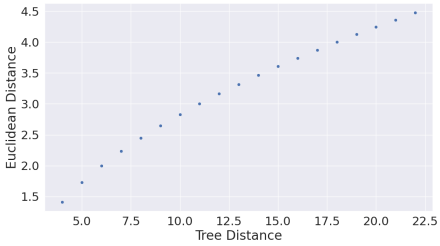

Figure 7: Monotonicity of LOINC concept embedding space.

a non-linear transformation into the dimensionality of the number of diagnosis codes in a specific level of ICD-9 aggregation, such that we end up with one output node for each predictable diagnosis code. We split patients into training validation and test sets with sizes 80/10/10 and used early stopping based on validation loss.

To evaluate and compare across experimental settings, we use the standard harmonic mean F1 value between the micro-averaged precision and recall as it is commonly used in the evaluation of multi-label classification tasks [13]. Furthermore, to investigate the robustness of pre-initializing features using *TreeEmb* embeddings, we evaluate the median over all 100 model iterations for each experiment. All experimental code and data are available online[3].

5.2. Results and Analysis

Figure 6 presents the results for each experimental setting over all iterations of the TPE. Experimental results in terms of the F1 value for the median and best-performing models are summarized in Table 5.

As illustrated in Figure 6, using *TreeEmb* embeddings for pre-initializing node features resulted in improved F1 scores compared to learning node embeddings as part of the training and pre-initialization using graphlet features. Furthermore, using unpaired t-test between **Rand** and **FeatInit** and between **Graphlet** and **FeatInit** results for any level of diagnosis code aggregation results in the two-tailed P value $p < .001$, which by conventional criteria indicates a statistically significant difference between the two groups.

As summarized in Table 5, for each setting, the best performing **FeatInit** model outperforms the best performing **Rand** and **Graphlet** models by $1.42 - 6.14$ and $6.80 - 12.30$ percentage points in terms of F1 score respectively. These results indicate that the initialization of node features using the hierarchical knowledge contained within domain taxonomies could provide valuable

[3] https://github.com/dkw-aau/graph_embedding_initialization

Table 5
Experimental results in terms of harmonic mean F1 scores for the experimental settings **Graphlet**, **Rand**, and **FeatInit** on five diagnosis code prediction problems with varying number of classes. Imp. presents the relative improvement in terms of F1 value for initializing concept embeddings using the *TreeEmb* embeddings.

	Median				Best			
Setting	**Graphlet**	**Rand**	**FeatInit**	**Imp.**	**Graphlet**	**Rand**	**FeatInit**	**Imp.**
L5 - 537 codes	26.54	30.30	32.84	**2.54**	27.21	30.98	34.56	**3.58**
L4 - 427 codes	25.27	32.58	35.60	**3.02**	26.01	32.87	37.56	**4.69**
L3 - 229 codes	34.00	40.01	45.40	**5.39**	34.81	40.97	47.11	**6.14**
L2 - 61 codes	47.18	55.25	59.30	**4.05**	48.21	56.41	59.69	**3.28**
L1 - 13 codes	66.72	72.69	74.58	**1.89**	68.25	73.63	75.05	**1.42**

knowledge for solving domain-specific problems such as the medical problem of patient diagnosis prediction.

The embeddings produced by *TreeEmb* should reflect the structure of the hierarchical taxonomy. Assuming that semantically similar concepts are close in the tree and disparate concepts far from each other, the distance between constructed embeddings should increase as the path length between nodes in the tree increases. To investigate this aspect of the *TreeEmb* embeddings, we compared the Euclidean distance between pairs of concept embeddings with the length of the shortest path on the tree between the pairs. As illustrated in Figure 7, the Euclidean distance between node embeddings is a monotonic increasing function given the length of the shortest path between nodes. This means that similar concepts will have similar embeddings while dissimilar concepts will have disparate embeddings.

6. Conclusion

In this work, we proposed that hierarchical medical taxonomies contain valuable knowledge that can be utilized by the pre-initialization of graph node embeddings. We then presented a method termed *TreeEmb* to do so. We evaluated the proposed method on the medical problem of multi-label diagnosis prediction by constructing *TreeEmb* embeddings for the pre-initialization of concept nodes in an EHR graph for the three medical hierarchical taxonomies ATC, LOINC, and ICD-9 Procedures. Experimental results from the prediction task on five different sets of diagnosis codes of varying difficulty demonstrate the superiority of *TreeEmb* embeddings over a transductive baseline of learned concept embeddings and an inductive baseline of pre-computed graphlet features. All experimental code and data are available online[3].

For future work, we aim to investigate the proposed method in domains beyond the medical. Furthermore, since not all levels of hierarchical domain taxonomies may be equally important for the given prediction task, we aim to investigate trainable attention mechanisms for constructing concept embeddings from only the most relevant hierarchical knowledge. We also aim to explore other graph convolution models, including attention techniques.

Acknowledgments

This work is partially supported by the Poul Due Jensen Foundation.

References

[1] P. C. Sen, M. Hajra, M. Ghosh, Supervised classification algorithms in machine learning: A survey and review, in: Emerging technology in modelling and graphics, Springer, 2020, pp. 99–111. doi:10.1007/978-981-13-7403-6_11.

[2] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, S. Y. Philip, A comprehensive survey on graph neural networks, IEEE transactions on neural networks and learning systems 32 (2020) 4–24. doi:10.1109/TNNLS.2020.2978386.

[3] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, M. Welling, Modeling relational data with graph convolutional networks, in: European semantic web conference, Springer, 2018, pp. 593–607. doi:{10.1007/978-3-319-93417-4_38}.

