Journal of the American Medical Informatics Association, 30(3), 2023, 475–484
https://doi.org/10.1093/jamia/ocac248
Advance Access Publication Date: 20 December 2022
Research and Applications

Research and Applications

A deep learning approach to identify missing *is-a* relations in SNOMED CT

Rashmie Abeysinghe[1], Fengbo Zheng[2], Elmer V. Bernstam[2,3], Jay Shi[4], Olivier Bodenreider[5], and Licong Cui[2]

[1]Department of Neurology, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, Texas, USA, [2]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, Texas, USA, [3]Division of General Internal Medicine, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, Texas, USA, [4]Intermountain Healthcare, Denver, Colorado, USA and [5]National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

Corresponding Author: Licong Cui, PhD, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, 7000 Fannin Street, Suite 600, Houston, TX 77030, USA; licong.cui@uth.tmc.edu

Received 27 September 2022; Revised 2 November 2022; Editorial Decision 1 December 2022; Accepted 7 December 2022

ABSTRACT

Objective: SNOMED CT is the largest clinical terminology worldwide. Quality assurance of SNOMED CT is of utmost importance to ensure that it provides accurate domain knowledge to various SNOMED CT-based applications. In this work, we introduce a deep learning-based approach to uncover missing *is-a* relations in SNOMED CT.
Materials and Methods: Our focus is to identify missing *is-a* relations between concept-pairs exhibiting a containment pattern (ie, the set of words of one concept being a proper subset of that of the other concept). We use hierarchically related containment concept-pairs as positive instances and hierarchically unrelated containment concept-pairs as negative instances to train a model predicting whether an *is-a* relation exists between 2 concepts with containment pattern. The model is a binary classifier leveraging concept name features, hierarchical features, enriched lexical attribute features, and logical definition features. We introduce a cross-validation inspired approach to identify missing *is-a* relations among all hierarchically unrelated containment concept-pairs.
Results: We trained and applied our model on the *Clinical finding* subhierarchy of SNOMED CT (September 2019 US edition). Our model (based on the validation sets) achieved a precision of 0.8164, recall of 0.8397, and F1 score of 0.8279. Applying the model to predict actual missing *is-a* relations, we obtained a total of 1661 potential candidates. Domain experts performed evaluation on randomly selected 230 samples and verified that 192 (83.48%) are valid.
Conclusions: The results showed that our deep learning approach is effective in uncovering missing *is-a* relations between containment concept-pairs in SNOMED CT.

Key words: ontology quality assurance, SNOMED CT, deep learning, missing *is-a* relations

INTRODUCTION

Biomedical ontologies or terminologies like SNOMED CT are widely used in biomedical research. They provide controlled vocabularies for annotating biomedical data, facilitating access to information, and supporting semantic interoperability.[1] Quality assurance of biomedical ontologies is a vital aspect of ontology management to ensure accurate representation of domain knowledge. However, this can be an arduous task due to the large size and complexity of ontologies making manual auditing unsustainable.[2] Therefore, automated approaches have been widely investigated to perform terminology quality assurance.

In this article, we introduce such an automated approach based on deep learning to identify missing *is-a* relations between containment concept-pairs in SNOMED CT. Leveraging hierarchically related and unrelated containment concept-pairs, we train a model to predict whether there is an *is-a* relation between a pair of concepts exhibiting a containment lexical pattern. The trained model is used to predict missing *is-a* relations among hierarchically unrelated containment concept-pairs. Randomly selected missing *is-a* predictions are evaluated by domain experts to confirm their validity.

BACKGROUND AND SIGNIFICANCE

SNOMED CT

SNOMED CT is the world's largest, most comprehensive clinical terminology with scientifically validated clinical content.[3–5] The September 2019 US edition of SNOMED CT contains over 350 000 concepts. SNOMED CT concepts are organized in 19 top-level subhierarchies including *Clinical finding*, *Procedure*, and *Body structure*. Each concept has a unique identifier, a fully specified name, a preferred name, synonyms, and definitional relations with other concepts.[6]

Quality assurance of SNOMED CT

Automated and semi-automated methods have been proposed for quality assurance or auditing of biomedical terminologies including SNOMED CT.[7] For example, summarization graphs called abstraction networks indicating various errors have been used to audit SNOMED CT.[8–11] Bodenreider introduced a method to identify missing *is-a* relations by comparing the SNOMED CT hierarchy against a hierarchy generated through reasoning on logical definitions derived from lexical features of concept labels.[12] Agrawal et al[13] leveraged lexical and structural indicators to detect inconsistent modeling among similar concepts. In previous work, we investigated lexical patterns in nonlattice subgraphs to identify missing *is-a* relations and missing concepts in SNOMED CT.[2] We then studied using enriched lexical attributes of concepts to uncover *is-a* inconsistencies in SNOMED CT.[14,15] All the above-mentioned methods are rule-based.

Recently, a few studies have leveraged machine learning techniques for auditing SNOMED CT. In one study, we explored 3 deep learning models to predict proper names of SNOMED CT concepts complying with the terminology's naming convention.[16] Other works have leveraged deep learning to place new concepts in the SNOMED CT hierarchy given that the new concept's name is known.[17–19] We also investigated whether deep learning could aid in automatically validating the suggested missing *is-a* relations in SNOMED CT obtained by nonlattice-based auditing approaches.[20] To the best of our knowledge, no other work has used deep learning techniques to detect actual missing *is-a* relations in SNOMED CT. Although Liu et al[21] have investigated a convolutional neural network (CNN) model to predict missing *is-a* relations in the National Cancer Institute thesaurus (NCIt), the results were not yet strong enough when their trained model was used to detect actual missing *is-a* relations (only 1 out of 20 random suggestions were valid). We hypothesize that deep learning approaches can be utilized to effectively detect missing *is-a* relations in biomedical terminologies such as SNOMED CT.

Specific contribution

The specific contribution of this work is to combine concepts' name features, hierarchical features, ancestor-related lexical features, and logical definition features, and train a deep learning model to automatically identify missing *is-a* relations between concept-pairs with a containment pattern in SNOMED CT. A pair of concepts (or concept-pair) is said to have a containment pattern if the set of words of one concept's name is a proper subset of that of the other concept's name (eg, concept *Chronic osteomyelitis* and concept *Chronic osteomyelitis of right ankle*). Due to the discovery nature of identifying missing *is-a* relations for quality assurance, it is important to assess the model's ability to uncover valid missing *is-a* relations. Therefore, manual evaluation by domain experts is performed on the missing *is-a* relations predicted by our model. Notably, this learning-based approach identifies missing *is-a* relations that were not discovered by previous rule-based approaches. To the best of our knowledge, this is the first work on automated identification of missing *is-a* relations in SNOMED CT using deep learning techniques.

