Leveraging graph convolutional networks for semi-supervised fault diagnosis of HVAC systems in data-scarce contexts

Leveraging graph convolutional networks for semi-supervised fault diagnosis of HVAC systems in data-scarce contexts

Cheng Fan[1,2,3], Yuanyuan Ma[2,3], Marco Savino Piscitelli[4], Roberto Chiosa[4], Huilong Wang[1,2,3]* Alfonso Capozzoli[4]

[1]Key Laboratory for Resilient Infrastructures of Coastal Cities, Ministry of Education, Shenzhen University, Shenzhen, China

[2]Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, China

[3]College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, China

[4]Politecnico di Torino, Department of Energy, TEBE research group, BAEDA Lab, Politecnico di Torino, Torino, Italy

*E-mail: wanghuilong@szu.edu.cn

Abstract

The continuous accumulation of operational data has provided an ideal platform to devise and implement customized data analytics for smart HVAC fault detection and diagnosis. In practice, the potentials of advanced supervised learning algorithms have not been fully realized due to the lack of sufficient labeled data. To tackle such data challenges, this study proposes a graph neural network-based approach to effectively utilizing both labeled and unlabeled operational data for optimum decision-makings. More specifically, a graph generation method is proposed to transform tabular building operational data into association graphs, based on which graph convolutions are performed to derive useful insights for fault classifications. Data experiments have been designed to evaluate the values of the methods proposed. Three datasets on HVAC air-side operations have been used to ensure the generalizability of results obtained. Different data scenarios, which vary in training data amounts and imbalance ratios, have been created to comprehensively quantify behavioral patterns of representative graph convolution networks and their architectures. The research results indicate that graph neural networks can effectively leverage associations among labeled and unlabeled data samples to achieve an increase of 2.86-7.30% in fault classification accuracies, providing a novel and promising solution for smart building management.

Keywords: Fault detection and diagnosis; Graph convolutional networks; Semi-supervised

learning; HVAC systems; Machine learning.

Nomenclature

FDD	fault detection and diagnosis
HVAC	heating, ventilation and air-conditioning
AHU	air handling units
GNN	graph neural network
GCN	graph convolution network
Cheb	Chebyshev network
SAGE	graph sample and aggregate network
GAT	graph attention network
GCN2	graph convolution network with initial residual connections and identity mapping
FCNN	fully connected neural network

1. Introduction

Building automation technologies have been widely adopted in the building field to enable intelligent monitoring and controls over various building services systems. The continuous accumulation of building operational data provides an ideal platform to devise and implement smart data-driven methods for optimum decision-makings. As a prominent example, by leveraging advances in machine learning and data science, data-driven classification models can be developed to describe underlying relationships between operating parameters and faulty conditions, which helps to enhance the efficiency and effectiveness for real-time and large-scale HVAC fault detection and diagnosis tasks (i.e., FDD) (Li and O'Neill 2018; Fan et al. 2021a; Piscitelli et al. 2021).

In general, data-driven FDD methods can be developed by firstly defining model inputs and outputs, and then applying statistical and machine learning algorithms for predictive modeling. For instance, chiller operating conditions can be well described through a set of physical parameters, such as the temperatures, pressures and flowrates of refrigerants and circulating water (Fan et al. 2019; Gao et al. 2022). Such variables can be used as model inputs to classify chiller conditions into either normal or faulty conditions. Considering the nonlinearity and complexity between model inputs and outputs, previous studies have explored the potentials of various machine learning techniques in fault classification tasks. The modeling approaches

used vary in their learning natures, ranging from linear (e.g., principal component analysis (Li and Wen 2014; Xia et al. 2021)) to nonlinear algorithms (e.g., support vector machines (Zhao et al. 2013; Han et al. 2019)), and single- (e.g., decision trees (Li et al. 2018)) to ensemble-model architectures (e.g., extreme gradient boosting trees and random forests (Chakraborty and Elzarka 2019; Yao et al. 2022)). Compared with conventional human-centric FDD methods, data-driven FDD can be fully automated and highly accurate given sufficient training data. It is therefore regarded as a more promising and scalable solution for intelligent building management. However, at present, the industrial applications of data-driven FDD methods are rather limited (Mirnaghi and Haghighat 2020; Chen et al. 2022). The main reason is that high-quality labeled data are not always available in practice. On the one hand, high-level domain expertise is needed to assign labels (e.g., *Normal* or *Faulty*) for data samples, making it labor-intensive and time-consuming to prepare sufficient labeled dataset for FDD modeling. On the other hand, operation faults are typically rare events and it is unlikely to encounter all possible faults in various operating conditions to ensure the generalization performance of data-driven fault classification models.

To tackle the above-mentioned data challenges in building operations, research efforts have been spent to apply novel machine learning paradigms for HVAC FDD tasks. As building systems and their operating behaviors share similarities, transfer learning methods can be developed to customize FDD solutions based on the knowledge learnt or data collected from other buildings. Zhu et al. proposed a model-based transfer learning framework for migrating chiller FDD knowledge between two screw chillers (Zhu et al. 2021). Liu et al. devised a convolution neural network-based transfer learning strategy for diagnosing chiller faults (Liu et al. 2021). Considering potential discrepancies in data variables collected from different domains, a two-dimensional image-based transfer learning framework was proposed to enable multi-source data integration and FDD knowledge sharing for air handling units (i.e., AHU) (Fan et al. 2022). Transfer learning has shown promising results on enhancing data-driven FDD performance. Nevertheless, such approach is a typically regarded as a heavyweight solution as it assumes the availability of source domain data. By contrast, semi-supervised learning is a rather lightweight solution as it focuses on the efficient utilization of both labeled and unlabeled data in individual buildings. Yan et al. compared the performance of various semi-supervised

learning algorithms for AHU fault diagnosis, such as semi-supervised support vector machines, decision trees and random forests (Yan et al. 2018). The results validated the potential of semi-supervised learning, as it could significantly lower the labeled data requirement for reliable fault modeling. Fan et al. proposed a semi-supervised learning framework for AHU fault detection and diagnosis (Fan et al. 2021b; Fan et al. 2021c). Semi-supervised neural networks were developed in an iterative fashion by utilizing pseudo-label information predicted for unlabeled data, which led to significant improvements on fault classification accuracies. Another approach is to adopt generative modeling to address data challenges in HVAC FDD tasks. Li et al. adopted generative adversarial networks to explore unlabeled data distributions for improving fault diagnosis performance of packaged rooftop units (Li et al. 2021). The results reported show that at most 10% improvements in fault diagnosis accuracies could be achieved given limited labeled data. Generative adversarial networks have also been used to generate synthetic faulty data samples for reducing imbalance ratios in training data and thereby, resulting in more reliable data-driven models for fault inferences (Yan et al. 2020a; Fan et al. 2021d).

