



## OPEN ACCESS

## EDITED BY

Bo Wang,  
Wuhan University, China

## REVIEWED BY

Ge Cao,  
Xi'an University of Technology, China  
Guangchen Liu,  
Inner Mongolia University of Technology, China

## \*CORRESPONDENCE

Jie Hu,  
✉ Jhon.Hu@outlook.com

RECEIVED 16 November 2023

ACCEPTED 28 December 2023

PUBLISHED 12 January 2024

## CITATION

Hu J, Xu G, Qi L and Qie X (2024), Knowledge reasoning in power grid infrastructure projects based on deep multi-view graph convolutional network.

*Front. Energy Res.* 11:1339416.

doi: 10.3389/fenrg.2023.1339416

## COPYRIGHT

© 2024 Hu, Xu, Qi and Qie. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Knowledge reasoning in power grid infrastructure projects based on deep multi-view graph convolutional network

Jie Hu<sup>1\*</sup>, Gang Xu<sup>1</sup>, Lizhong Qi<sup>2</sup> and Xin Qie<sup>1</sup>

<sup>1</sup>School of Electrical and Electronic Engineering, North China Electric Power University, Beijing, China,

<sup>2</sup>State Grid Economic and Technological Research Institute Co Ltd., Beijing, China

With the rapid development of power grid infrastructure, especially the increasing number of ultra-high voltage (UHV) projects, knowledge extracted from historical engineering data is collected and can be potentially used to assist in the review of power transmission and transformation projects. However, conventional knowledge modeling and knowledge reasoning methods cannot meet the current needs of power grid construction. In this paper, considering the more supernumerary and distinctive information brought by multi-view data which could be beneficial for feature representation and knowledge reasoning from the constructed knowledge base, a multi-view graph convolutional network (GCN) based on knowledge graph is proposed to make classification for power grid infrastructure projects. Specifically, several views are constructed based on attribute information of a knowledge graph. In addition, a Haar convolution-based pooling mechanism is employed to capture the structural features represented by a chain of subgraphs. And then an aggregator that combines both attribute and structural information is used to classify UHV projects. Results from both UHV and NCI-1 datasets indicate that our proposed method is more has higher accuracy and generalization ability.

## KEYWORDS

review of power transmission and transformation, knowledge graph, graph classification, graph convolutional network, knowledge reasoning

## 1 Introduction

With the proposal of carbon peaking and carbon neutralization (J. Liu et al., 2021; Luo et al., 2023; Ren et al., 2021), a new power system based on clean energy consumption has become an important part of achieving dual-carbon goals. In this context, photovoltaic and wind power from the western region of China have become important power sources with a total scale of 450 million kilowatts and increasing. Due to the long distance of electricity transmission, UHV projects are expected to enter a larger-scale construction stage. Unlike conventional power transmission and transformation projects, UHV projects are larger in construction scale with multiple companies involved, which lead to a huge amount of accumulated historical data in terms of quantity and complexity. With the development of digital technology and artificial intelligence and the deep integration of technology and information technology in the field of power grid engineering construction, the concept of digital infrastructure has emerged. As an important part of engineering construction management and control, engineering review is still in an inefficient mode that relies mainly on expert experience and offline review, which makes it difficult to meet the analysis

requirements of the current ever-expanding construction scale of power transmission and transformation projects. Therefore, it is urgent to build an auxiliary review system with knowledge storage and reasoning components based on historical data of power transmission and transformation projects, which leads to an improvement of the digitalization level and the efficiency of power grid infrastructure project review.

At present, the studies of auxiliary review platforms of power transmission and transformation projects are still in the early stages. Browser/Server(B/S) architecture and SQL server databases are two key technologies of these systems that have already been put into the production phase. Built for various voltage levels and different workflows, three key functions are implemented such as data entry, key factor extraction, and data searching (Huang, 2018; Li et al., 2021). Although the application of these platforms has simplified the review process and improved the efficiency of data processing, there are three difficulties yet not be solved: 1) the knowledge extracted from heterogeneous data is stored in relational databases and their connection relationships have not been modeled; 2) knowledge reasoning is difficult based on existing models and data storage methods; 3) knowledge extraction methods are heavily based on expert experiences and human labor. Therefore, knowledge graph technology, which has the advantages of high scalability, high query efficiency, and good visualization, has become one of the excellent technologies to choose from in building the next-generation power transmission and transformation engineering auxiliary review platform.

The knowledge graph is a structured semantic knowledge base that integrates knowledge extraction, data storage, reasoning, and analysis capabilities (Ji et al., 2022). It has been widely used in many fields (Yang et al., 2022; Zou and Lu, 2022; Wu et al., 2023). At present, the application of knowledge graphs in the power field is mainly oriented to aspects such as dispatching, operation and maintenance, and fault handling, and has achieved good results (Pu et al., 2021; Tian et al., 2022; Liu et al., 2023). These researches mainly focus on construction methods such as named entity recognition and relation extraction algorithms of different domains. However, the methodologies and applications of knowledge reasoning technology based on domain knowledge graphs in electric power systems are still in an early stage. Knowledge graphs can be represented as semantic triples or attributed networks (Gao et al., 2023), which are all non-European structural graphs. Traditional deep learning methods such as convolutional neural network(CNN) and recurrent neural network(RNN) cannot be used in such scenarios because representations of graph-structured data is generally irregular, therefore a new deep learning mechanism is needed to process graph structure. In this context, the graph neural network (GNN) is proposed to get the latent representations embedded from nodes, attributes, and structural information of graphs. Compared to the basic network structure of the neural network, the fully connected layer (MLP), which multiplies the feature matrix by the weight matrix, the graph neural network considers structural information, and adds an adjacency matrix as input. With the development of the study, several GNN variants have been proposed. For the first time, graph convolution networks(GCN) introduce convolution operations in image processing to process graph-structured data (Kipf and Welling, 2016). Various experiments show the effectiveness of