MATERIALS AND METHODS

In this work, we focus on identifying missing *is-a* relations in the *Clinical finding* subhierarchy, the largest subhierarchy of SNOMED CT (September 2019 US edition). More specifically, we focus on relations between concepts exhibiting a containment pattern, because this pattern is often indicative of hierarchical relations. We used hierarchically related containment concept-pairs as positive instances and hierarchically unrelated containment concept-pairs as negative instances to train a deep learning model predicting whether there is an *is-a* relation between 2 concepts with containment pattern. This model combines 4 kinds of concept features: concept name features, hierarchical features, enriched lexical attribute features, and logical definition features. Since our goal is to identify missing *is-a* relations, we applied the trained model to identify such relations from hierarchically unrelated containment concept-pairs.

Containment concept-pair generation

Hierarchically unrelated concept-pairs exhibiting a containment pattern may represent missing *is-a* relations.[2] In other words, given 2 hierarchically unrelated concepts C_1 and C_2, if the set of words of C_1 is a proper subset of that of C_2, then there might be a missing *is-a* relation between C_2 and C_1 (ie, C_2 *is-a* C_1). Therefore, we targeted such unrelated containment concept-pairs to identify missing *is-a* relations.

In subsequent text, for a pair of concepts with containment pattern, we refer to the concept whose set of words subsumes that of the other as the container concept, and the other as the containee concept. For example, for concept *Open fracture of metaphysis of first metatarsal bone (disorder)* and concept *Open fracture of metaphysis of metatarsal bone (disorder)*, we refer to the former as the container concept and the latter as the containee concept.

To train a model for *is-a* prediction, we generated both hierarchically related concept-pairs with containment pattern (considered as positive instances) and unrelated concept-pairs with containment pattern (considered as negative instances). For example, concept *Open fracture of metaphysis of first metatarsal bone (disorder)* and concept *Open fracture of metaphysis of metatarsal bone (disorder)* are hierarchically related (ie, having an *is-a* relation) and considered as a positive instance; while concept *Acute gastric ulcer with hemorrhage but without obstruction (disorder)* and concept *Acute gastric ulcer with obstruction* are hierarchically unrelated and considered as a negative instance.

There were 10 820 unrelated concept-pairs and 24 407 hierarchically related concept-pairs with containment pattern in the *Clinical finding* subhierarchy of SNOMED CT. However, in a particular situation where the container concept's name includes "*and*," "*or*," or "*and/or*," it is highly likely that the containment concept-pair would not suggest a valid missing *is-a* relation. For example, since the set of words of concept *Contusion of neck (disorder)* is a proper subset of that of concept *Contusion of head and/or neck (disorder)*, the containment concept-pair heuristic would wrongly suggest *Contusion of head and/or neck (disorder)* is a subtype of *Contusion of neck (disorder)*, while the valid case is that *Contusion of neck (disorder)* is a subtype of *Contusion of head and/or neck (disorder)*. Therefore, we disregard those cases except in the following 2 scenarios:

1. if the containee concept also includes "*and*," "*or*," or "*and/or*," and the containee concept has the same number of {"*and*," "*or*," "*and/or*"} words as the container concept, for example, containee *Carcinoma liver and/or biliary system (disorder)* and container *Carcinoma in situ of liver and/or biliary system (disorder)*;
2. if the containee concept appears as a substring of the container concept, and the word immediately following the substring is not any of {"*and*," "*or*," "*and/or*"}, for example, containee *Neuropathic ulcer (disorder)* and container *Neuropathic ulcer of midfoot AND/OR heel due to type 2 diabetes mellitus (disorder)*.

Disregarding such cases resulted in a total of 8804 unrelated containment concept-pairs that were used as negative instances. A randomly selected 8804 related containment concept-pairs were used as positive instances. Therefore, our dataset included 17 608 instances of containment concept-pairs.

Concept feature representation

We leverage several features of SNOMED CT concepts in our model, including lexical features (ie, concept name), structural features (hierarchical relations), and logical features (logical definitions). While all features are extracted from the concept under investigation, lexical features can also be extracted from its ancestors (see *Enriched lexical attribute features*).

Concept name features

The Fully Specified Name (FSN) of a concept in SNOMED CT is a unique description that encodes the meaning of the concept in an unambiguous manner.[22] To use FSNs as features in our model, we vectorized them by training a Doc2vec embedding model using Gensim Python library.[23] Doc2Vec is an unsupervised framework that learns fixed-length feature representations from variable-length pieces of text.[24] We set the size of the embeddings of FSNs to 100 and window size to 5, and trained using the distributed bag-of-words model of Doc2vec.

Hierarchical features

The hierarchical *is-a* or subsumption relationship is formed between 2 concepts such that one concept (child) is a subtype of the other (parent). In this work, we leveraged a Graph Convolutional Network (GCN) to automatically learn hierarchical features of a concept. GCN learns structural features for each graph node by performing neighborhood aggregation.[25,26]

To learn hierarchical features of a concept with GCNs, we extracted an ancestor subgraph containing parents and grandparents of the concept and their relations. We generated ancestor subgraphs for concepts in both positive and negative instances. Figures 1 and 2A show the ancestor subgraphs obtained for a positive instance with container concept *Open fracture of metaphysis of first metatarsal bone (disorder)* and containee concept *Open fracture of metaphysis of metatarsal bone (disorder)*, respectively. Figures 3 and 4A display the ancestor subgraphs obtained for a negative instance with container concept *Acute gastric ulcer with hemorrhage but without obstruction (disorder)* and containee concept *Acute gastric ulcer with obstruction (disorder)*, respectively.

In a positive instance, the ancestor subgraph of the container concept includes the containee concept and its parent(s), and there is an *is-a* link between the container concept and the containee concept; while in a negative instance, the ancestor subgraph of the container concept may not include the containee concept and its parent(s), and there is no *is-a* link between the container concept and the containee concept. To make the container concept's subgraphs include similar information for positive and negative instances, we performed the following modifications. For positive instances, we modified the ancestor subgraph of the container concept by removing the *is-a* relation between the container and containee as shown in Figure 2B. For negative instances, we added the containee concept and its parent(s) to the ancestor subgraph of the container concept as displayed in Figure 4B. These steps were necessary to ensure that positive and negative instances have the same link existence information and the model will not optimize on this part of information to classify.

Note that in some positive instances, the container does not have any additional *is-a* relations other than the one between the container and the containee. In such instances, removing the *is-a* relation between the 2 would produce an ancestor subgraph without any ancestors leading to no hierarchical features being learned. We did not consider such instances for training.