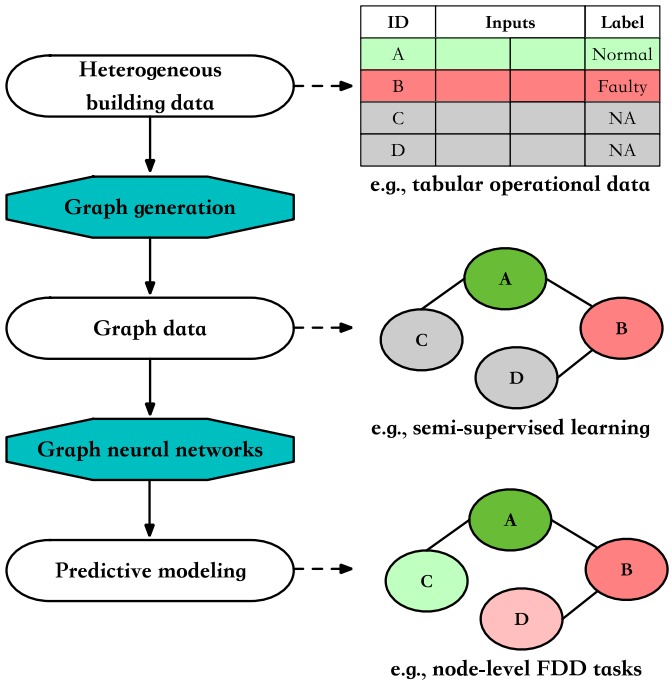

Fig. 1 Roadmap for GNN-based predictive modeling tasks in the building field

Existing studies have proved that semi-supervised learning is extremely promising to ensure the feasibility and reliability of data-driven FDD methods. Despite the encouraging results

obtained, existing solutions are only applicable for analyzing data with conventional Euclidean or grid-like structures. A natural question arises as how to generalize the fault modeling methods to effectively integrate and utilize multi-source and non-Euclidean information embedded in building operations for fault inferences, such as the geometric locations of sensors, the hierarchical information of HVAC systems, and possible connections or linkages among data samples. Such non-Euclidean information is ubiquitous in building system operations and can be very valuable to further increase the theoretical performance limits of data-driven FDD methods. Considering that graphs are compatible with both Euclidean and non-Euclidean information, graph neural networks (i.e., GNNs) have drawn increasing interests from both academia and industries (Keramatfar et al. 2022; Jiang and Luo 2022). GNN-based methods have been developed in diagnosing faults in various industrial processes (Zhang and Yu 2022) and mechanical systems, such as motors, gears and bearings (Li et al. 2022). Nevertheless, few studies have been conducted to exploit its value in smart building system management. It would be of great significance to develop GNN-based solutions for semi-supervised HVAC FDD tasks, as it enables the efficient integration of multi-source information with heterogeneous structures and utilization of both labeled and unlabeled operational data. To bridge the knowledge gap, as shown in Fig. 1., this study proposes a novel GNN-based semi-supervised learning framework for HVAC FDD tasks and evaluates its performance across various datasets. The paper is organized as follows. Section 2 introduces the research methodology with theoretical background provided. Section 3 describes details on data experiments and results are discussed in Section 4. Conclusions are drawn in Section 5.

2. Theoretical background

2.1 Graph-structure data

Graphs are natural and flexible representations of information with heterogenous structures. A graph G consists of a number of vertices and edges, which are denoted as $V(G)$ and $E(G)$ respectively. The vertices or nodes are typically used to represent a set of entities, while their relationships can be described using either directed or undirected edges or links. Detailed information on nodes, edges or the entire graph can be stored as attributes or embeddings. Graphs have been widely used to describe heterogenous information in various fields, such as social networks, molecule structures, academic citations, and commercial transactions

(Sanchez-Lengeling et al. 2021).

Fig. 2 presents two example types of graphs for HVAC systems. As shown in Fig. 2(a), vertices or nodes represent different entities in a chiller plant while undirected and unweighted edges are used to represent hierarchical and connection information among system components. Each node has several attributes describing physical measurements at different time steps, such as chiller plant COP and part-load ratios, and operating frequencies of pumps. By contrast, Fig.2(b) treats each data sample as a node and uses weighted edges to depict similarities among data samples. Each node has three attributes representing temperature, flowrate and pressure measurements, while each edge has one attribute describing inter-node similarities. In addition, global attributes can be designed to represent global information of a graph, such as node and edge numbers. As we shall introduce in the following two subsections, graph convolutions can be performed to aggregate information among graph nodes and therefore, providing a valuable approach to integrating both "self-information" and "neighboring-information" from tabular data for predictive modeling. More specifically, conventional machine learning algorithms for tabular data analysis, such as fully connected neural networks, typically treat each data sample as an independent observation and therefore, is making predictions based on the "self-information" of each data sample in essence. Even though Euclidean distance is commonly used to define weighted edges in Fig. 2(b), such graph layout allows information passing among graph nodes, which enables heterogenous information integration for both inductive and transductive decision makings in the building field.

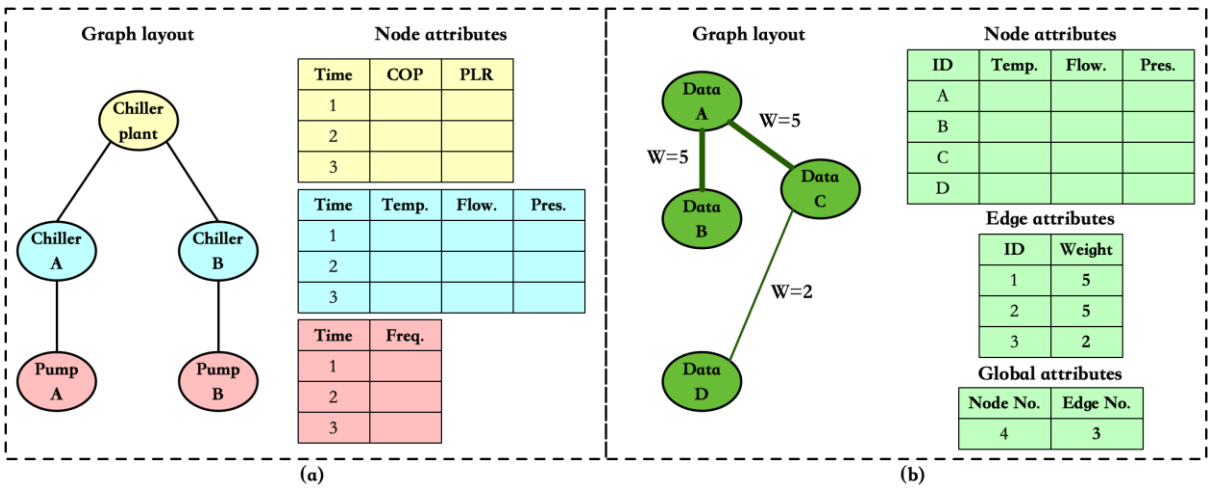

Fig.2 Example graphs for HVAC systems

2.2 Graph learning problems

As shown in Fig. 3, three types of prediction tasks can be performed based on graph data, i.e., graph-level, node-level and edge-level (Daigavane et al. 2021). Graph-level tasks aim to predict a single property, which is of either regression or classification nature, from a whole graph. For instance, as shown in Fig. 3(a), a binary classification problem can be formulated to classify the COP of HVAC system operations into either *Low* or *High* based on the information contained in the whole graph. Node-level tasks aim to predict certain property for each node and example applications include node-level classification and node-level clustering. As an example, node-level binary classification problems can be formulated to classify whether a node represent *Normal* or *Faulty* operations, which are colored as red and yellow respectively in Fig. 3(b). Edge-level tasks try to predict the property or presence of edges in a graph. As shown in Fig. 3(c), a link prediction task can be formulated as a binary classification problem to predict whether an edge should be added between Nodes *A* to *D* and Nodes *B* to *D*.