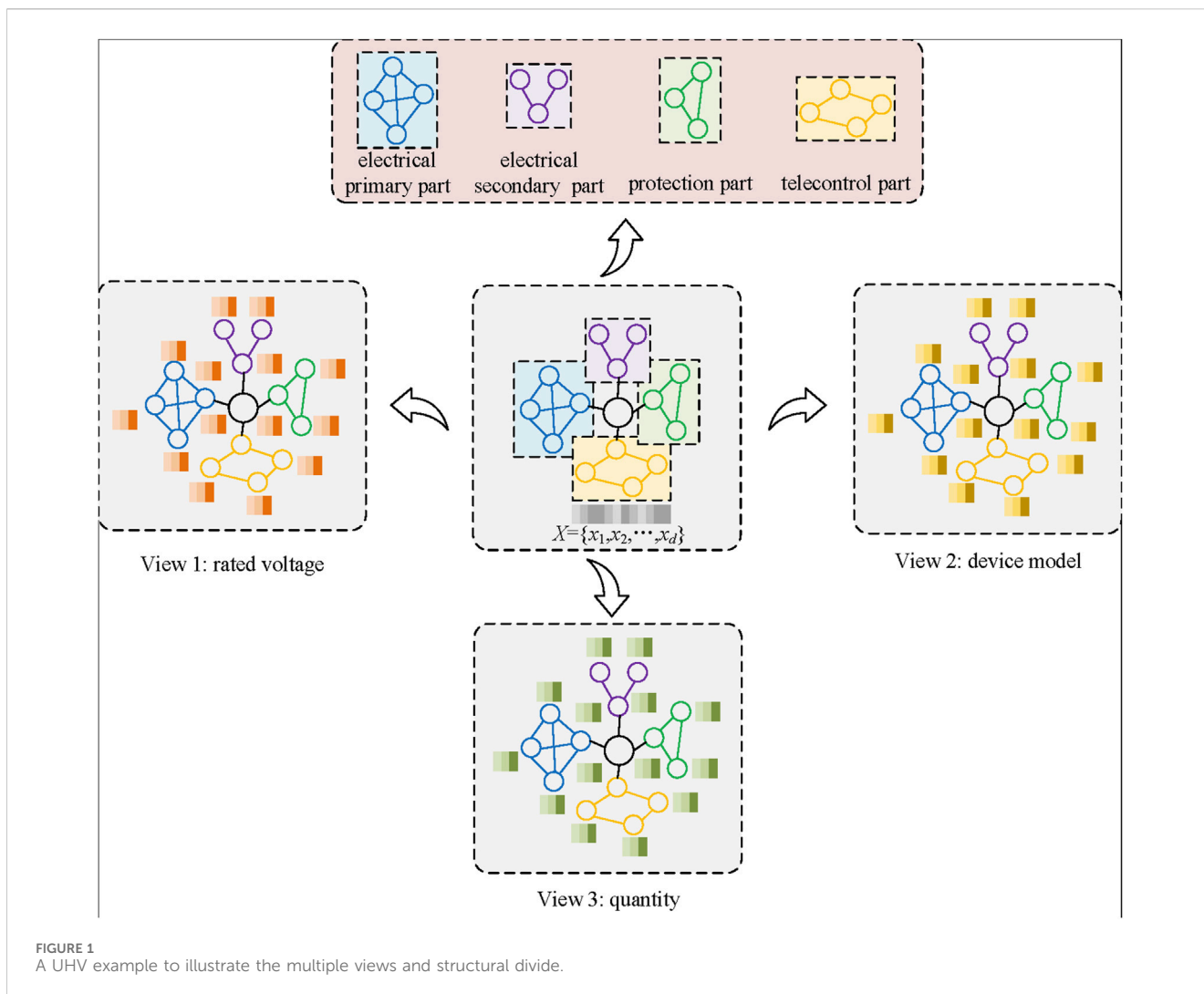
GCN due to the ability to encode the structural information of the graph. However, the shortcomings of GCN are also obvious: 1) GCN needs to put the entire graph data into memory and graphics processing unit(GPU), which requires high-performance equipment when dealing with large graphs. 2) GCN has high computational complexity due to the eigendecomposition operation of graph Laplacian. To handle the problems mentioned above, graph sample and aggregate(GraphSAGE) are proposed (Hamilton et al., 2017). GraphSAGE is an inductive learning framework. In practice, it only retains the training sample to the edge of the training sample during training, and then includes the two major steps which are sample and aggregate. Then, to solve the problem that GNN does not take into account the different importance of different neighbor nodes when aggregating neighbor nodes, graph attention networks (GAT) take the idea of Transformer and introduce the masked self-attention mechanism (Veličković et al., 2017). When calculating the representation of each node in the graph, different weights are assigned to neighbor nodes based on their different characteristics.

However, there are different types of attributes of a node in the knowledge graph. For instance, a transformer could have three attributes such as device model, quantity, and rated voltage. Both of the latter two are usually numeric properties and the first one is a combination of letters and numbers. Conventional GNN-based knowledge modeling and reasoning methods take these attributes as a single matrix, which may lead a confusion about attribute characteristics (Peng et al., 2020). To solve the knowledge reasoning problem in power grid infrastructure projects, a deep multi-view graph convolutional network is introduced considering both the attribute and structural information. This model mainly contains 3 components: a multi-view graph encoder, an aggregator, and classification module. The key contributions are as follows:

- 1) Multi-view information is used separately in knowledge reasoning of UHV projects. It can provide a more accurate feature embedding than a single-view latent representation.
- 2) A more effective pooling mechanism based on the Haar convolution method is introduced considering the structural information of the knowledge graph in the UHV projects.
- 3) Based on the construction of knowledge graphs in the UHV projects domain and the proposition of a deep multi-view graph convolutional network, the problem of knowledge reasoning on project classification is solved, which lays the foundation for the downstream application of the power transmission and transformation auxiliary review platform.

## 2 Problem formulation

The knowledge graph of UHV projects can be divided into two categories based on the attributes of the nodes and the structural information of the graph. Specifically, nodes of knowledge graph can be seen as nodes in the attribute matrix, and the structure information can be modeled by adjacency matrix. As shown in Figure 1, three views are generated based on different attribute types, with each view representing a single aspect of the node attributes, such as rated voltage, device model, and quantity. Several subgraphs are also produced based on the structural information of the graph,



such as the electrical primary part, electrical secondary part, protection part, telecontrol part, and so on.

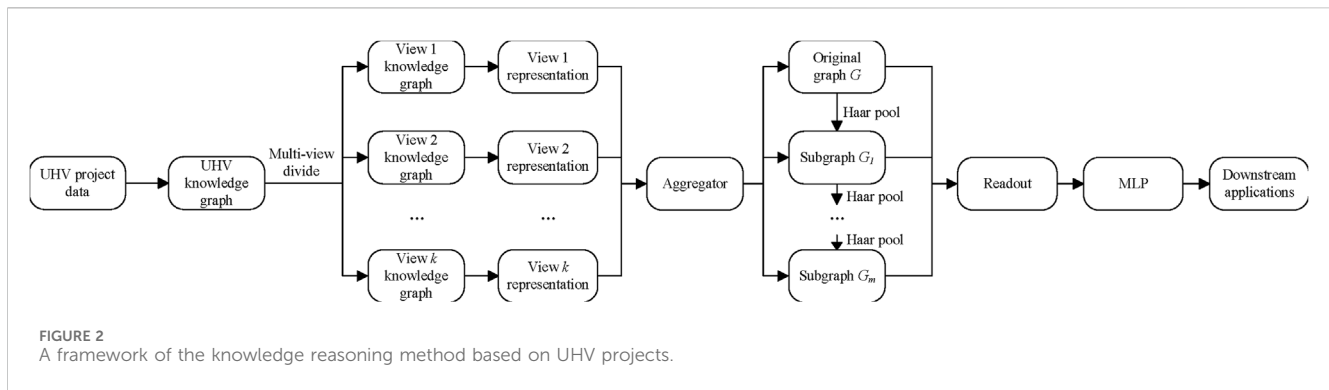
Following the commonly used notations, a knowledge graph is denoted as  $G(V, A, X)$ , where  $V$  denotes the set of nodes,  $A$  denotes the adjacency matrix and  $X$  denotes the attribute matrix. Thus, a multi-view graph of the graph  $G$  can be represented as  $G^{mg}=(V, A, X^{mg})$ , where  $X^{mg} \in \mathbb{R}^{n \times d^{mg}}$  is a submatrix of  $X$  and  $n = |V|$  is the number of nodes,  $d^{mg}$  contains  $k$  distinct attribute features of the graph. As is illustrated above, a multi-view graph has the same structure as the original graph which  $V$  and  $A$  remain unchanged, while the attributes of the nodes are part of  $X$ . On the other hand, a subgraph of  $G$  can be denoted as  $G^s=(V^s, A^s, X)$ , where  $V^s$  and  $A^s$  are the submatrix of the original  $V$  and  $A$ , representing a divide from the structural aspect.