Enriched lexical attribute features

We previously showed that enriched lexical attributes of concepts are helpful for identifying missing *is-a* relations.[14,15,27] Given a concept, such attributes are obtained from leveraging words and noun phrases in the FSNs of the concept and its ancestors. To vectorize these attributes, we first constructed documents for each concept by creating a sentence from each attribute of the concept preceded by the words "has attribute." Then we trained a distributed bag-of-words Doc2vec model with such documents created for all the concepts. We set the size of embeddings in this step to 200 and window size to 3.

Logical definition features

In SNOMED CT, logical definitions provide a structured representation for concepts based on their clinical meanings, descriptions, and how they are related to other concepts.[28,29] For instance, Figure 5 shows the logical definition of the concept *Aneurysm of conjunctiva (disorder)*. With such information, SNOMED CT can be considered as a knowledge graph with a collection of facts in the form of triples consisting of a head, a relation, and a tail.[30] We used a knowledge graph embedding scheme called TransE to generate vector representations for the elements of the knowledge graph.[31,32] We set the size of

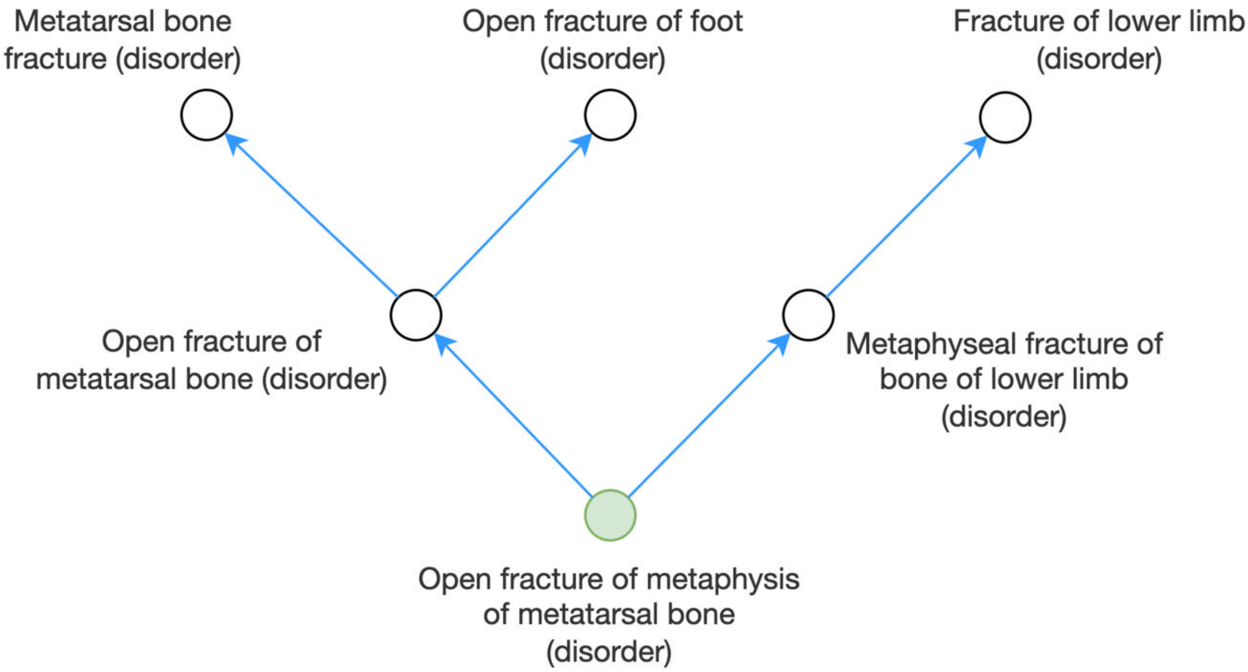

Figure 1. Ancestor subgraph of concept *Open fracture of metaphysis of metatarsal bone (disorder).*

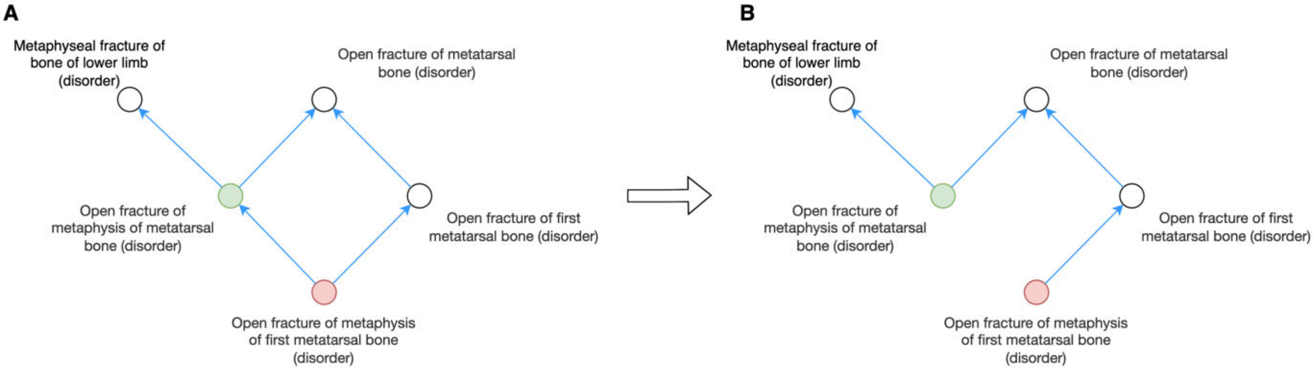

Figure 2. (A) Ancestor subgraph of a positive instance's container concept *Open fracture of metaphysis of first metatarsal bone (disorder).* (B) Modified ancestor subgraph with the link between the container concept and the containee concept *Open fracture of metaphysis of metatarsal bone (disorder)* removed.

embeddings to 400 and learning rate to 0.001, and used 400 negative instances per positive instance to train the TransE model.

Architecture of the model

Figure 6 illustrates the architecture of our model. It consists of 2 GCN layers that separately learn hierarchical features for container and containee concepts (both layers share the same weights), followed by 3 fully connected (FC) layers that perform the classification. The final FC layer uses a sigmoid activation function, and other layers use ReLU activation functions. The output of the model is the probability of the concept-pair having an *is-a* relation. Batch normalization is also performed and dropouts are used after the first 2 FC layers so that the model does not overfit the data. Table 1 summarizes the hyperparameters used for our model. We used the default values for embedding size and window size in Gensim Doc2vec for concept name feature embeddings. Since the documents created for enriched lexical attributes are usually much larger than concept names, we increased the embedding size of these to 200. Since a majority of sentences in these documents are 3-word sentences, we set the window size to 3. The network layer sizes, dropouts, learning rate, and batch size were selected through manual hyperparameter tuning.