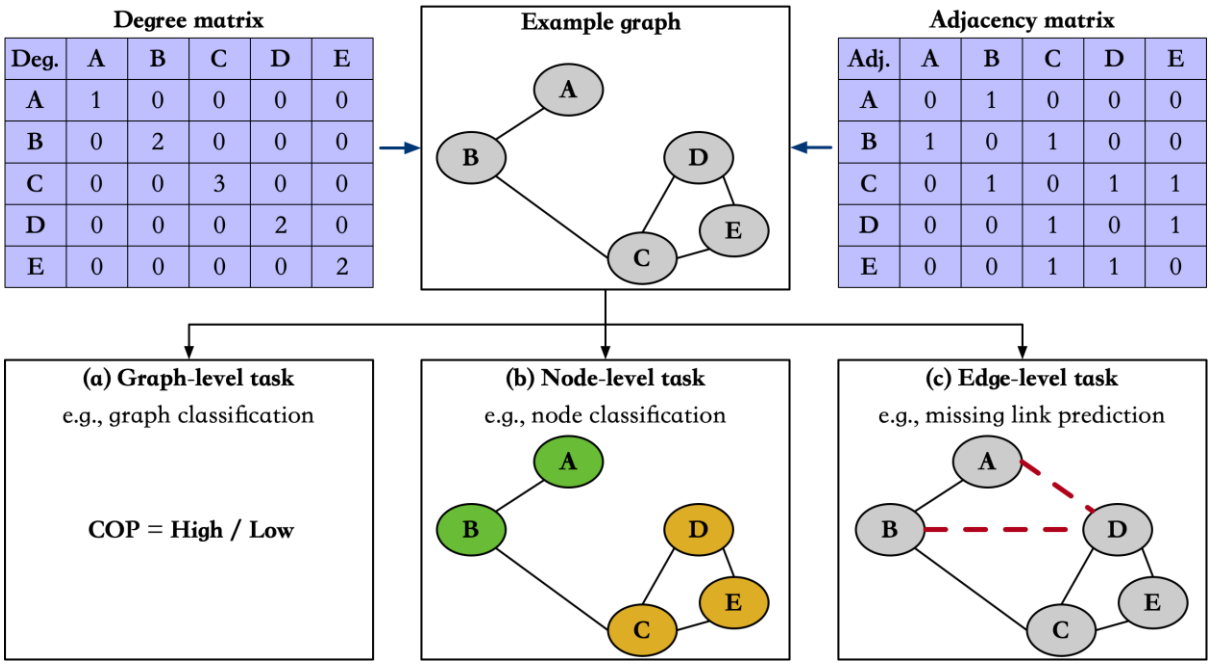

Fig. 3 Typical graph learning problems

In essence, four types of information can be utilized for graph learning problems, i.e., nodes, edges, global-context and connectivity (Daigavane et al. 2021; Sanchez-Lengeling et al. 2021). The former three types of information are typically represented as conventional Euclidean formats. For instance, node information can be stored using a node feature matrix, where each row corresponds to a node and each column represents a node attribute. Such data formats are compatible with existing machine learning algorithms. However, it is non-intuitive to represent

and process connectivity information using conventional machine learning algorithms. The connectivity among nodes can be represented as adjacency matrices or adjacency lists. As shown in Fig. 3, the dimension of an adjacency matrix is $n \times n$, where n represents the total node number. The entry value is either 1 or 0, indicating whether two nodes are connected or not. A diagonal degree matrix can also be defined indicating the total degree of each node. Graph neural network or GNN is a special kind of neural networks designed to process all the above-mentioned four types of information. Existing studies have shown that GNNs have better performance in analyzing graph-structured data compared with other methods, such as the graph kernel (Vishwanathan et al. 2010) and random-walk methods (Grover and Leskovec 2016). The rationale behind GNNs is introduced in the following subsection.

2.3 Basics on graph neural networks

The main intuition of graph neural networks is to perform optimizable transformations on graph attributes while preserving graph symmetries. As introduced in Section 2.1, graph attributes can be embedded at node-, edge- or global-level. The backbone of GNNs adopts the concept of graph convolution for message or information processing. More specifically, graph data are processed through graph convolution operations using a three-step approach. Firstly, the neighboring information of each node or edge is gathered considering graph connectivity. Secondly, an aggregation function, such as the mean or sum operation, is used to aggregate neighboring information gathered. Thirdly, an update function, which is typically learnt and optimized through the use of a fully connected neural network, is used to transform the information aggregated to updated embeddings.

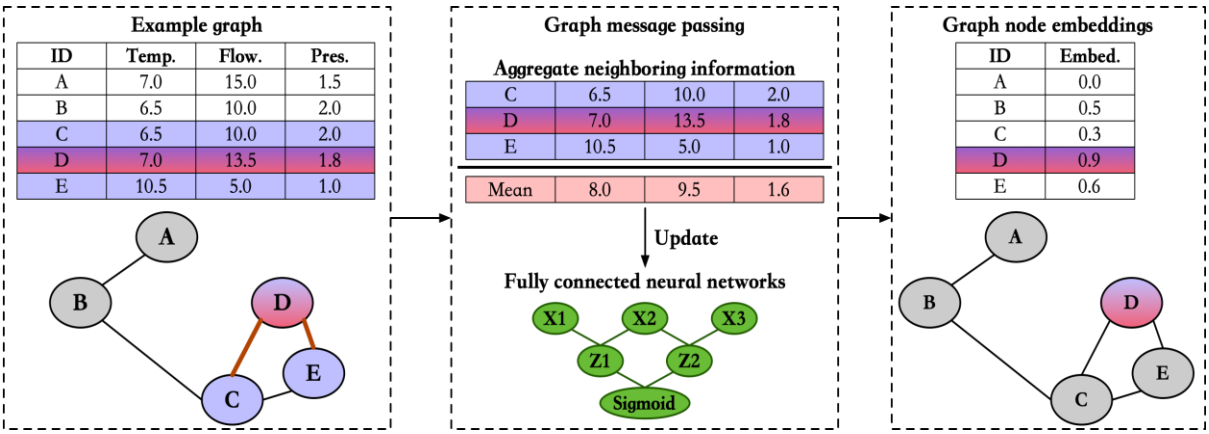

Fig. 4 An example of graph message passing process

As an example, Fig. 4 presents the message passing process for Node D. The example graph has three node-level attributes, i.e., the temperature, flowrate and pressure. Node D has connections with Nodes C and E. Besides information contained in Node D itself, the information gathered for aggregation can be stored as a 3×3 matrix. Using the mean operation as the column-wise aggregation function, the node attributes can be readily calculated as 8.0, 9.5 and 1.6 respectively. A three-layer fully connected neural network is then used for information update. Assuming the graph learning problem is a node-level binary classification task, the output layer of fully connected neural networks has one neuron with a Sigmoid activation function. As a result, the embedding of each node is a scalar which ranges between zero and one. Similar steps can be performed to update edge information if needed. It should be noted that multiple graph convolution layers can be designed to enable node-level or edge-level information exchanges with greater connectivity.

Various graph convolution models exist differing in their learning mechanisms. There are two general graph convolution approaches, i.e., spatial- and spectral-based approaches (Wu et al. 2021). The spectral-based approach defines graph convolutions by utilizing polynomial filters from the perspective of graph signal process (Zhou et al. 2020). Spectral-based methods can be computational extensive as graphs become larger. Research efforts have been made to propose more efficient solutions by making approximations to polynomial filters and simplifications in eigen-decomposition process (Daigavane et al. 2021; Sanchez-Lengeling et al. 2021). Graph Convolution Network (i.e., denoted as GCN) and Chebyshev Network (i.e., denoted as Cheb) are two representatives of spectral-based graph convolution models. As an example, Eq. 1 presents the details of embedding computation for GCN. More specifically, $h_v^{(m)}$ is the embedding of Node *v* at step *m* and it represents the original information contained in Node *v* when *m* equals to 0. $N(v)$ represents the neighbors of Node *v* while $|N(v)|$ is the number of neighbors. $W^{(m)}$ and $B^{(m)}$ are graph convolution parameters which are used for neighboring information and the last-step embedding of Node *v* respectively. $f^{(m)}$ represents a certain activation function, such as ReLU.

$$h_v^{(m)} = f^{(m)}\left(W^{(m)} \cdot \frac{\sum_{\mu \in N(v)} h_\mu^{(m-1)}}{|N(v)|} + B^{(m)} \cdot h_v^{(m-1)}\right) \quad \text{Eq. 1}$$