[4] W. L. Hamilton, R. Ying, J. Leskovec, Inductive representation learning on large graphs, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 1025–1035.

[5] S. Ji, S. Pan, E. Cambria, P. Marttinen, P. S. Yu, A survey on knowledge graphs: Representation, acquisition, and applications, IEEE Trans. Neural Networks Learn. Syst. 33 (2022) 494–514. doi:10.1109/TNNLS.2021.3070843.

[6] J. D. Bossér, E. Sörstadius, M. H. Chehreghani, Model-centric and data-centric aspects of active learning for deep neural networks, in: IEEE BigData, IEEE, 2021, pp. 5053–5062.

[7] Z. Zhao, H. Zhou, L. Qi, L. Chang, M. Zhou, Inductive representation learning via cnn for partially-unseen attributed networks, IEEE Transactions on Network Science and Engineering 8 (2021) 695–706. doi:10.1109/TNSE.2020.3048902.

[8] R. A. Rossi, R. Zhou, N. K. Ahmed, Deep inductive network representation learning, in: Companion Proceedings of the The Web Conference 2018, 2018, pp. 953–960. doi:10.1145/3184558.3191524.

[9] S. A. Alhosseini, R. B. Tareaf, P. Najafi, C. Meinel, Detect me if you can: Spam bot detection using inductive representation learning, in: WWW (Companion Volume), ACM, 2019, pp. 148–153. doi:10.1145/3308560.3316504.

[10] Y. Si, J. Du, Z. Li, X. Jiang, T. Miller, F. Wang, W. J. Zheng, K. Roberts, Deep representation learning of patient data from electronic health records (ehr): A systematic review, Journal of Biomedical Informatics 115 (2021) 103671. doi:{10.1016/j.jbi.2020.103671}.

[11] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, 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, 2017, pp. 787–795. doi:10.1145/3097983.3098126.

[12] Z. Sun, H. Yin, H. Chen, T. Chen, L. Cui, F. Yang, Disease prediction via graph neural networks, IEEE Journal of Biomedical and Health Informatics 25 (2020) 818–826. doi:10.1109/JBHI.2020.3004143.

[13] T. Pham, X. Tao, J. Zhang, J. Yong, Y. Li, H. Xie, Graph-based multi-label disease prediction model learning from medical data and domain knowledge, Knowledge-Based Systems (2021) 107662. doi:10.1016/j.knosys.2021.107662.

[14] N. K. Ahmed, J. Neville, R. A. Rossi, N. Duffield, Efficient graphlet counting for large networks, in: 2015 IEEE International Conference on Data Mining, IEEE, 2015, pp. 1–10. doi:10.1109/ICDM.2015.141.

[15] M. Ronning, A historical overview of the atc/ddd methodology, WHO drug information 16 (2002) 233.

[16] WHO, et al., International classification of diseases:[9th] ninth revision, basic tabulation list with alphabetic index, World Health Organization, 1978.

[17] C. J. McDonald, S. M. Huff, J. G. Suico, G. Hill, D. Leavelle, R. Aller, et al., Loinc, a universal standard for identifying laboratory observations: a 5-year update, Clinical chemistry 49 (2003) 624–633.

[18] A. Johnson, L. Bulgarelli, T. Pollard, S. Horng, L. Celi, R. Mark, Mimic-iv (version 0.4). physionet, 2020.

[19] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, et al., Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals, Circulation 101 (2000) e215–e220.

[20] G. Hripcsak, J. D. Duke, N. H. Shah, C. G. Reich, V. Huser, M. J. Schuemie, et al., Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers., Studies in health technology and informatics 216 (2015) 574–8.

[21] S. J. Nelson, K. Zeng, J. Kilbourne, T. Powell, R. Moore, Normalized names for clinical drugs: Rxnorm at 6 years, Journal of the American Medical Informatics Association 18 (2011) 441–448.

[22] A. Hosseini, T. Chen, W. Wu, Y. Sun, M. Sarrafzadeh, Heteromed: Heterogeneous information network for medical diagnosis, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 763–772.

[23] Z. Liu, X. Li, H. Peng, L. He, S. Y. Philip, Heterogeneous similarity graph neural network on electronic health records, in: 2020 IEEE International Conference on Big Data (Big Data), IEEE, 2020, pp. 1196–1205. doi:10.1109/BigData50022.2020.9377795.

[24] D. J. Cartwright, Icd-9-cm to icd-10-cm codes: what? why? how?, in: Advances in wound care, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, 2013, pp. 588–592. doi:10.1089/wound.2013.0478.

[25] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.

[26] W. L. Hamilton, R. Ying, J. Leskovec, Representation learning on graphs: Methods and applications, arXiv preprint arXiv:1709.05584 (2017).

[27] J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization, in: NIPS, 2011, pp. 2546–2554.

[28] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: ICLR (Poster), 2015.

[29] T. Sagi, E. R. Hansen, K. Hose, G. Y. Lip, T. B. Larsen, F. Skjøth, Towards assigning diagnosis codes using medication history, in: International Conference on Artificial Intelligence in Medicine, Springer, 2020, pp. 203–213. doi:10.1007/978-3-030-59137-3_19.

[30] E. R. Hansen, T. Sagi, K. Hose, G. Y. Lip, T. B. Larsen, F. Skjøth, Assigning diagnosis codes using medication history, Artificial Intelligence in Medicine (2022) 102307. doi:10.1016/j.artmed.2022.102307.