Given a concept-pair instance, the nodes in the ancestor subgraphs of both container and containee concepts are assigned with Doc2vec embeddings of the FSNs of the relevant concepts as their initial features. Then, both ancestor subgraphs are passed through the corresponding GCN layers so that hierarchical features can be learned. Next the learned hierarchical features of container and containee concepts are multiplied with each other. Similarly, the concept name features, enriched lexical attribute features, and logical definition features of container and containee concepts are multiplied, respectively. Then the resulting vectors of these multiplications are concatenated and forwarded through FC layers to perform the classification.

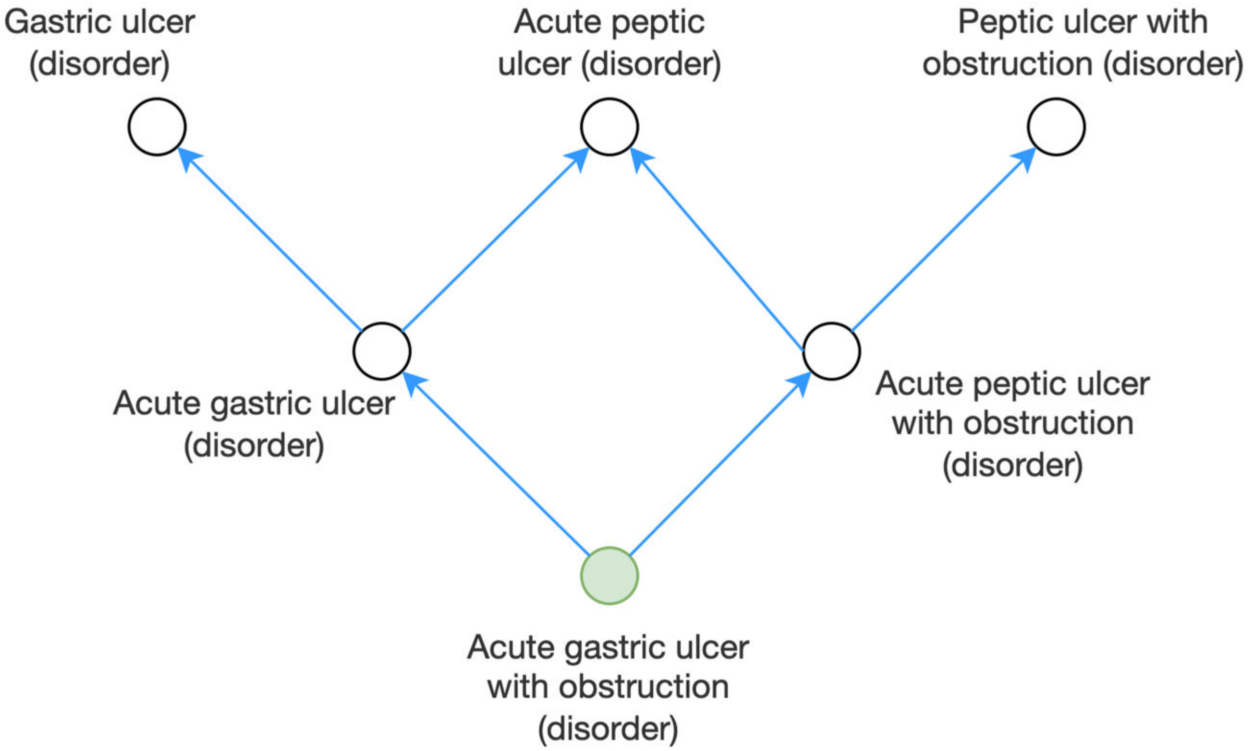

Figure 3. Ancestor subgraph of a negative instance's containee concept *Acute gastric ulcer with obstruction (disorder)*.

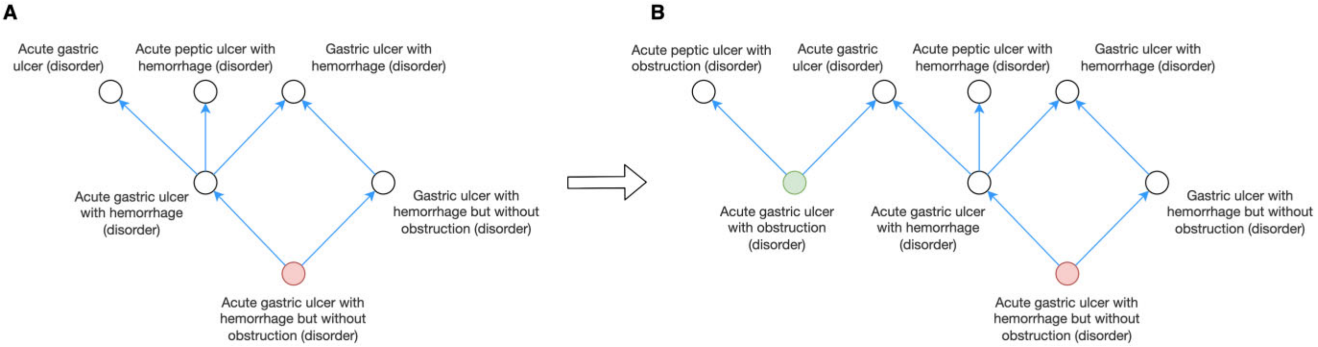

Figure 4. (A) Ancestor subgraph of a negative instance's container concept *Acute gastric ulcer with hemorrhage but without obstruction (disorder)*. (B) Modified ancestor subgraph with containee concept *Acute gastric ulcer with obstruction (disorder)* and its parent(s) added.

Model implementation

We implemented our model with PyTorch using the Graph Neural Network package in the Deep Graph Library.[33] We used Binary Cross Entropy as the loss function and Adamax as the optimizer. A learning rate of 0.0001 and a batch size of 512 was used. The training was performed for 20 epochs. The model was trained on an NVIDIA GeForce RTX 2080 Ti graphics card.

A cross-validation approach for predicting missing *is-a* relations

As mentioned earlier, hierarchically unrelated containment concept-pairs not only serve as negative instances to train our model, but also the candidates of potentially missing *is-a* relations. To identify missing *is-a* relations from the entire set of 8804 unrelated containment concept-pairs, we introduce an approach derived from cross-validation. More specifically, we use a subset of these instances to train the model while leaving the rest to identify missing *is-a* relations. We randomly divided negative instances to 10-folds (N_1, N_2, N_3, …, N_{10}) and positive instances into 10-folds (P_1, P_2, P_3, …, P_{10}) respectively. In each training round, 8 negative folds together with 8 positive folds were used for training, while a single negative fold and a single positive fold were used for validation. After training, the trained model was applied to the remaining negative fold to predict missing *is-a* relations. For example, in the first round, we may use N_3, N_4, N_5, N_6, N_7, N_8, N_9, N_{10}, P_3, P_4, P_5, P_6, P_7, P_8, P_9, and P_{10} folds for training, N_2 and P_2 for validation, and N_1 for predicting missing *is-a* relations. Note that concept-pairs are not shared between training, validation, and prediction sets in any round. This approach would allow us to identify missing *is-a* relations among the entire 8804 unrelated concept-pairs with a containment pattern.