Our purpose is to classify the knowledge graph subgraphs, which is a fundamental function for the review of power transmission and transformation projects. Therefore, a function to generate a probability of each graph should be learned with both  $G^{mg}$  and  $G^s$  as input, which could be represented by the following Equation (1), where  $Z$  denotes the probability of the graph,  $f$  is a trainable function of the deep multi-view graph convolutional network,  $X_{mg}$  and  $X_s$  denotes the embeddings of  $G^{mg}$  and  $G^s$  respectively.

$$Z = f(X_{mg}(G^{mg}), X_s(G^s)) \tag{1}$$

### 3 Framework and methodologies

Considering both the attribute and structural information of the UHV knowledge graph, the knowledge reasoning framework mainly takes five steps. Firstly, with the analysis of historical structured, semi-structured, and unstructured data, a UHV knowledge graph is constructed. Secondly, considering the attribute information of the knowledge graph, a series of multi-view graphs are generated and GCN is leveraged to learn the latent representation of each graph. Thirdly, an aggregator is used to combine multi-view representations. Fourthly, a hierarchical Haar graph pooling method is adopted to replace the graph Laplacian-based GNN considering the structural information of the graph. Finally, a unified representation of the knowledge graph is generated and ready for graph classification and other downstream applications. The framework of the knowledge reasoning method is shown in Figure 2. More details will be elaborated in the following sections.



### 3.1 Knowledge graph construction based on UHV projects

Data from the UHV projects mainly contains three parts: structural, semi-structural, and non-structural data. Project investment table, equipment list are typical structural data. Equipment inventory is semi-structural data because the properties can be put into the same cell and break the structural feature of the data. Preliminary design instructions, feasibility study reports are some of the non-structural data, which account for a large proportion. The knowledge graph is a structured semantic knowledge base used to describe concepts in the physical world and their relationships in symbolic form. In this section, a two-stage method is proposed to construct a UHV knowledge graph.

#### 1) Knowledge graph ontology construction

Ontology is a model and a pattern constraint on the data that constitutes the knowledge graph. Building an ontology in a specific domain requires cooperation with multiple experts in vertical fields. The inputs for constructing ontology include domain knowledge, terminology dictionary, experience of experts, etc. The output includes the entity categories that constitute the knowledge graph, as well as the relationships between categories, the set of attributes that entities of a specific category, and so on. There are generally three methods for constructing knowledge graph ontology: top-down, bottom-up, and a combination of the two. Among them, the top-down approach is to define the most common top-level concepts in the field and then expand downwards in sequence. It requires a thorough understanding of a specific domain both in the business aspect and the data aspect as well. The bottom-up method is just the opposite. It starts with entities, summarizes and organizes entities to form low-level concepts, and then gradually abstracts upward to form upper-level concepts.

In the UHV projects knowledge domain, heterogeneous data comes from different sources. On the one hand, existing data models such as traditional power system engineering systems and expert knowledge bases can provide equipment information and other constraints. On the other hand, rich knowledge is buried in various types of data such as unstructured, semi-structured, and structured data, which needs to be added to the knowledge base using data mining methods. To ensure the integrity of UHV knowledge graph ontology, a combination of top-down and bottom-up is used. Expert experiences are regarded as the guidelines of the construction, while

knowledge extracted from data is a supplement to the ontology. These two aspects together form the conceptual part of the knowledge graph.

#### 2) Knowledge graph construction for UHV projects

Knowledge construction for UHV projects constitutes the data layer of the knowledge graph. It mainly contains 2 components: knowledge extraction and knowledge fusion.

**Knowledge extraction** is a technology that automatically extracts structured information such as entities, relationships, and entity attributes from heterogeneous data. Conventional named entity recognition (NER) algorithm includes BiBERT-LSTM-CRF (Huang et al., 2015), RoBERTa-CRF (Liu et al., 2019), and other deep neural network models. To solve the nested entity problem, methods based on entity matrix such as GlobalPointer are proposed. Other methods such as TPLinker, Tencent Multi-head, and Deep Biaffine are also suitable for solving nested NER problems. As for relation extraction, recent research mainly focuses on joint extraction methods, which can be divided into two categories: sequence annotation-based methods which converts joint extraction problem into sequence annotation problem when decoding, and sequence-to-sequence based methods (Zeng et al., 2018; Takanobu et al., 2019). For constructing a UHV knowledge graph, due to the fact that there exists a nested NER phenomenon, a combination of manual annotation and GlobalPointer is recommended. For example, given a sentence “The Beijing 1000 kV substation project requires four new main transformers”, which describes the quantity of transformers in the project. To recognize the nested entity “Beijing 1000 kV substation project”, GlobalPointer first lists all the entity candidates and then Pick out the real entities with entity labels. It can be concluded that a sentence with  $n$  words could generate  $\frac{n(n+1)}{2}$  candidates. With  $m$  labels, the NER problem is then converted to a multi-label classification problem. Although the time complexity is  $\sigma(n^2)$ , with the use of Transformer and the design of parallel computing, it can be reduced to  $\sigma(1)$ .

**Knowledge fusion** refers to the fusion of description information about the same entity or concept from multiple data sources and the integration and disambiguation of heterogeneous data under unified standards for knowledge graphs. It requires two processes such as implement entity linking and knowledge merging. The process of entity linking is to use a given entity referent to

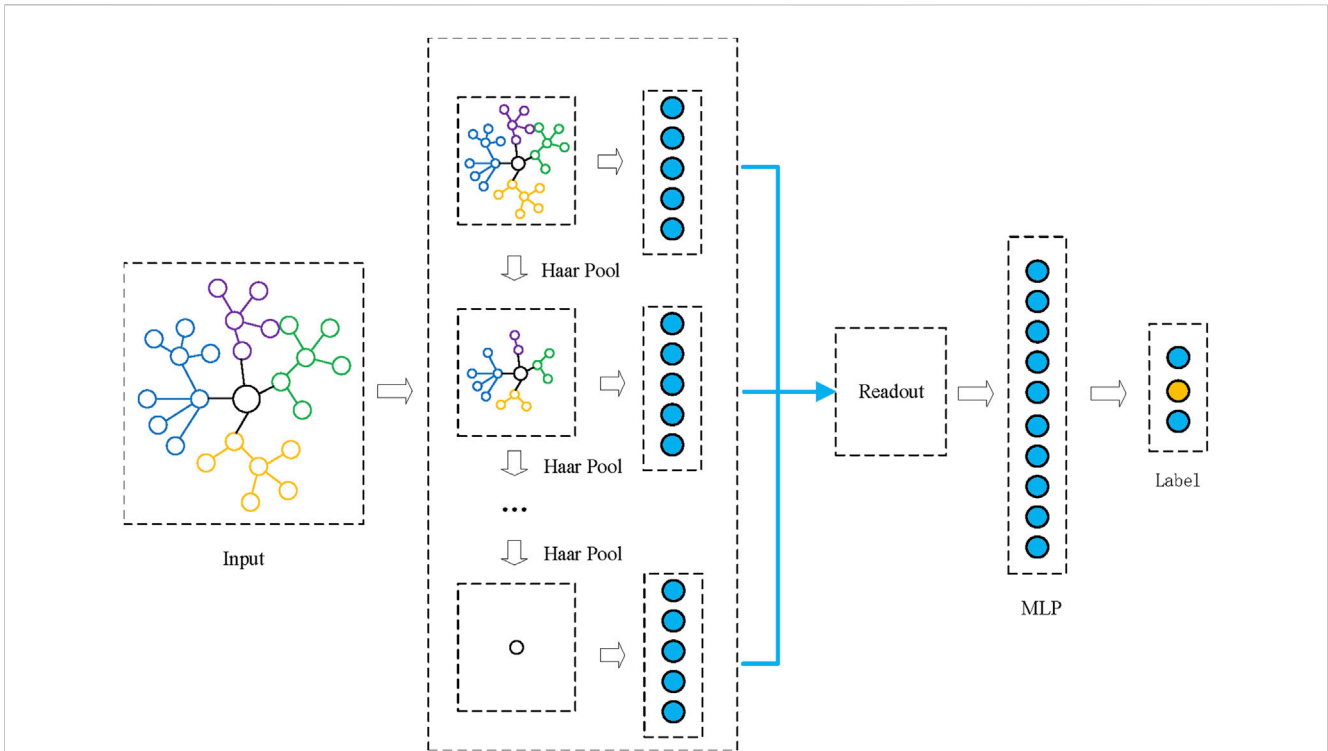


FIGURE 3 Part of a knowledge graph based on a UHV project.