By contrast, the spatial-based approach, which mimics the 2D image convolution operation, performs graph convolutions by propagating information among neighboring nodes and edges. Graph Sample and Aggregate (i.e., denoted as GraphSAGE) and Graph Attention Networks (GAT) are two widely used spatial models. Existing studies indicate that spatial-based models are theoretically more efficient, scalable and flexible than spectral-based models, especially analyzing dynamic and large-scale graphs (Zhou et al. 2020; Wu et al. 2021). Eq. 2 describes the embedding calculation methods for GraphSAGE models, where aggregation of Node v's embeddings at step m-1 are concatenated with Node v embedding at step m-1 before parameterized and activated by $W^{(m)}$ and $f^{(m)}$ respectively. Eq. 3 presents an example of dimension-wise maximum aggregation function, where σ is the Sigmoid activation function, $W_{pool}^{(m)}$ and b represents pooling weights and bias at step m. Eq. 4 describes the embedding calculation method for GAT models, where $\alpha_{v\mu}^{(m)}$ represents attention weights generated between Node v and Node μ through attention mechanisms. As shown in Eq. 5, $A^{(m)}$ represents the attention mechanism which essentially normalizes attention weights of Node v neighbors to have a sum of one. Interested readers are directed to (Daigavane et al. 2021; Sanchez-Lengeling et al. 2021) for further theoretical details.

$$h_v^{(m)} = f^{(m)}\left(W^{(m)} \cdot \left[AGG_{\mu \in N(v)}\left(\left\{h_\mu^{(m-1)}\right\}\right), h_v^{(m-1)}\right]\right) \quad \text{Eq. 2}$$

$$AGG_{\mu \in N(v)}\left(\left\{h_\mu^{(m-1)}\right\}\right) = max_{\mu \in N(v)}\{\sigma(W_{pool}^{(m)} h_\mu^{(m-1)} + b)\} \quad \text{Eq. 3}$$

$$h_v^{(m)} = f^{(m)}\left(W^{(m)} \cdot \left[\sum_{\mu \in N(v)} \alpha_{v\mu}^{(m-1)} h_\mu^{(m-1)} + \alpha_{vv}^{(m-1)} h_v^{(m-1)}\right]\right) \quad \text{Eq. 4}$$

$$\alpha_{v\mu}^{(m)} = \frac{A^{(m)}(h_v^{(m)}, h_\mu^{(m)})}{\sum_{w \in N(v)} A^{(m)}(h_v^{(m)} + h_w^{(m)})} \quad \text{Eq. 5}$$

3. Research methodology

3.1 Research outline

This research investigates the potentials of graph neural networks for semi-supervised HVAC FDD tasks. In this study, the HVAC fault diagnosis task is formulated as a semi-supervised node-level classification problem, where each node represents a data sample either with or without labeling information, and their possible linkages are described as graph edges. Unlike inductive learning approaches, such node-level task adopts a transductive learning paradigm for fault classification, where both training and testing data are utilized to construct a single

graph for inferences and the testing data are treated as unlabeled data. Compared with conventional data-driven methods, GNN-based methods have unique abilities in exploring and integrating neighboring information on intra-data structures or similarities and thereby, providing effective tools to enhancing data-driven fault classification performance.

The research outline is depicted as Fig. 5. Graph generation methods are firstly proposed to transform tabular building operational data into graphs. Data experiments are then designed to evaluate the potentials of graph neural networks for semi-supervised HVAC fault diagnosis. More specifically, both fully connected neural networks and graph convolution networks are developed to validate the advantages of GNN-based semi-supervised learning. In essence, the former corresponds to conventional induction reasoning as prediction models are developed using labeled training data only. By contrast, the latter enables semi-supervised learning through transductive reasoning, where both labeled and unlabeled nodes presented in a graph are utilized to make predictions on unlabeled testing data. In addition, this study also investigates the performance of various GNN architectures in terms of graph convolution types and hidden layers under different data availabilities, based on which useful guidelines are obtained practical applications.

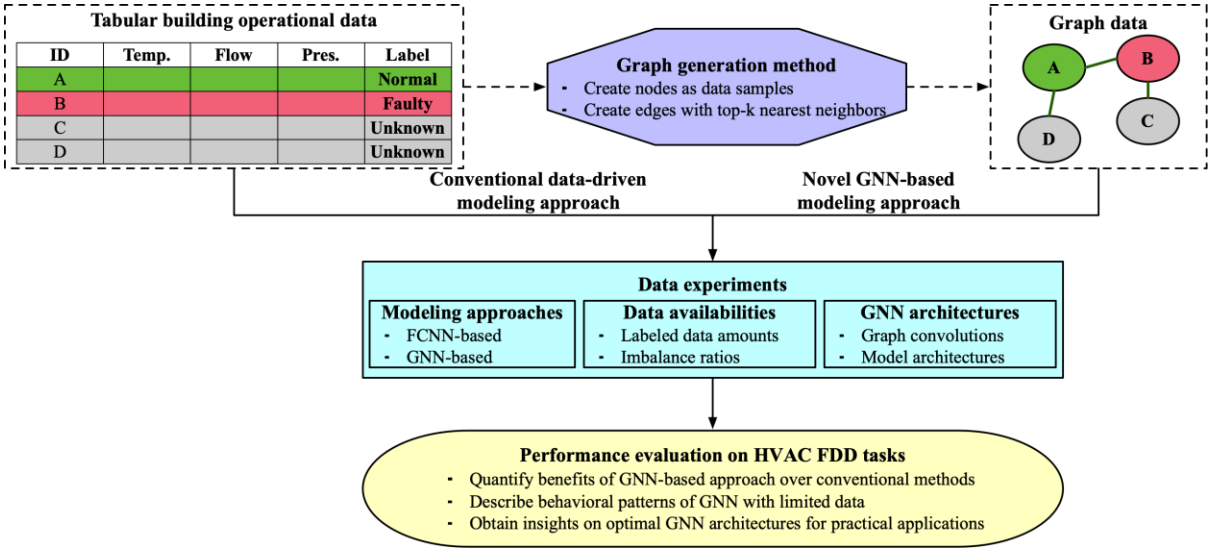

Fig. 5 Research outline

3.2 The *k*NN-based graph generation method for HVAC fault diagnosis

In this study, the *k*-nearest neighbor (i.e., *k*NN) algorithm is used to create associations among data samples and thereby, transforming tabular building operational data into graphs. It can be

summarized as a three-step method. Firstly, domain expertise or statistical methods can be applied to choose a set of input variables for *k*NN computations. Such variables should be sufficiently expressive to describe operating characteristics of system operations. For instance, the supplied/returned temperatures and flowrates of chilled and condensing water, the ambient temperature and relative humidity can be used to describe the operating conditions of water-cooled chillers. As another example, the supplied, returned and mixed air temperatures can be used to describe general operating patterns of AHUs. Secondly, as illustrated in Fig. 6, distance measures, such as the Euclidean distance, are utilized to calculate pairwise distances among data samples. Data normalization or standardization should be performed to ensure the validity of distance calculation. In addition, the number of variables used should be kept relatively small and data dimensionality reduction methods (e.g., principal component analysis) can be applied to avoid the curse of dimensionality. Thirdly, by setting a certain k value, graph edges can be generated considering the top-k similar neighbors to each data sample. It should be mentioned that such edges serve as bridges for node-level information exchange during graph convolution operations. To reduce computation burdens and avoid possible disturbances from varying operating conditions, the k value is suggested to be relatively small, e.g., 5 or 10. As an example, Fig. 6 presents the graph generated when k equals to two. The adjacency matrix is created in a symmetric manner without distinguishing edge directions and therefore, some nodes may have more than k edges, e.g., Node B. It should be mentioned that the graph created for transductive learning may become larger and larger as more data are available for analysis. In practice, it is suggested to select partial yet representative data samples for analysis given redundant and large-scale building operational data.