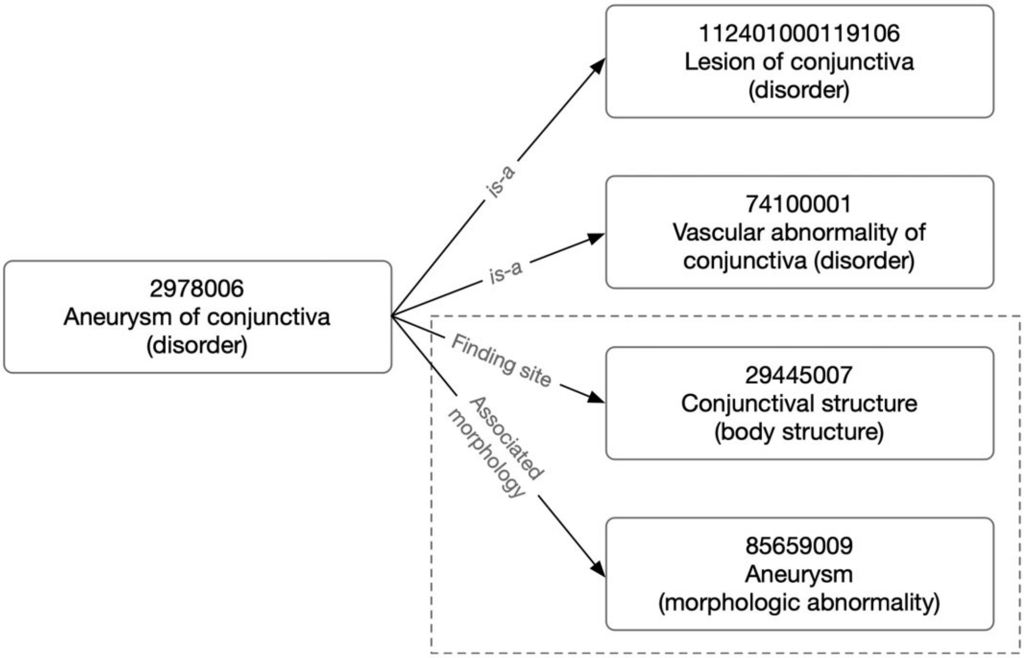

Figure 5. The logical definition of the concept *Aneurysm of conjunctiva (disorder)*. The "Finding site" and "Associated morphology" relationships in this logical definition form an attribute group.[35]

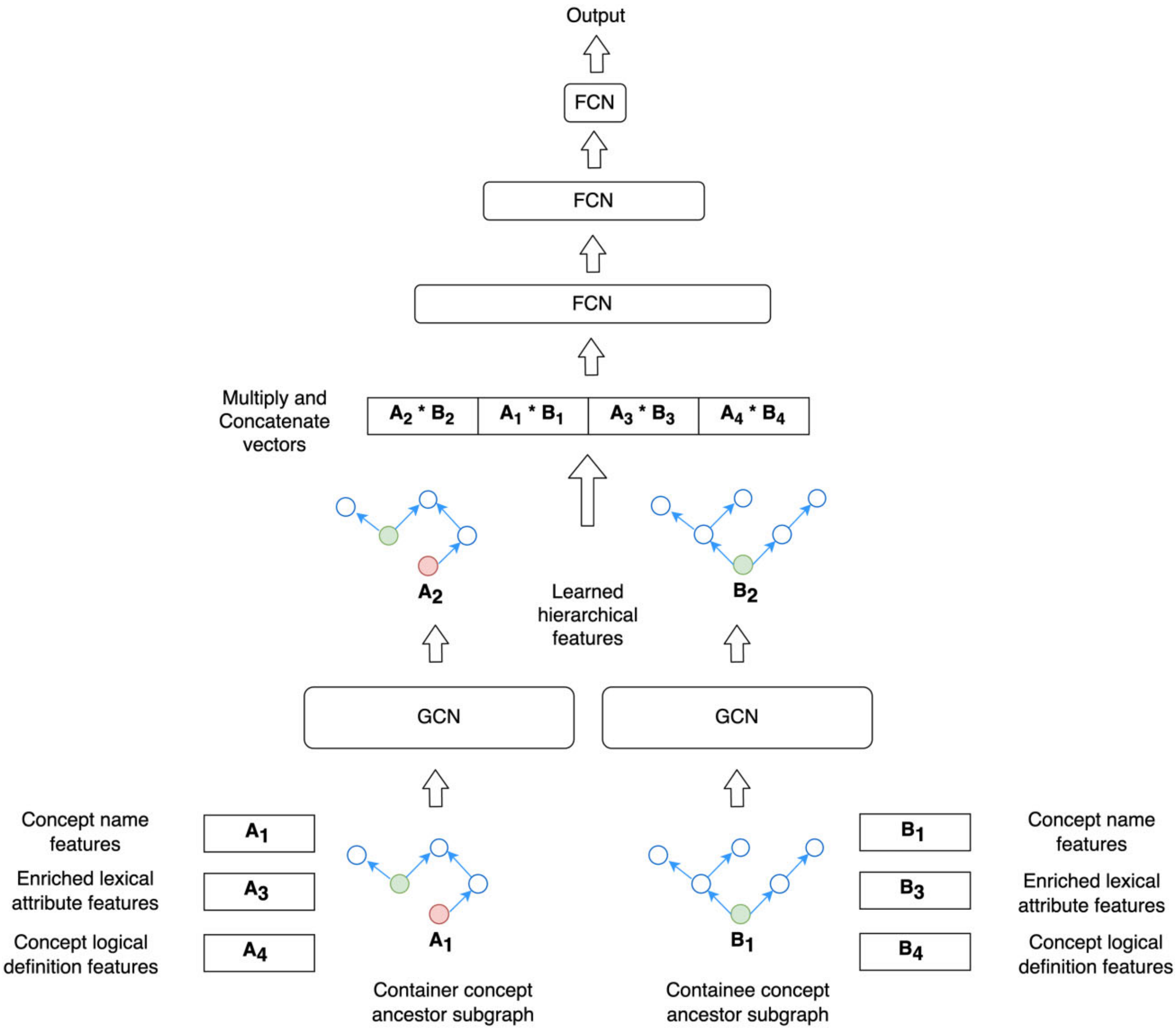

Figure 6. The architecture of our model. The 2 GCN layers share the same weights.