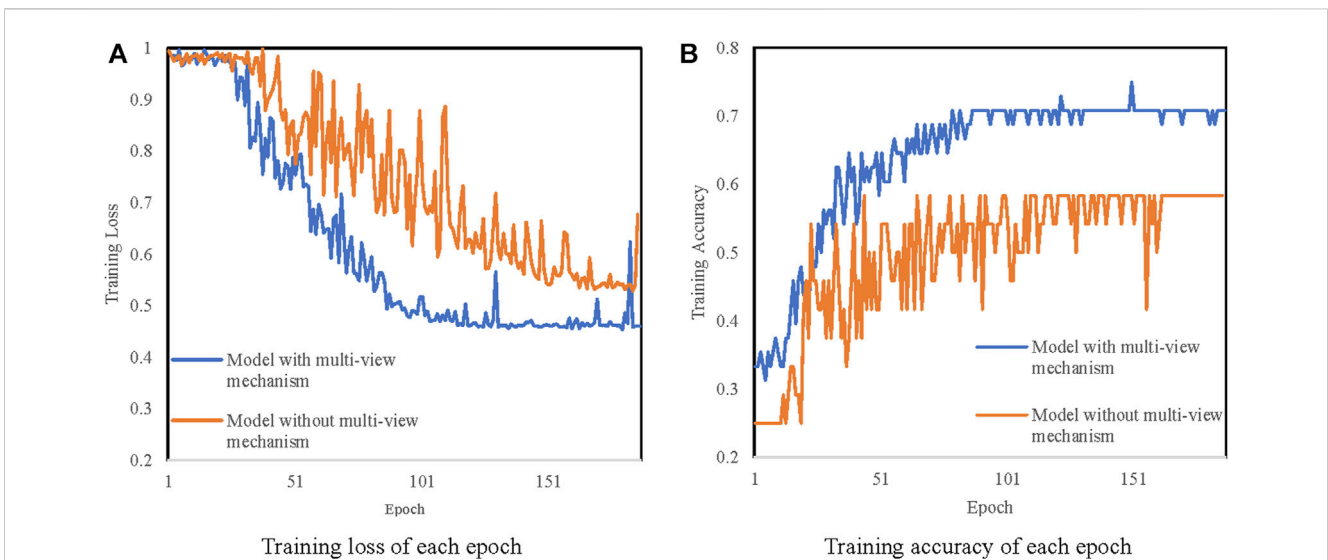


FIGURE 4 A subgraph chain of UHV knowledge graph.

perform entity disambiguation and coreference resolution through similarity calculation. After confirming the correct entity object, the entity referent is linked to the corresponding entity in the knowledge graph. Among them, entity disambiguation solves the problem of ambiguity of entities with the same name, and coreference resolution solves the problem of multiple references corresponding to the same entity object.

Based on the data extracted from a UHV project, Figure 3 shows the constructed knowledge graph. As can be seen, a UHV project mainly contains six parts: electrical primary part, electrical secondary part, hydraulic part, HVAC part, remote control part and protection part, which is represented as orange nodes in the figure. Each part contains multiple equipment, with components followed.

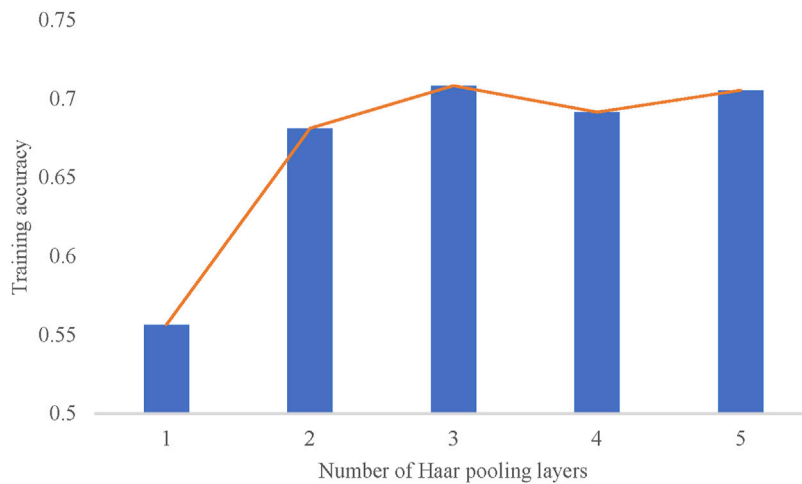


FIGURE 5  
A framework of graph classification based on the Haar pooling mechanism.

### 3.2 Multi-view graphs encoder

After obtaining the knowledge graph of UHV projects, we use attributed network model to analyze the knowledge reasoning problem. To transform knowledge graphs to attributed networks, we take entities as nodes and edges between entities become edges between nodes. And attributes of each node are embedded before concatenated as a vector. For a specific UHV graph, we divide it into multi-view graphs considering the difference of multiple attributes of a node. Taking the accuracy and effectiveness of the multi-view divide into account, there are two principles: 1) attributes with different units need to be divided into different views in order to avoid confusion of attribute information due to excessive numerical differences; 2) pure numeric properties and mixed text and numeric properties should be separated into separate views.

After the construction of multi-view graphs, a GCN-based encoder is adopted to get the latent representation of each graph. GCN is a convolutional neural network that can directly act on graphs and utilize their attribute and structural information. To get the most precious embedding of the graph, several layers of GCN are usually used. Given a graph  $G(V, A, X)$  with  $k$  views, each of the multi-view graph can be denoted as  $G_i^{mg}(V, A, X_i^{mg})$ , where  $X = (X_1^{mg}, X_2^{mg}, \dots, X_k^{mg})$ . The embedding of each layer can be denoted as:

$$\mathbf{H}^{l+1} = f(\mathbf{H}^l, \mathbf{A}, \mathbf{W}^l) \tag{2}$$

where  $\mathbf{H}^l$  is the input of the layer  $l$ , and  $\mathbf{W}^l$  is the trainable parameters.