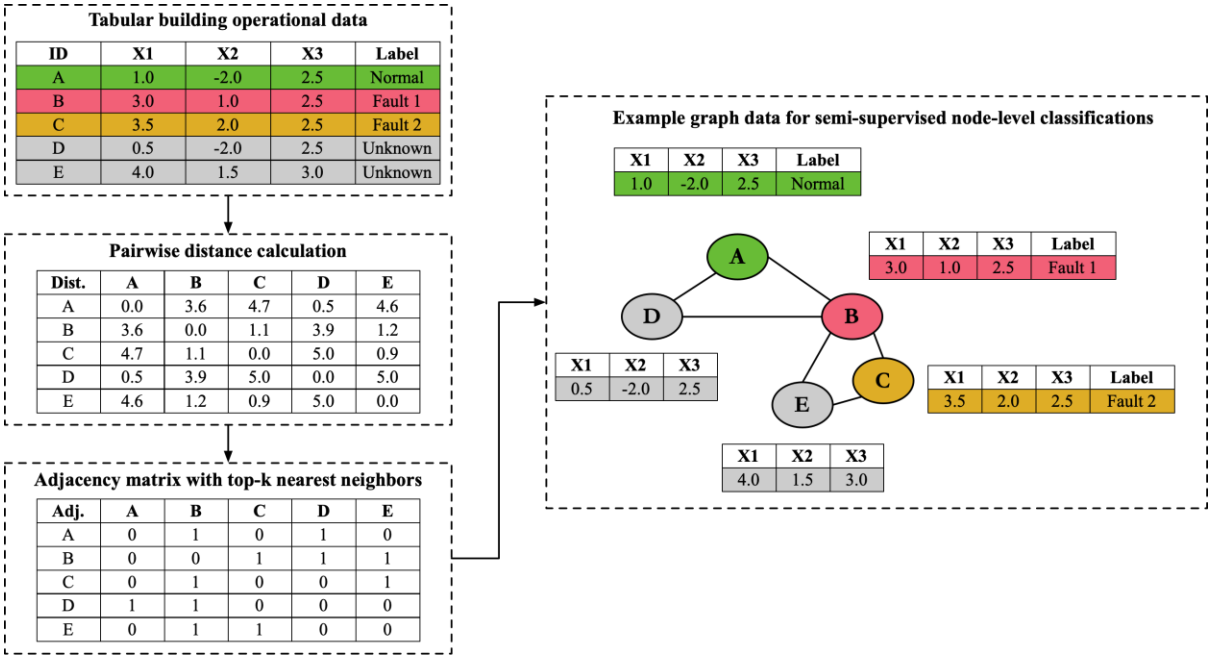

Fig. 6 The graph generation method for semi-supervised node-level classifications

3.3 Data experiments

3.3.1 Experimental data descriptions

Once graph data are generated, data experiments can be conducted to comprehensively evaluate the value of GNN-based method for HVAC fault diagnosis. To ensure the generalizability of research results, three experimental datasets describing operations of HVAC air-side systems are used for analysis. The first dataset comes from the ASHRAE 1312-RP (Wen and Li 2011) while the latter two come from the automated fault detection and diagnosis project conducted by the Lawrence Berkeley National Laboratory (Granderson and Lin 2019).

As summarized in Table-1, the first dataset, which is denoted as the ASHRAE data in this study, contains operational data collected in three testing periods with a collection interval of one-minute, i.e., the summer of 2007, and the spring and winter of 2008. In this study, 11 variables are selected as input variables for graph generation and fault modeling, including the temperatures of supplied, returned, mixed and outdoor air, the flowrates of supplied, returned and outdoor air, the water temperatures at chilled and hot water coils, and the differential pressure of supplied and returned air fans. The fault diagnosis task is formulated as a 16-class classification problem, including one *Normal* class and five major faults (i.e., damper stuck faults at exhaust air and outdoor air dampers, valve stuck or leaking faults at cooling and heating coils, and fixed speed fault at returned air fans) each with three severity levels.

The second and third datasets are denoted as the SZCAV and SZVAV, as they are experimental data collected at one-minute interval from single-zone constant air volume and variable air volume AHU systems respectively. The system operates 7-day a week with a predefined operating time schedule, i.e., 6 a.m. to 6 p.m. In total, 11 variables are selected as inputs, including the temperatures of supplied, returned, mixed and outdoor air, damper positions of returned, outdoor and exhaust air dampers, valve positions of cooling and heating coils, heating and cooling temperature setpoints of supplied air. There are 14 and 7 faults recorded in SZCAV and SZVAV datasets respectively, making it a 15-class and 8-class classification problem for fault diagnosis. All these faults take place at three components, i.e., outdoor air damper, cooling and heating coil valves. The numbers of fault classes vary as different fault severity levels were introduced during experiments.

Table-1 A summary of experimental data

Datasets	**ASHRAE**	**SZCAV**	**SZVAV**
Total data samples	34,560	10,800	7,690
Input variable numbers	11	11	11
Output class numbers	16	15	8

3.3.2 Experiment setups

To comprehensively evaluate the performance of different data-driven fault classification methods, 30 data scenarios are created considering five labeled data availabilities and six imbalance ratios. Five levels of labeled data availabilities are specified referring to 25, 50, 75, 100 and 125 data samples for each class under balanced data scenarios. The numbers of labeled data for each class may vary when different imbalance ratios (i.e., 10, 20, 30, 40 and 50) are introduced into data experiments. For instance, given a labeled data availability of 50 per class and an imbalance ratio of 10, the ASHRAE data will have $50 \times 16 = 800$ labeled training data in total, leading to $\frac{800}{10+15} = 32$ data samples per *Faulty* class and $32 \times 10 = 320$ data samples for the *Normal* class. Balanced testing datasets are selected for performance evaluation by randomly selecting 200 data samples per class. Considering that the original data were collected at one-minute intervals, it is possible to have similar data samples in both training and testing data and thereby, leading to biased accuracy metrics. In this study, the training and

testing data amounts were set relatively small compared to the overall data size, which may help to alleviate such data overlapping issues. In addition, data experiments are repeated 10 times for each data scenario with averaged performance reported to ensure the result robustness given randomness in data sampling. The training and testing data are fixed for all data-driven models during each experiment run, making it a fair game for performance evaluation. As a result, the possible biased accuracy metrics obtained may not conflict with research conclusions.

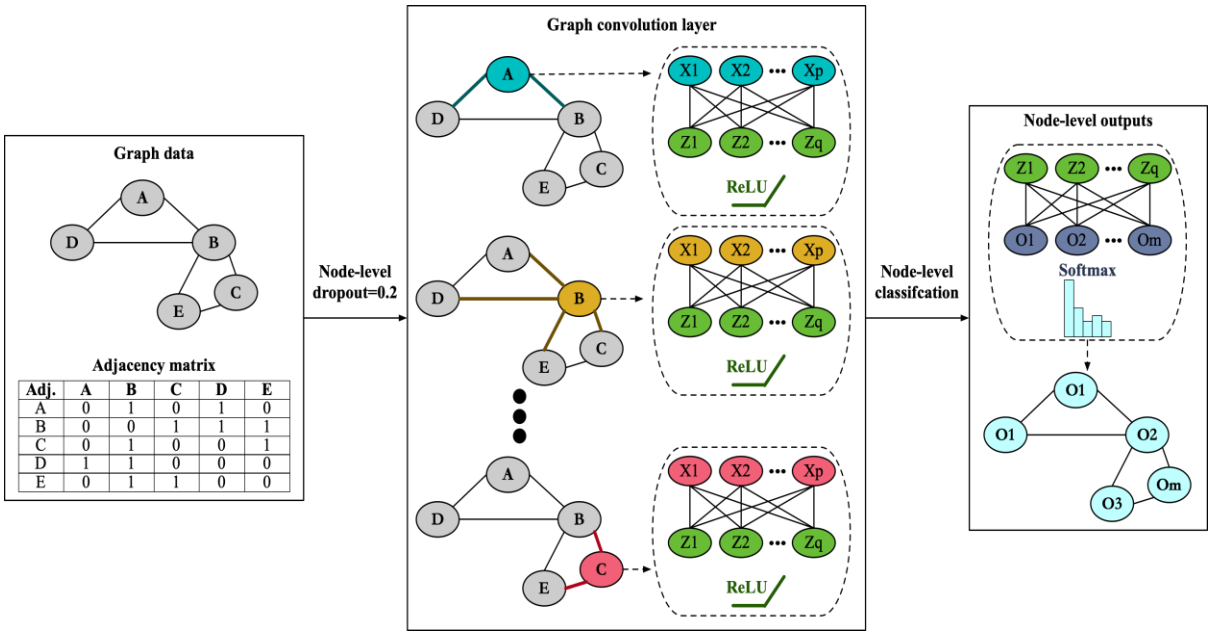

Fig. 7 Semi-supervised node-level graph convolution networks for fault classifications