Table 1. Model hyperparameters

Parameter	Value
Doc2vec vector size (concept names)	100
Doc2vec vector size (enriched lexical attributes)	200
Knowledge graph embedding vector size	400
GCN layer size	1700
Fully connected layer 1 (FC1) size	1000
Fully connected layer 2 (FC2) size	32
Fully connected layer 3 (FC3) size	1
Activation for FC1 and FC2	ReLU
Activation for FC3	Sigmoid
Dropouts for FC1 and FC2	0.6
Optimizer	Adamax
Learning rate	0.0001
Loss function	Binary cross entropy
Batch size	512
Epochs	20

Ablation studies

Ablation studies were performed to investigate the respective effectiveness and utility of each of the 4 input features used in our model: hierarchical features (H), concept name features (CN), enriched lexical attribute features (ELA), and concept logical definition features (CD). In each ablation study, we removed one type of input features and performed the cross-validation approach on the training and validation folds.

Evaluation

To evaluate our model's performance in uncovering actual missing *is-a* relations in the *Clinical finding* subhierarchy of SNOMED CT, we selected a random collection of predicted missing *is-a* samples for expert review and verification. Two domain experts (authors EVB and JS) independently performed the evaluation and provided comments. We considered a sample as a valid missing *is-a* relation only if both experts agreed.

As it is labor-intensive for domain experts to manually discover missing *is-a* relations from a large collection of unrelated containment concept-pairs, it is infeasible to create a gold standard that can be used to comprehensively assess the precision, recall, and F1 score of the model. Therefore, we only evaluated whether the randomly chosen samples predicted by the model were valid missing *is-a* relations, that is, only the precision was assessed. And it is worth noting that given the discovery nature of missing *is-a* identification, any valid *is-a* relations uncovered are useful.

For the remaining unrelated containment concept-pairs that were not predicted by our model as potentially missing *is-a* relations, some of them may still be valid missing *is-a* relations. Therefore, we randomly picked another collection of them for experts' further evaluation.

RESULTS

Ablation studies

The average precision, recall, and F1 score of each ablation experiment can be found in Table 2.

It can be seen that the H + CN + ELA + CD model achieved the best F1 score, while H + CN + CD had the best precision and H + CN + ELA model had the best recall. Comparing F1 scores, it is evident that the introduction of hierarchical features improves the model performance considerably.

Table 2. Ablation study

Model	Average precision	Average recall	Average F1 score
CN+ELA+CD	0.7597	0.8133	0.7855
H+CN+CD	**0.8339**	0.7973	0.815
H+CN+ELA	0.7678	**0.8789**	0.8196
H+ELA+CD	0.7964	0.846	0.8204
H+CN+ELA+CD	0.8164	0.8397	**0.8279**

H: hierarchical features; CN: concept name features; ELA: enriched lexical attribute features; CD: concept logical definition features. The best performance is marked in bold.

Missing *is-a* relation identification

Using the best performing model in terms of F1 score (H + CN+ ELA + CD model), a total of 1661 potentially missing *is-a* relations were identified in the *Clinical finding* subhierarchy. Table 3 presents 10 valid missing *is-a* relations identified by our model. For instance, *Residual interatrial communication following procedure (disorder)* is a valid subtype of *Residual interatrial communication (disorder)*.

Evaluation

Out of 1661 potentially missing *is-a* relations predicted by our model, we randomly selected 230 samples for evaluation (see Supplementary Appendix SI). One domain expert (EVB) confirmed that 199 cases are valid missing *is-a* relations, while the other domain expert (JS) confirmed 208 cases are valid. The 2 domain experts agreed on 192 cases out of 230 as valid, indicating a precision of 83.48% achieved by our model. In addition, both experts agreed on 15 cases as invalid missing *is-a* relations. This means that the 2 experts agreed on a total of 207 (90%) out of 230 cases that they evaluated. The prevalence-adjusted bias-adjusted kappa (PABAK) coefficient for the inter-rater agreement was 0.8.[34] Additionally, the PABAK between the model and expert EVB was 0.7304, and the PABAK between the model and expert JS was 0.8087. Thus, the model differed from the experts approximately as much as the experts differed from each other.

As for the remaining 7143 ones that were not predicted by our model as potentially missing *is-a* relations, we randomly picked 50 samples from them for further evaluation (see Supplementary Appendix SII). The 2 experts agreed on 24 samples as valid missing *is-a* relations (a precision of 48%).

DISCUSSION

Model performance

It is worth noting that domain expert evaluation showed a slightly higher precision (83.48%) than the best model in ablation studies (81.64%). This is because the potentially missing *is-a* relations identified in this work originate from hierarchically unrelated containment concept-pairs, which were also leveraged as negative instances during the training process. Our assumption was that a minority of these cases would be valid missing *is-a* relations (ie, they are actually positive instances) and hence using them as negative instances during training would not have a major impact on the model. However, existence of such valid missing *is-a* cases in the validation set would result in reduced performance metrics (Table 2) than the actual performance. For instance, even if the model correctly identifies a valid missing *is-a* case in the validation set (ie, true positive), since it is a negative instance in the validation set, it is counted as a false positive

Table 3. Ten examples of valid missing *is-a* relations that were identified by our model and confirmed by domain experts

Subtype concept	Supertype concept
735579001: Residual interatrial communication following procedure (disorder)	449351005: Residual interatrial communication (disorder)
2978006: Aneurysm of conjunctiva (disorder)	432119003: Aneurysm (disorder)
92542002: Carcinoma *in situ* of ascending colon (disorder)	269533000: Carcinoma of colon (disorder)
86731008: Multiple open fractures of pelvis with disruption of pelvic circle (disorder)	263222005: Multiple pelvic fractures (disorder)
1088171000119103: Recurrent acute serous otitis media of left middle ear (disorder)	194287004: Recurrent acute otitis media (disorder)
724590008: Benign osteogenic neoplasm of articular cartilage of clavicle (disorder)	92061008: Benign neoplasm of clavicle (disorder)
26171000119109: Pyogenic bacterial arthritis of wrist (disorder)	24761000119107: Pyogenic bacterial arthritis (disorder)
232069006: Retinal pigment epithelial detachment with tear of retinal pigment epithelium (disorder)	95690009: Retinal tear (disorder)
403496009: Discoid lupus erythematosus of genital mucous membranes (disorder)	403494007: Discoid lupus erythematosus of mucous membranes (disorder)
402270001: Irritant contact dermatitis of hands caused by friction (disorder)	735758003: Irritant contact dermatitis caused by friction (disorder)

and hence the reported precision of the model is lowered. Nevertheless, this only affects the performance scores for ablation studies. The manual evaluation is unaffected as it is performed by domain experts on a random collection of missing *is-a* relations.