More specifically, the function  $f(\mathbf{H}^l, \mathbf{A}, \mathbf{W}^l)$  can be expressed as follows:

$$f(\mathbf{H}^l, \mathbf{A}, \mathbf{W}^l) = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^l\mathbf{W}^l\right) \tag{3}$$

where  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ ,  $\tilde{\mathbf{D}}$  is the degree matrix of the nodes with  $\tilde{D}_{i,i} = \sum_j \tilde{A}_{i,j}$ ,  $\sigma$  is an activation function such as *Relu*. To simplify the calculation,  $\mathbf{W}^l$  is shared for all the nodes. It is worth noticing that GCN can only capture the information of the neighborhood for a specific node. In order to take long-range information into consideration, a deep GCN network with  $k$  convolutional layers is

required. However due to an over-smoothing problem, the number of layers is no more than three. In this paper, a UHV knowledge graph is built with a depth of three, thus a three-layer GCN network is used. The final feature representation can be expressed as:

$$\mathbf{Z}_i^{mg} = Relu(Relu(Relu(\mathbf{X}^{mg}, \mathbf{A}, \mathbf{W}^0), \mathbf{A}, \mathbf{W}^1)\mathbf{A}, \mathbf{W}^2) \tag{4}$$

### 3.3 Multi-view graphs aggregator

For every low-dimensional feature representation, it is important to understand that each view only contains part of the attribute information of the nodes. To be more specific, in the UHV knowledge graph,  $\mathbf{Z}_1^{mg}$ ,  $\mathbf{Z}_2^{mg}$  and  $\mathbf{Z}_3^{mg}$  represent the embedding of rated voltage, quantity and device model respectively. To get the full description of the nodes, an aggregator gathering all information from every view seems sensible. There are usually two methods when it comes to representation aggregation, concatenation and weighted addition. As each attribute is equally important to a specific node, we apply concatenation to polymerize each representation. The unified representation of the original knowledge graph can be denoted as follows:

$$\mathbf{Z} = concat(\mathbf{Z}_1^{mg}, \mathbf{Z}_2^{mg}, \mathbf{Z}_3^{mg}) \tag{5}$$

### 3.4 Construction of subgraphs based on Haar pooling

After obtaining the unified representation of the UHV knowledge graph, traditional algorithms take it as an input to a multilayer perceptron (MLP) and output a possibility score of each category. Because the electric power grid infrastructure domain has a strong hierarchical relationship, especially in the equipment selection area, we can use a clustering method to get the node aggregation features in the knowledge graph. Further

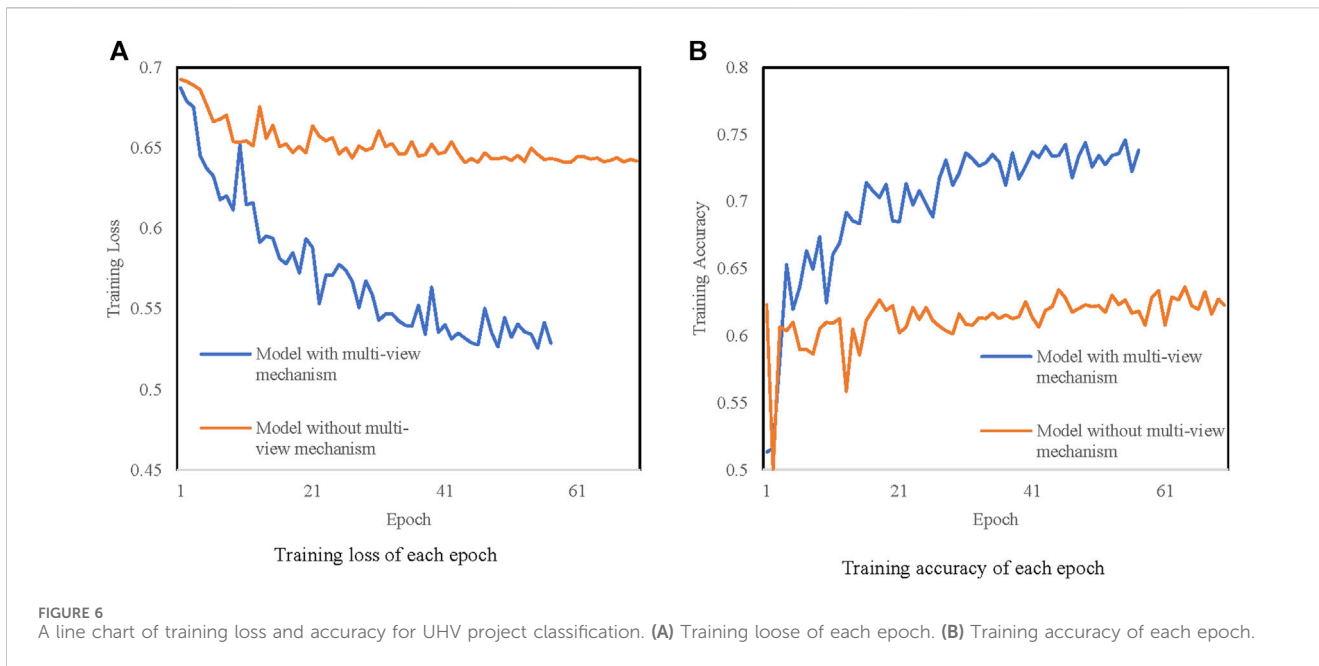
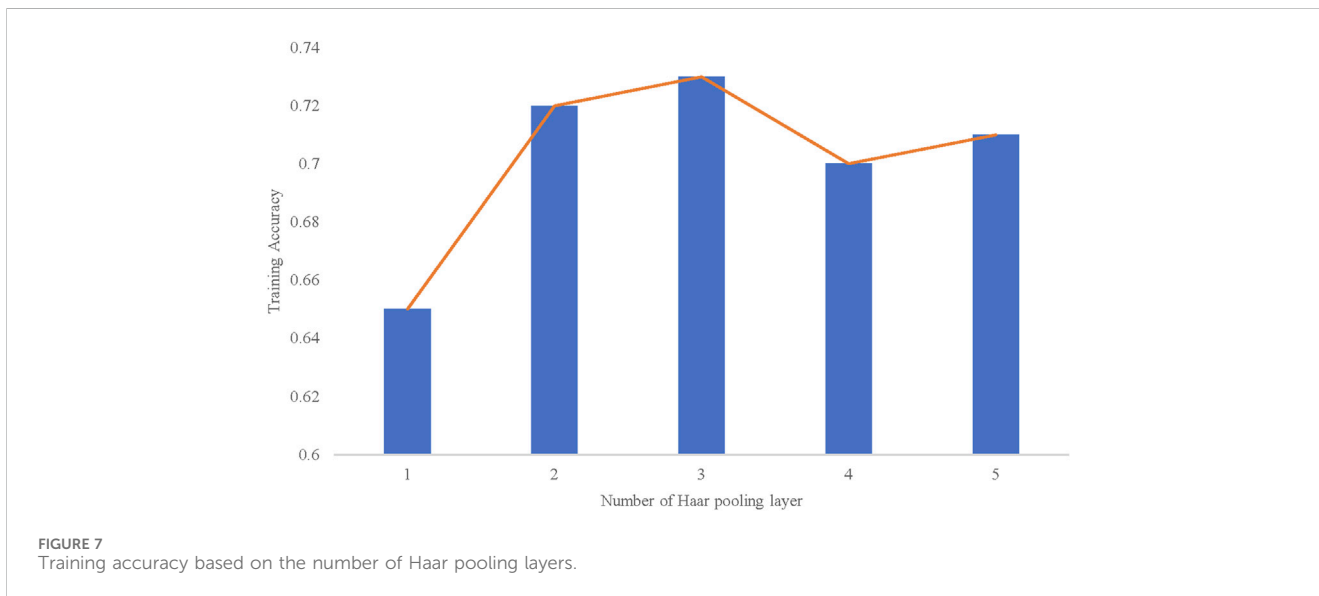


TABLE 1 UHV projects classification considering multi-view mechanism.