As benchmarks, fully connected neural networks (i.e., FCNNs) are developed using labeled data only. Besides the four representative graph convolutions introduced in Section 2.3, another graph convolution, i.e., GCN2, which enhances the learning ability of deeper graph convolution networks through the residual connection (Chen et al. 2020), is also explored in this study. The general architecture of GNNs with one graph convolution layer is shown in Fig. 7. During experiments, all data-driven models are developed considering one to four hidden layers. The maximum training epochs is set as 2000. To avoid possible overfitting problems, all models are developed using the early-stopping training scheme and a dropout of 20% is applied after the input layer. The initial learning rate is fixed as 0.1 and the Reduce on Plateau method is used to dynamically reduce the learning rate by a factor of two for better model convergence. In addition, to alleviate the data imbalance problem, higher training weights are assigned for data samples associated with the minority class. In this study, the weights are set as the

reciprocals of class frequencies, e.g., the weights of *Faulty* data samples are 10 times larger than those of *Normal* data samples when the imbalance ratio is 10. All the data experiments and result analysis are conducted using the Pytorch Geometric library (Fey and Lenssen 2019) and the R programming language (R development core team, 2008).

4. Results and discussions

4.1 Benchmark fault diagnosis performance

4.1.1 Benchmark performance using fully connected neural networks

Figs. 8 to 10 depict the fault diagnosis performance of fully connected neural networks (i.e., FCNNs) on *ASHRAE*, *SZCAV* and *SZVAV* datasets respectively, where *N*, *IR* and *L* represent the number of training data samples per class, the imbalance ratio and the number of hidden layers. As mentioned in Section 3, the model performance is tested against balanced data and therefore, the accuracy is selected as the evaluation metric. It is evident that the increase in training data samples will enhance the fault diagnosis performance, which is expected as data-driven models typically benefit from more information-rich datasets. The results indicate that higher imbalance ratios would lead to poorer fault diagnosis performance, even though minority classes have been assigned with much higher weights during model training. For instance, given balanced training data, the fault diagnosis accuracies on the *ASHRAE* data of single hidden layer FCNNs are 80.8%, 82.9%, 83.6%, 83.7% and 84.1% when the training data amount increases from 400 to 2000 (i.e., the training data amounts range between 25 to 125 per class in balanced data scenarios, resulting in 400 to 2000 training data in total as there are 16 classes in *ASHRAE* dataset). By contrast, the fault diagnosis performance will dramatically degrade to 70.7%, 75.8%, 77.2%, 79.8% and 80.8% respectively when the imbalance ratio increases to 50. In addition, the results indicate that more hidden layers do not necessarily improve fault diagnosis performance using fully connected neural networks, especially when the training data suffer from severe imbalance data problems. Such observations are in accordance with expectations, as complicated data-driven models are more vulnerable to overfitting issues and will become less competent in identifying faults with minority nature. For instance, given an imbalance ratio of 50, the maximal and minimal fault diagnosis accuracies on the *ASHRAE* data are 80.8% and 70.7% respectively when FCNNs have one hidden layer, while decrease to 77.5% and 61.0% respectively using FCNNs with four hidden

layers. It should be mentioned that higher accuracies have been reported in existing studies using the ASHRAE data (Piscitelli et al. 2020; Yan et al. 2020b). Nevertheless, such performance differences may result from different training and testing data splits, data preprocessing techniques used and intrinsic model complexities. In this study, training data have been deliberately designed to simulate insufficient and imbalanced data scenarios for model training, while testing data are kept balanced to avoid possible biased results towards the majority *Normal* class. Rather than focusing on the absolute fault classification accuracies, this study tries to capture general model behaviors or performance trends to draw insightful conclusions for practical applications.