Out of 8804 hierarchically unrelated concept-pairs with containment pattern, our deep learning-based approach suggested 1661 potentially missing *is-a* relations with a precision of 83.48% based on the expert evaluation of 230 samples. For the remaining 7143 ones, expert evaluation of 50 samples revealed a precision of 48%. This indicates that our deep learning-based approach built on top of the simple containment baseline is able to more precisely identify missing *is-a* relations compared to using the simple containment baseline alone. However, it also shows that the deep learning-based approach misses some valid cases of *is-a* relations (ie, lower recall compared with the simple baseline as expected).

Supporting evidence from SNOMED CT evolution

In this work, we used the September 2019 US edition of SNOMED CT, after which several new editions have been released. Such terminology evolution enables us to find additional supporting evidence of potentially missing *is-a* relations identified by our model. Out of 1661 missing *is-a* relations predicted by our model, 513 have already been fixed in the September 2022 US edition of SNOMED CT. For instance, our model predicted *Acute bronchitis caused by chemical fumes (disorder)* should be a subtype of *Acute chemical bronchitis (disorder)*, and this has been reflected in the newer edition.

In addition, at least one of the concepts in 150 missing *is-a* predictions have been removed in the newer edition. For example, our model predicted that *Inflammatory hereditary disorder (disorder)* should be a subtype of *Inflammatory disorder (disorder)*. However, *Inflammatory hereditary disorder (disorder)* is no longer an active concept in the newer edition.

Distinction with related work

In Liu et al's[21] work using CNNs to predict missing *is-a* relations in the Neoplasm subhierarchy of NCIt, it was concluded that the model performance was not yet sufficient for detecting missing *is-a* relations. In comparison, our model achieved much better performance and was applied to a different and larger terminology: SNOMED CT.

The containment lexical pattern was originally introduced in our previous work where we considered only hierarchically unrelated concept-pairs within lower and upper bounds of nonlattice subgraphs in SNOMED CT.[2] In contrast, our deep learning-based model is able to explore all the hierarchically unrelated containment concept-pairs. Out of 1606 potentially missing *is-a* relations identified in this work, 1208 (75%) were not captured in the upper/lower bounds of nonlattice subgraphs.

Examples of false positive cases

Although the model predicted valid cases of missing *is-a* among containment concept-pairs, there were a few cases of invalid suggestions made by the model. For instance, the model predicted that *Acute gastric ulcer with hemorrhage but without obstruction (disorder)* should be a subtype of *Acute gastric ulcer with obstruction (disorder)*. This is invalid as the former concept is regarding "without obstruction" while the latter is "with obstruction." Similarly, our model predicted that *Single coronary artery fistula (disorder)* should be a subtype of *Single coronary artery (disorder)*. This is invalid as *coronary artery fistula* and *coronary artery* are distinct entities.

Limitations and future work

As mentioned earlier, some of the negative instances in the training data could be valid missing *is-a* relations, posing a unique challenge for learning-based quality assurance approaches like this work. It would be interesting to explore whether manual curation of a subset of the negative training data by removing valid *is-a* relations would affect the model performance. In addition, we balanced our dataset by down-sampling the positive samples. A future direction is to perform resampling with different ratios of positive to negative samples and find out which ratio provides optimum performance for this particular task of identifying missing *is-a* relations.

In this work, we have focused on investigating concept-pairs with containment pattern. A more challenging task for future work is to predict missing *is-a* relations among any concept-pair without restricting to containment pattern. The major hurdle is how to select negative instances. Using full set of unconnected pairs would yield a heavily imbalanced training dataset. Random generation of concept-pairs as negative instances may not be effective, since this may lead to distant concept-pairs in the terminology, while missing *is-a* relations are observed among close concepts.

Additionally, our approach was only applied to the largest subhierarchy of SNOMED CT (*Clinical finding*). A major challenge in applying this approach to other subhierarchies is the smaller number of training instances available. In the future, we will explore the effectiveness of predicting missing *is-a* relations in a subhierarchy using a model trained with instances from some other subhierarchies. We also plan to apply our approach to other biomedical terminologies such as NCIt.

For the valid missing *is-a* relations identified in this work, we will submit them to SNOMED International for consideration of incorporation into a new release of the SNOMED CT.

CONCLUSION

In this article, we introduced a deep learning approach to predict missing *is-a* relations among hierarchically unrelated concept-pairs that exhibit a containment lexical pattern. This approach identified 1661 potentially missing *is-a* relations in the *Clinical finding* subhierarchy of SNOMED CT (September 2019 US edition). Domain experts' evaluation of 230 random samples showed the effectiveness of this approach (a precision of 83.48%). Additional supporting evidence showed that 513 potentially missing *is-a* relations predicted by our approach have been reflected in the newer edition.

FUNDING

This work was supported by the National Science Foundation (NSF) through grant 2047001, National Institutes of Health (NIH) National Library of Medicine through grant R01LM013335, and National Institute of Neurological Disorders and Stroke through grant R01NS116287 received by LC. This work was also supported in part by the National Center for Advancing Translational Sciences (NCATS) under awards UL1TR000371 and U01TR002393; the Cancer Prevention and Research Institute of Texas (CPRIT), under award RP170668 and the Reynolds and Reynolds Professorship in Clinical Informatics. This work was also supported in part by the Intramural Research Program of the NIH, National Library of Medicine and by an appointment to the National Library of Medicine Research Participation Program administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the National Library of Medicine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NSF.

AUTHOR CONTRIBUTIONS

LC, OB, and RA conceptualized and designed this study. RA developed the deep learning pipeline, performed the ablation study experiments, generated the results, and prepared the evaluation samples. FZ constructed the set of enriched lexical attributes that were used as features. EVB and JS performed the evaluation. RA and LC analyzed the evaluation results. RA and LC wrote the manuscript with input from EVB and OB.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

We sincerely thank Dr Xiaoqian Jiang at UTHealth School of Biomedical Informatics for providing the access to NVIDIA GPUs used in this work. This work also utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov) and we thank Dr Vinh Nguyen for her computational support.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

The data underlying this article are described in the article and in its online Supplementary Material.