Algorithm	Epochs	Test accuracy	Test loss	Time
Haar pool with multi-view mechanism	188	0.7083	0.4600	0.2645
Haar pool without multi-view mechanism	187	0.5833	0.6774	0.2436



analysis of the knowledge graph shows a sparse characteristic which means the number of edges is far less than the number of nodes. Traditional algorithms such as the k-nearest neighbors algorithm(KNN) use the distance between nodes as the basis for classification, which is not suitable for graph data because it ignores the topology information. Therefore, considering the clear hierarchy relationship in components, parts, and

equipment in the electric power equipment knowledge graph, we can obtain a chain of subgraphs based on the facts of the business situation without adopting a clustering algorithm. In this context, getting a series of representation of subgraphs effectively becomes the focus of this problem. In this paper, a Haar convolution-based pooling mechanism is leveraged to the graph unified representation.

TABLE 2 Test results of different pooling mechanisms.

Algorithm	Test accuracy	Test loss	Time
Multi-view mechanism with Haar pool	0.7083	0.4600	0.2645
Multi-view mechanism with SAGPool	0.6753	0.7521	0.2015
Multi-view mechanism with CGIPool	0.6861	0.8652	0.3054
Multi-view mechanism with GSAPool	0.6675	0.8534	0.2465

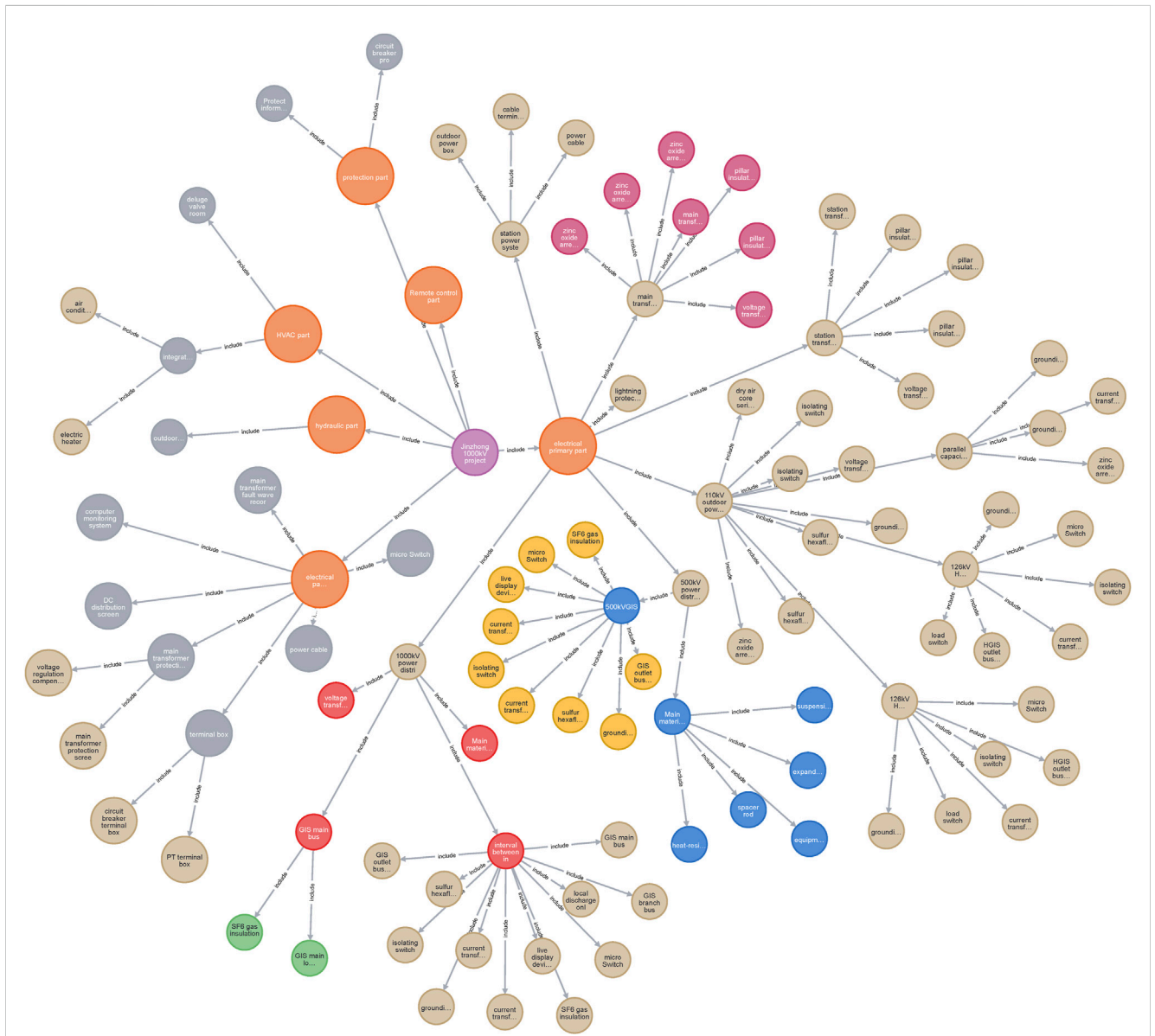


FIGURE 8 A line chart of training loss and accuracy for NCI-1 classification. (A) Training loose of each epoch. (B) Training accuracy of each epoch.

Given a chain of the subgraphs ( $G_0, G_1, \dots, G_m$ ), each node in  $G_i$  is a set of nodes in  $G_{i+1}$ . As shown in Figure 4,  $G_0$  represents the original knowledge graph, while  $G_3$  can be a single node that represent the whole knowledge graph. The Haar convolutional-based pooling mechanism should be used in every layer of the

chain, getting a smaller graph from  $G_i$  to  $G_{i+1}$ . Since a readout module is used afterward to integrate every representation of subgraphs, it is important to guarantee each matrix has the same output dimension. The output of a specific subgraph  $G_i(V^i, E^i)$  is defined as:



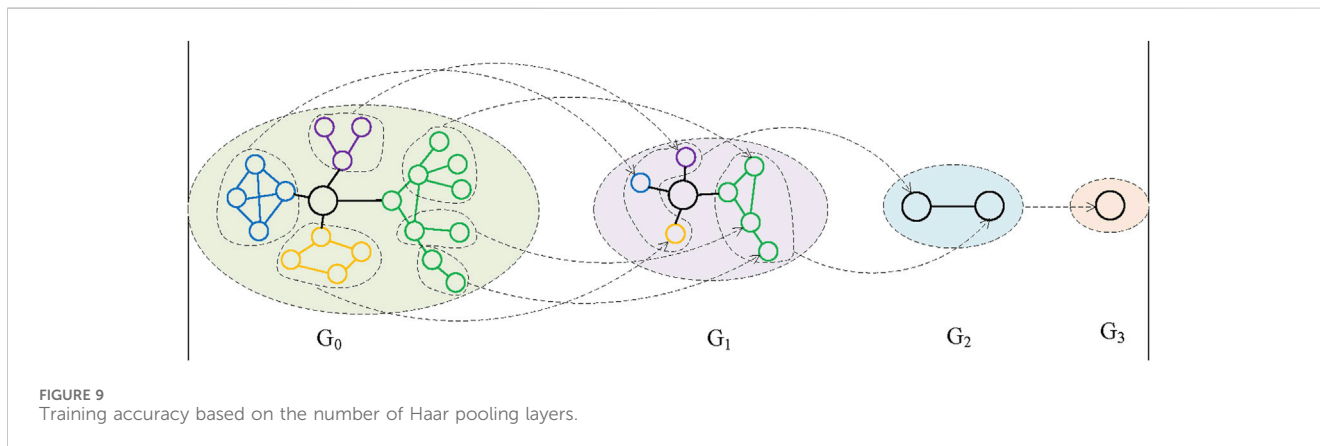


FIGURE 9 Training accuracy based on the number of Haar pooling layers.

$$X_i^{output} = \Phi_i^T X_i^{input} \tag{6}$$

$$Y = \text{softmax}(\text{MLP}(X_{\text{Readout}})) \tag{9}$$

where  $\Phi_i^T$  is the transpose of Harr basis matrix of the  $i$ th subgraph,  $V^i$ ,  $E^i$  represents the nodes and edges of  $G_i$  respectively.