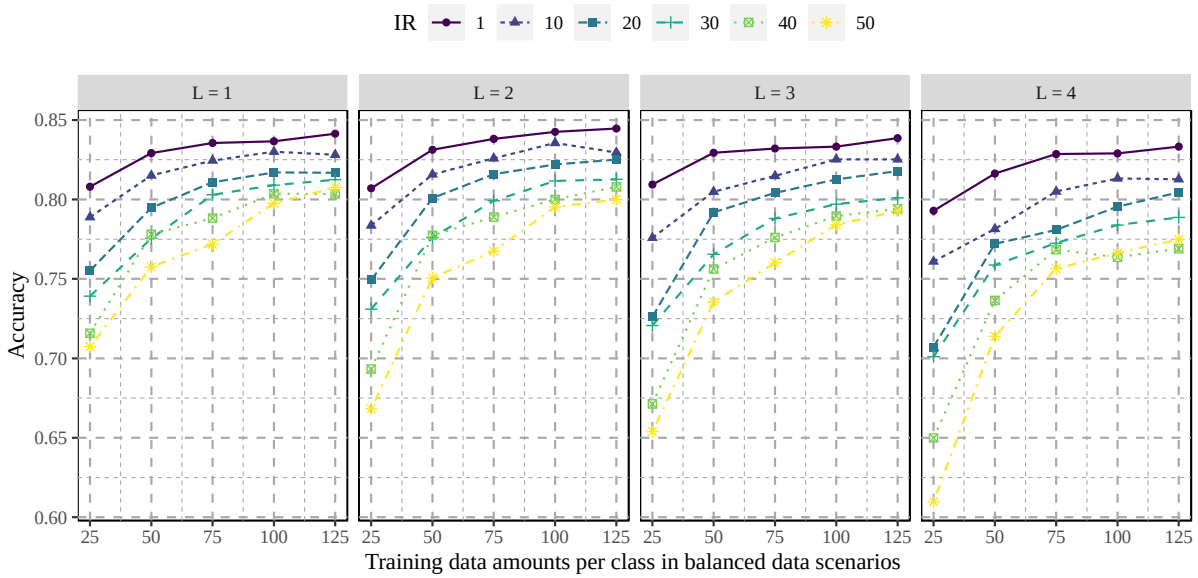

Fig. 8 The fault diagnosis performance of FCNNs on ASHRAE data

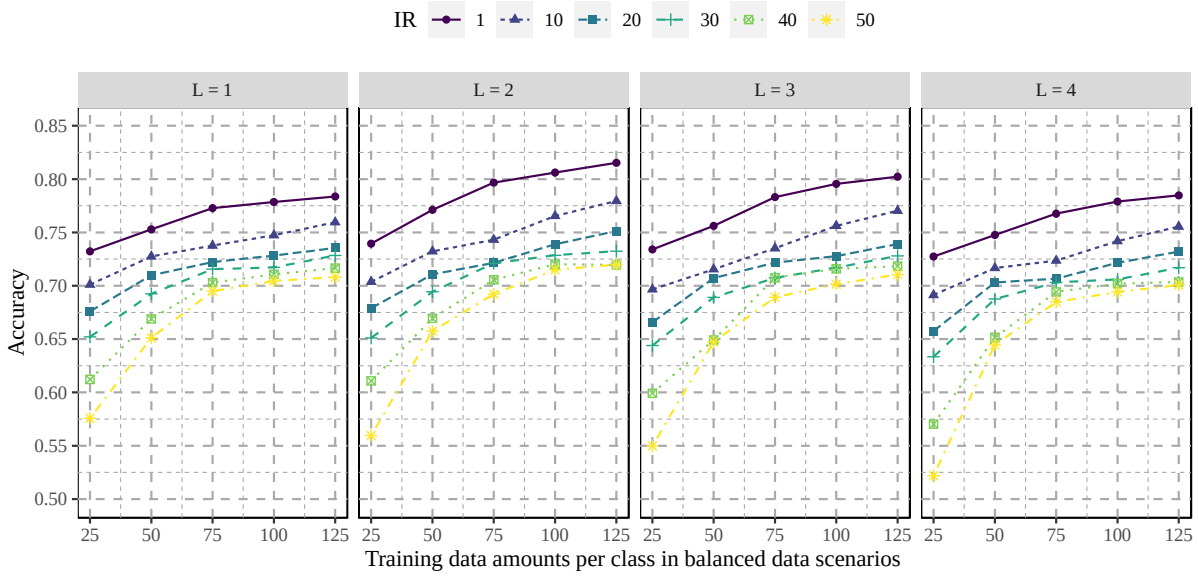

Fig. 9 The fault diagnosis performance of FCNNs on SZCAV data

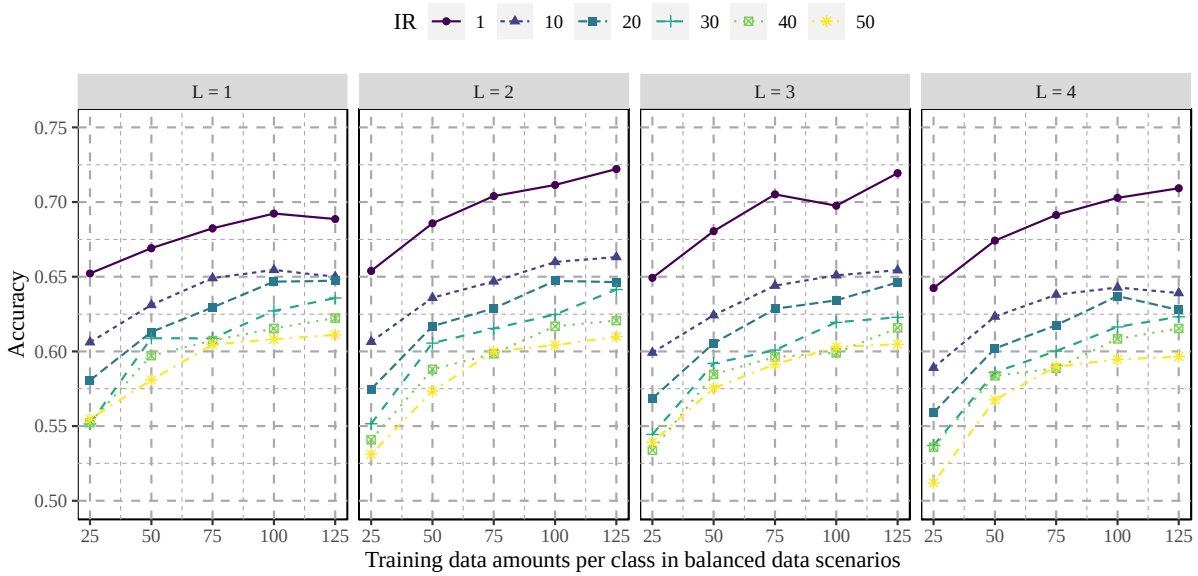

Fig. 10 The fault diagnosis performance of FCNNs on SZVAV data

To further explore FCNN behavioral patterns on fault diagnosis, multiple linear regression and Shapley values are utilized for in-depth interpretation considering possible interactions among training data amounts per class in balanced data scenarios (i.e., denoted as *N*), imbalance ratios (i.e., denoted as *IR*) and hidden layer numbers (i.e., denoted as *L*). More specifically, multiple linear regression models have been developed using fault diagnosis accuracies as dependent variables, while *N*, *IR* and *L* are used as independent variables. The resulting model coefficients are used to explain possible impacts on FCNN model performance, assuming linear relationships exist between *N*, *IR*, *L* and fault diagnosis accuracies. Table-2 summarizes model

coefficients in their original and standardized forms. The signs of model coefficients reflect correlations between input variables and fault diagnosis accuracies. The results indicate that training data amounts positively affect model accuracies, while imbalance ratios and hidden layer numbers have negative impacts on model accuracies. The P-values shown in brackets suggest that these three factors have statistically significant linear impacts given a significant level of 1%. The standardized model coefficients are helpful for comparing the effect strengths of different factors on model accuracies. It is shown that training data amounts and imbalance ratios have similar effect strengths on model accuracies, as their standardized coefficients share similar magnitudes. By contrast, the standardized coefficients of hidden layer numbers are much smaller across different datasets, indicating smaller effect strengths on FCNN model performance.

Table-2 Multiple linear regression results on FCNN model performance

Settings	**Coefficients**	**ASHRAE**	**SZCAV**	**SZVAV**
Training data amounts	Original	7.57×10^{-4} (P=0.00)	8.87×10^{-4} (P=0.00)	6.58×10^{-4} (P=0.00)
	Standardized	0.61	0.61	0.53
Imbalance ratios	Original	-1.61×10^{-3} (P=0.00)	-2.13×10^{-3} (P=0.00)	-1.93×10^{-3} (P=0.00)
	Standardized	-0.61	-0.69	-0.74
Hidden layer number	Original	-9.62×10^{-3} (P=0.00)	-4.36×10^{-3} (P=0.01)	-3.93×10^{-3} (P=0.01)
	Standardized	-0.24	-0.09	-0.10

Considering nonlinear interactions may exist between fault diagnosis accuracies and the above-mentioned experiment settings (i.e., the training data amount *N*, imbalance ratio *IR* and hidden layer number *L*), surrogate models using LightGBM have been developed for predictive modeling and interpreted through the KernelSHAP local explanation algorithm. In such a case, Shapley values obtained are used to quantify possible impacts of *N*, *IR*, *L* on FCNN model performance. More specifically, the global importance of *N*, *IR* and *L* on fault diagnosis accuracies can be calculated as the sum or mean of absolute values of their individual Shapley

values. As shown in Table-3, *N* and *IR* have similar importance on FCNN model performance across different datasets, while *L* has much smaller impacts on fault diagnosis accuracies. Figs. 11 to 13 illustrate the Shapley values of different settings using the *ASHRAE*, *SZCAV* and *SZVAV* datasets respectively. Similar behavioral patterns can be observed across different datasets, i.e., the increase in training data amounts, the decrease in imbalance ratios and hidden layer numbers typically result in higher Shapley values, referring to possible enhancements in fault diagnosis accuracies. For instance, Fig. 11 presents that Shapley values generally range between -0.04 and -0.08 when *N* is set to 25. It is in accordance with domain expertise as 25 may be too small to guarantee the generalization performance of complicated data-driven models. As *N* increases, the resulting Shapley values also increase and become the highest given *N* equals to 125, indicating that FCNN model accuracies do benefit from more training data.