REFERENCES

1. Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb Med Inform* 2008; 17 (01): 67–79.
2. Cui L, Zhu W, Tao S, *et al.* Mining non-lattice subgraphs for detecting missing hierarchical relations and concepts in SNOMED CT. *J Am Med Inform Assoc* 2017; 24 (4): 788–98.
3. Bodenreider O, Cornet R, Vreeman DJ. Recent developments in clinical terminologies—SNOMED CT, LOINC, and RxNorm. *Yearb Med Inform* 2018; 27 (1): 129–39.
4. Benson T, Grieve G. *Principles of Health Interoperability: SNOMED CT, HL7 and FHIR*. London, UK: Springer International Publishing; 2016.
5. SNOMED—5-Step Briefing. https://www.snomed.org/snomed-ct/five-step-briefing. Accessed December 15, 2020.
6. Definitions—SNOMED CT Editorial Guide. https://confluence.ihtsdotools.org/display/DOCEG/Definitions. Accessed December 15, 2020.
7. Zhu X, Fan J-W, Baorto DM, *et al.* A review of auditing methods applied to the content of controlled biomedical terminologies. *J Biomed Inform* 2009; 42 (3): 413–25.
8. Ochs C, Geller J, Perl Y, *et al.* A tribal abstraction network for SNOMED CT target hierarchies without attribute relationships. *J Am Med Inform Assoc* 2015; 22 (3): 628–39.
9. Wei D, Bodenreider O. Using the abstraction network in complement to description logics for quality assurance in biomedical terminologies—a case study in SNOMED CT. *Stud Health Technol Inform* 2010; 160 (Pt 2): 1070–4.
10. Ochs C, Geller J, Perl Y, *et al.* Scalable quality assurance for large SNOMED CT hierarchies using subject-based subtaxonomies. *J Am Med Inform Assoc* 2015; 22 (3): 507–18.
11. Ochs C, Perl Y, Geller J, *et al.* Scalability of abstraction-network-based quality assurance to large SNOMED hierarchies. *AMIA Annu Symp Proc* 2013; 2013: 1071–80.
12. Bodenreider O. Identifying missing hierarchical relations in SNOMED CT from logical definitions based on the lexical features of concept names. In: *6th International Conference on Biomedical Ontology (ICBO 2016) Proc*; 2016; Corvallis, OR. http://ceur-ws.org/Vol-1747/IT601_ICBO2016.pdf. Accessed October 21, 2022.
13. Agrawal A, Perl Y, Ochs C, *et al.* Algorithmic detection of inconsistent modeling among SNOMED CT concepts by combining lexical and structural indicators. *IEEE Int Conf Bioinformatics Biomed* 2015; 2015: 476–83.

14. Cui L, Bodenreider O, Shi J, *et al.* Auditing SNOMED CT hierarchical relations based on lexical features of concepts in non-lattice subgraphs. *J Biomed Inform* 2018; 78: 177–84.
15. Zheng F, Shi J, Cui L. A lexical-based approach for exhaustive detection of missing hierarchical IS-A relations in SNOMED CT. *AMIA Annu Symp Proc* 2020; 2020: 1392–401.
16. Zheng F, Cui L. Exploring deep learning-based approaches for predicting concept names in SNOMED CT. *IEEE Int Conf Bioinformatics Biomed* 2018; 2018: 808–13.
17. Liu H, Geller J, Halper M, *et al.* Using convolutional neural networks to support insertion of new concepts into SNOMED CT. *AMIA Annu Symp Proc* 2018; 2018: 750–9.
18. Zheng L, Liu H, Perl Y, *et al.* Training a convolutional neural network with terminology summarization data improves SNOMED CT enrichment. *AMIA Annu Symp Proc* 2019; 2019: 972–81.
19. Liu H, Perl Y, Geller J. Transfer learning from BERT to support insertion of new concepts into SNOMED CT. *AMIA Annu Symp Proc* 2019; 2019: 1129–38.
20. Sun Q, Zhang G-Q, Zhu W, *et al.* Validating auto-suggested changes for SNOMED CT in non-lattice subgraphs using relational machine learning. *Stud Health Technol Inform* 2019; 264: 378–82.
21. Liu H, Zheng L, Perl Y, *et al.* Can a convolutional neural network support auditing of NCI thesaurus neoplasm concepts? In: *International Conference on Biomedical Ontologies Proceedings*; 2018; Corvallis, OR.
22. Fully Specified Name. https://confluence.ihtsdotools.org/display/DOCEG/Fully+Specified+Name. Accessed May 4, 2021.
23. Gensim: Topic modelling for humans. https://radimrehurek.com/gensim/. Accessed October 21, 2022.
24. Le Q, Mikolov T. Distributed representations of sentences and documents. In: *International Conference on Machine Learning Proceedings*, Vol. 2014; 2014: 1188–96; Beijing, China.
25. Zhou J, Cui G, Hu S, *et al.* Graph neural networks: a review of methods and applications. *AI Open* 2020; 1: 57–81.
26. Wu Z, Pan S, Chen F, *et al.* A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst* 2021; 32 (1): 4–24.
27. Abeysinghe R, Brooks MA, Cui L. Leveraging non-lattice subgraphs to audit hierarchical relations in NCI thesaurus. *AMIA Annu Symp Proc* 2019; 2019: 982–91.
28. SNOMED CT Logical Model. https://confluence.ihtsdotools.org/display/DOCSTART/5.+SNOMED+CT+Logical+Model. Accessed May 6, 2021.
29. Nikiema JN, Jouhet V, Mougin F. Integrating cancer diagnosis terminologies based on logical definitions of SNOMED CT concepts. *J Biomed Inform* 2017; 74: 46–58.
30. Chang D, Balazevic I, Allen C, *et al.* Benchmark and best practices for biomedical knowledge graph embeddings. *Proc Conf Assoc Comput Linguist Meet* 2020; 2020: 167–76.
31. Zheng D, Song X, Ma C, *et al.* DGL-KE: Training knowledge graph embeddings at scale. In: *43rd International ACM SIGIR Conference on Research and Development in Information Retrieval Proceedings*, Virtual Event; 2020: 739–48; China.
32. Bordes A, Usunier N, Garcia-Duran A, *et al.* Translating embeddings for modeling multi-relational data. *Adv Neural Inf Process Syst* 2013; 26: 2787–95.
33. Wang M, Zheng D, Ye Z, *et al.* Deep graph library: A graph-centric, highly-performant package for graph neural networks. *arXiv:1909.01315* 2019.
34. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993; 46 (5): 423–9.
35. SNOMED CT—Aneurysm of Conjunctiva (Disorder). https://browser.ihtsdotools.org/?perspective=full&conceptId1=2978006&edition=MAIN/SNOMEDCT-US/2022-09-01&release=&languages=en. Accessed October 20, 2022.