Then the Haar basis could be obtained using the following steps: 1) order  $V^i$  by their degrees of nodes as  $\{v_1^i, v_2^i, \dots, v_N^i\}$ , where  $N$  denotes the number of nodes in  $G_i$ ; 2) the  $N$  vectors can be calculated by the following equations:

$$\begin{aligned} \Phi_1^i &= \frac{1}{\sqrt{N}} \quad l = 1, \\ \Phi_l^i &= \sqrt{\frac{N-l+1}{N-l+2}} \left( x_{i-1}^i - \frac{\sum_{j=l}^N x_j^i}{N-l+1} \right) \end{aligned} \tag{7}$$

where  $x_j^i$  is a function for  $j$ th node, which could be expressed as:

$$x_j^i(v) = \begin{cases} 1, & v = v_j^i \\ 0, & v \in V^i \setminus \{v_j^i\} \end{cases} \tag{8}$$

To speed up the calculation process, a compressed Haar basis matrix is used to replace the original  $\Phi_i$ . Considering the fact that the first  $N_{i+1}$  represents the low frequency coefficients which contains the majority information of the graph, we adopt a tailor operation to reduce the size of  $\Phi_i$ . The Haar convolutional-based pooling mechanism requires less computational complexity than the traditional graph Fourier transform. The use of fast algorithms further improves calculation speed. The key to the fast algorithms is the Haar basis, which are sparse matrixes to which a compressive method can be adopted. Specifically, given a chain of subgraphs  $(G_0, G_1, \dots, G_m)$ , where  $G_m$  is the coarsest graph with only one node. Each subgraph could generate a set of  $n$  orthogonal vectors called Haar basis, where  $n$  is the number of nodes in the specific subgraph. They are the representations of the input in the Haar domain. Analysis shows that these matrixes are sparse and several low-frequency coefficients contains most of the information of the input. Therefore, we use first  $n_{j+1}$  Haar basis vector in our pooling layer, while the other components are aborted.

In order to minimize errors caused by information loss, a hierarchical mechanism is adopted. As shown in Figure 5, every low-dimensional representation is aggregated together using a Readout module and then an MLP is used to get the probability of each label for subgraph  $G_i$ .

## 4 Case studies

In this section, we tested our proposed method on two datasets along with several traditional methods as comparison algorithms. The two datasets are UHV\_Projects and NCI-1. Among them, UHV\_Projects is constructed on historical engineering data, and NCI-1 is an open-source dataset widely used to test the classification performance of different algorithms. All these experiments are performed on a laptop with Intel Core i7-11800H CPU and NVIDIA GeForce RTX 3070 Laptop GPU. Some of the hyperparameters are set as follows: batch size is fixed to 50; learning rate is fixed to 0.001.

### 4.1 UHV projects data classification

The construction of the UHV projects knowledge graph contains heterogeneous data from multiple data sources such as Preliminary Design Instructions, Equipment Inventory, Specialized Reports, and so on. It requires efforts from experts from both the power electrical domain and computer domain. Due to the complexity of different business needs, we only construct a dataset based on the requirements of power transmission and transformation project review with a focus on equipment. Therefore, the dataset mainly contains four levels: the first level is the name of the projects; the second level contains six professional fields based on construction guidelines, such as electrical primary part, electrical secondary part, protection part, telecontrol part, hydraulic part and fire-controlling part; the third level mainly refers to major equipment of each part of second level, such as main transformers, 1000 kV power distribution unit; the tertiary level contains parts of the major equipment, such as cable, voltage transformer, current transformer, breaker and so on.

Taking each project as a subgraph, a label is assigned to each graph indicating the type of the project, which are the new substation project, substation expansion project, and bay expansion project. The difference between the three kinds of projects mainly lies on the quantity and equipment models. For

TABLE 3 NCI-1 classification considering multi-view mechanism.

Algorithm	Epochs	Test accuracy	Test loss	Time
Haar pool with multi-view mechanism	57	0.7385	0.5286	5.5319
Haar pool without multi-view mechanism	70	0.6223	0.6419	4.8624

TABLE 4 Test results of different pooling mechanisms.

Algorithm	Test accuracy	Test loss	Time
Multi-view mechanism with Haar pool	0.7385	0.5286	5.5319
Multi-view mechanism with SAGPoll	0.6967	0.6519	5.1864
Multi-view mechanism with CGIPool	0.7058	0.7622	5.9571
Multi-view mechanism with GSAPool	0.6885	0.6681	5.3748

example, a new substation project usually has more main transformers than substation expansion projects and a bay expansion project may only have some in-bay equipment such as isolating switches and lightning protectors. Thus, we constructed 120 graphs, with an average of 130 nodes and 128 edges per graph. We first test our model and our model without multi-view mechanism on this dataset and the results are illustrated in Figure 6 and Table 1. It can be seen that the accuracy of our model with a multi-view mechanism is improved by 21.43% with a little more time consumption.

We also add an analysis regarding the number of Haar pooling layers. It describes the number of subgraphs which is an indicator of different clustering methods. It should be noted that 1 layer represents the model without the Haar pooling strategy, and 2 layers only consist of two subgraphs which are the original knowledge graph and the final knowledge graph with only one node. The results are shown in Figure 7. Due to the fact that the UHV knowledge graph could be divided into three levels which are professional domains, equipment, and parts, a three-level clustering is most proper and the result of the experience also proves the conclusion.

To test the performance of the Haar pool, we selected three traditional pooling methods as the control group such as SAGPoll (Lee et al., 2019), CGIPool (Pang et al., 2021), and GSAPool (Zhang et al., 2020). The results are illustrated in Table 2. According to test results of different pooling mechanisms, our proposed model shows excellent performance both in accuracy and loss. Although a multi-view mechanism with SAGPoll achieves the highest computational efficiency, the accuracy and the loss of our proposed model outperforms it by 4.88% and 63.5%. It shows that a little sacrifice of the time complexity returns good performance.

## 4.2 NCI-1 data classification

To test the generalizability of our proposed model, we use the NCI-1 dataset as our input. NCI-1 is a popular open-source benchmark dataset mainly focused on chemical and medical domain. It contains 4,100 compounds, each one of which could be seen as a graph sample. Among them, nodes represent atoms and edges represent chemical bonds. The task

of the dataset is to determine whether the compound has properties that hinder the growth of cancer cells. As experiments in Section 4.1, we first test multi-view mechanism on this dataset. The results are shown in Figure 8 and Table 3. It shows an increase in test accuracy by 18.67% when comparing the model with a multi-view mechanism to the one without. As the number of samples increases, the running time is also extended. The difference between both algorithms is extended when compared to UHV datasets, which implies that traditional GCN has a non-linear computational complexity.

As for the number of Haar pooling layers, there is no obvious structure of the NCI-1 dataset. We make a test from layer 1 to layer 5. The results are shown in Figure 9. According to training accuracy, apart from the results for 1 layer, the rest of the results are almost the same.

To test the performance of the Haar pool, we also employ the traditional pooling mechanism as in Section 4.1. And the results are shown in Table 4. It demonstrates that our proposed method is best-considering test accuracy and test loss. Due to the multi-view mechanism, the running time is not the shortest of all experiments, but considering the increase in performance, it is completely acceptable.

## 5 Conclusion

This paper proposed a novel multi-view knowledge reasoning method that takes both attribute and structural characteristics into consideration. Firstly, a knowledge graph construction method is proposed based on UHV project data. Secondly, considering the difference of multiple attributes, a series of multi-view graphs are constructed and represented using traditional GCN. Thirdly, a Haar convolutional-based pooling method is leveraged to deal with the structural information with high efficiency. Results from the UHV dataset and NCI-1 dataset prove the feasibility of our algorithm. In general, our contributions are as follows:

- 1) The introduction of a multi-view mechanism to the knowledge reasoning framework improves the accuracy of graph representation learning.

- 2) A Haar convolutional-based pooling mechanism is used in the UHV knowledge graph, which proposes a better way when analyzing knowledge graphs with hierarchical structures.
- 3) Although running time increases, our proposed method shows various improvements in accuracy in different datasets.

In the future, knowledge graphs will become more and more popular in the electric power domain. Knowledge reasoning methods should surely serve in various downstream applications. Our proposed method mentioned above mainly focus on graph classification and could be used in power transmission and transformation review platform.

## Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

## Author contributions

JH: Conceptualization, Writing—original draft. GX: Supervision, Writing—review and editing. LQ: Supervision, Writing—review and editing. XQ: Supervision, Writing—review and editing.

## References

- Gao, F., Li, G., Gu, J., Zhang, L., and Wang, L. (2023). GridOnto: knowledge representation and extraction for fault events in power grid. *IEEE Access* 11, 58863–58878. doi:10.1109/ACCESS.2023.3284839
- Hamilton, W., Ying, Z., and Leskovec, J. (2017). Inductive representation learning on large graphs. *Adv. neural Inf. Process. Syst.* 30. doi:10.48550/arXiv.1706.02216
- Huang, J. (2018). Research on intelligent review management platform for 110 kV and below power transmission and transformation project design. *M. Electr. Tech.* 41, 112–115. doi:10.16652/j.issn.1004-373x.2018.04.028
- Huang, Z., Xu, W., and Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging. Available at: <https://arxiv.org/abs/1508.01991>.
- Ji, S., Pan, S., Cambria, E., Marttinen, P., and Yu, P. S. (2022). A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 494–514. doi:10.1109/TNNLS.2021.3070843
- Kipf, T. N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. Available at: <https://arxiv.org/abs/1609.02907>.
- Lee, J., Lee, I., and Kang, J. (2019). “Self-attention graph pooling,” in Proceedings of the International conference on machine learning, Long Beach, United States, June 2019, 3734–3743.
- Li, G., Zhou, X., and Wang, B. (2021). Research on intelligent review and management platform for power transmission and transformation engineering design of 110kV and above. *Cn. M. Inf.* 24, 107–108.
- Liu, J., Tang, Q., Su, Y., Li, T., Wang, Y., and Zhu, M. (2021). “Economic analysis of solid oxide fuel cell and its role in carbon peak, carbon neutralization process,” in Proceedings of the 2021 4th International Conference on Energy, Electrical and Power Engineering (CEEPE), Chongqing, China, April 2021, 115–119.
- Liu, L., and Yan, X. (2023). Knowledge graph in distribution network fault handling: advances, challenges and prospects. *E. P. Info. Comm. Technol.* 21, 19–26. doi:10.16543/j.2095-641x.electric.power.ict.2023.07.03
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., et al. (2019). Roberta: a robustly optimized bert pretraining approach. Available at: <https://arxiv.org/abs/1907.11692>.
- Luo, Q., Mao, Z., Zhao, Q., Gong, S., Zhao, H., You, H., et al. (2023). “China’s energy evolving approach to carbon peaking by 2030 based on 23 provincial carbon peaking implementation programs,” in Proceedings of 2023 Panda Forum on Power and Energy (PandaFPE), Chengdu, China, April 2023, 1076–1080.
- Pang, Y., Zhao, Y., and Li, D. (2021). “Graph pooling via coarsened graph infomax,” in Proceedings of the 44th International ACM SIGIR Conference on

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Conflict of interest

Author LQ was employed by State Grid Economic and Technological Research Institute Co Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Research and Development in Information Retrieval, New York, USA, July 2021, 2177–2181.

Peng, Z., Luo, M., Li, J., Xue, L., and Zheng, Q. (2020). A deep multi-view framework for anomaly detection on attributed networks. *IEEE Trans. Knowl. Data Eng.* 34, 1–2552. doi:10.1109/TKDE.2020.3015098

Pu, T., Tan, Y., Zheng, G., Xu, H., and Zhang, Z. (2021). Construction and application of knowledge graph in the electric power field. *P. Sys. Technol.* 45, 2080–2091. doi:10.13335/j.1000-3673.pst.2020.2145

Ren, Y., Ma, C., Chen, H., and Huang, J. (2021). “Low-carbon power dispatch model under the carbon peak target,” in Proceedings of the 2021 IEEE Sustainable Power and Energy Conference (iSPEC), Nanjing, China, December 2021, 1–5.

Takanobu, R., Zhang, T., Liu, J., and Huang, M. (2019). “A hierarchical framework for relation extraction with reinforcement learning,” in Proceedings of the AAAI conference on artificial intelligence, Nanjing, China, January 2019, 7072–7079.

Tian, J., Song, H., Chen, L., Sheng, G., and Jiang, X. (2022). Entity recognition approach of equipment failure text for knowledge graph construction. *P. Sys. Technol.* 46, 3913–3922. doi:10.13335/j.1000-3673.pst.2021.1886

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. Available at: <https://arxiv.org/abs/1710.10903>.

Wu, X., Jiang, T., Zhu, Y., and Bu, C. (2023). Knowledge graph for China’s genealogy11.a shorter version of this paper won the best paper award at iee ickg 2020 (the 11th iee international conference on knowledge graph, ickg 2020.bigke.org). *IEEE Trans. Knowl. Data Eng.* 35, 634–646. doi:10.1109/TKDE.2021.3073745

Yang, J., Yang, L. T., Wang, H., Gao, Y., Liu, H., and Xie, X. (2022). Tensor graph attention network for knowledge reasoning in internet of things. *IEEE Internet Things J.* 9, 9128–9137. doi:10.1109/JIOT.2021.3092360

Zeng, X., Zeng, D., He, S., Liu, K., and Zhao, J. (2018). “Extracting relational facts by an end-to-end neural model with copy mechanism,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, July 2018, 506–514.

Zhang, L., Wang, X., Li, H., Zhu, G., Shen, P., Li, P., et al. (2020). “Structure-feature based graph self-adaptive pooling,” in Proceedings of The Web Conference 2020, Taipei, China, April 2020, 3098–3104.

Zou, Q., and Lu, C. (2022). An accurate matching query method of natural language knowledge graph based on hierarchical graph topological sequence. *IEEE Access* 10, 24080–24094. doi:10.1109/ACCESS.2022.3155520