Table-3 Shapley value-based global importance of different settings on FCNN performance

Shapley importance	**ASHRAE**	**SZCAV**	**SZVAV**
Training data amounts	2.92	3.44	2.47
Imbalance ratios	2.82	3.64	3.39
Hidden layer numbers	1.24	0.81	0.61

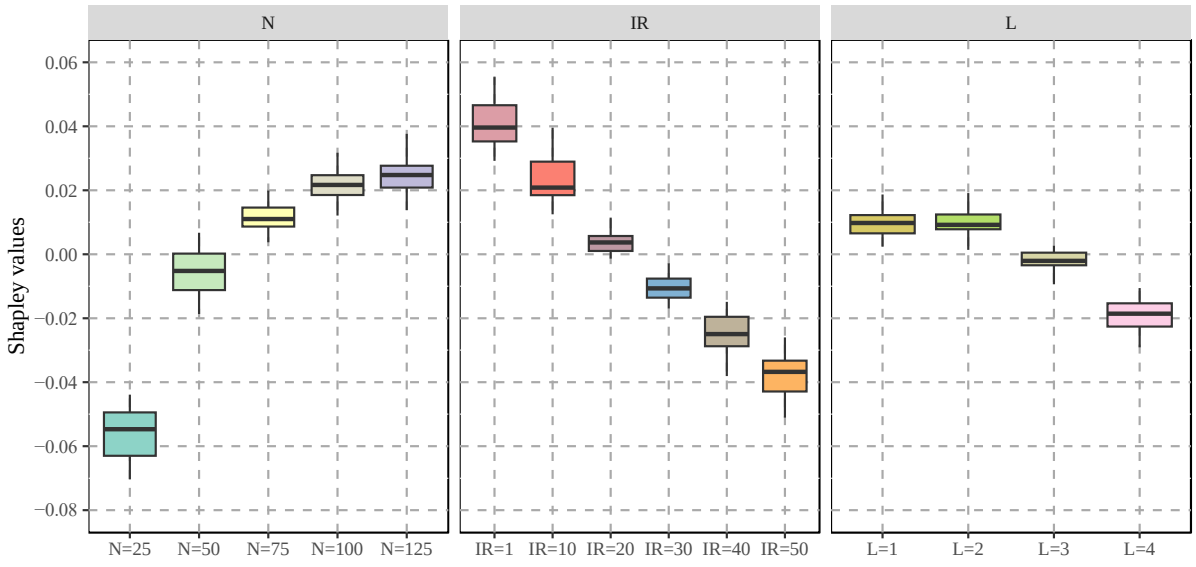

Fig. 11 Shapley values for FCNN model interpretation using *ASHRAE* data

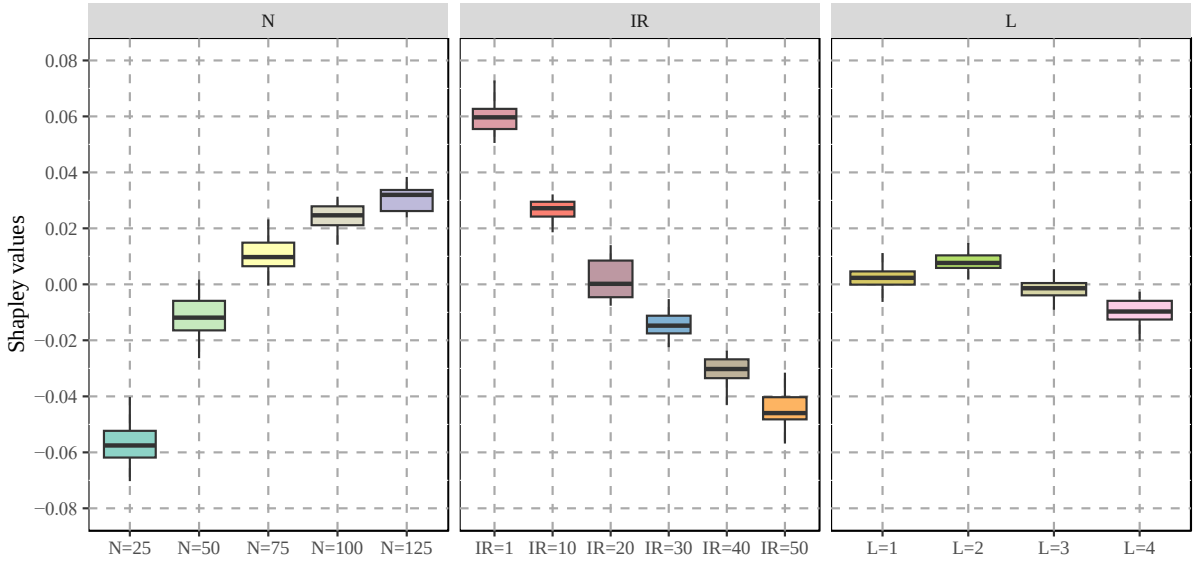

Fig. 12 Shapley values for FCNN model interpretation using *SZCAV* data

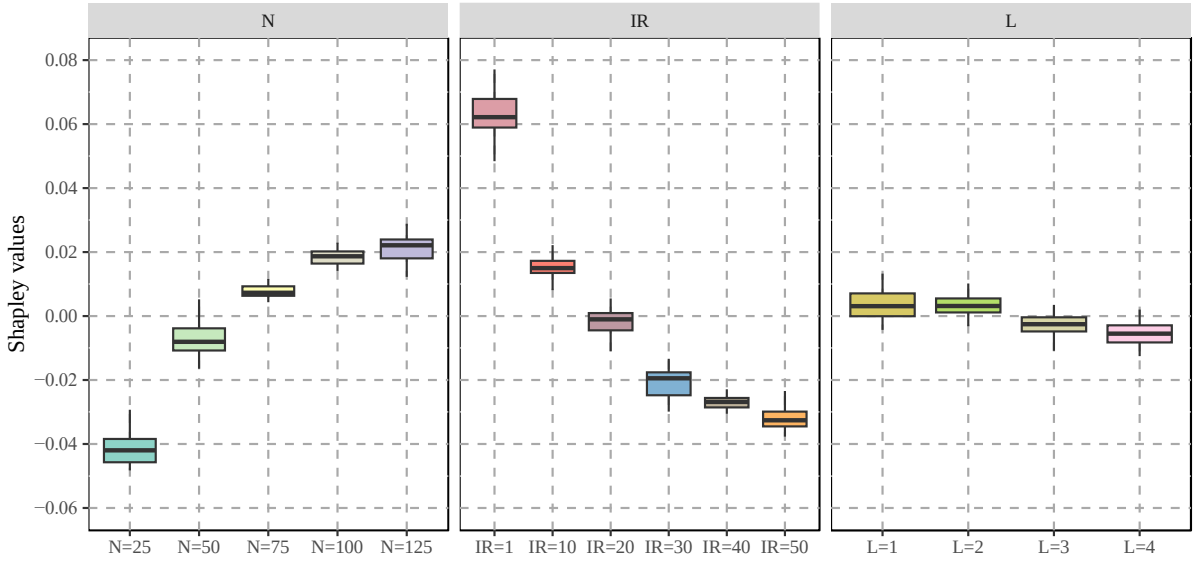

Fig. 13 Shapley values for FCNN model interpretation using *SZVAV* data

4.1.2 Benchmark performance using *k*-nearest neighbor ensemble models

The *k*-nearest neighbor (i.e., *k*NN) algorithm is one of the most classic non-parametric classification algorithms which utilizes neighboring information for predictions. As introduced in Section 3.2, *k*NN is used to generate associations or linkages among data samples and thereby, providing both self- and neighboring-information for graph neural networks. In such a case, it is intuitive to use *k*NN as a benchmark algorithm for performance comparison. In this study, ensemble *k*NN models based on the majority voting scheme have been developed for fault diagnosis considering four different *k* values of 5, 10, 15 and 20. Fig. 14 compares the

performance of *k*NN ensemble models and FCNN models with one hidden layer. In general, FCNN models are more robust against data scarce contexts with limited data and high imbalance ratios. Taking the ASHRAE data with an imbalance ratio of 50 for examples, the FCNN fault diagnosis accuracies are 80.7% and 70.7% given 125 and 25 training data per class, which are much higher than the corresponding accuracies of *k*NN ensemble models, i.e., 67.1% and 24.7% respectively. Nevertheless, *k*NN ensemble models do have better performance given sufficient training data with smaller imbalance ratios. For instance, when the training data are balanced (i.e., *IR*=1), *k*NN ensemble typically perform better than the single hidden layer FCNN models across all three datasets. It indicates that neighboring information can be very helpful for fault diagnosis, yet it is highly sensitive to training data quality and the performance may degrades dramatically given insufficient or imbalanced training data. Considering that data environment in practice may vary greatly, parametric data-driven methods, such as artificial neural networks, may provide more reliable and robust results.

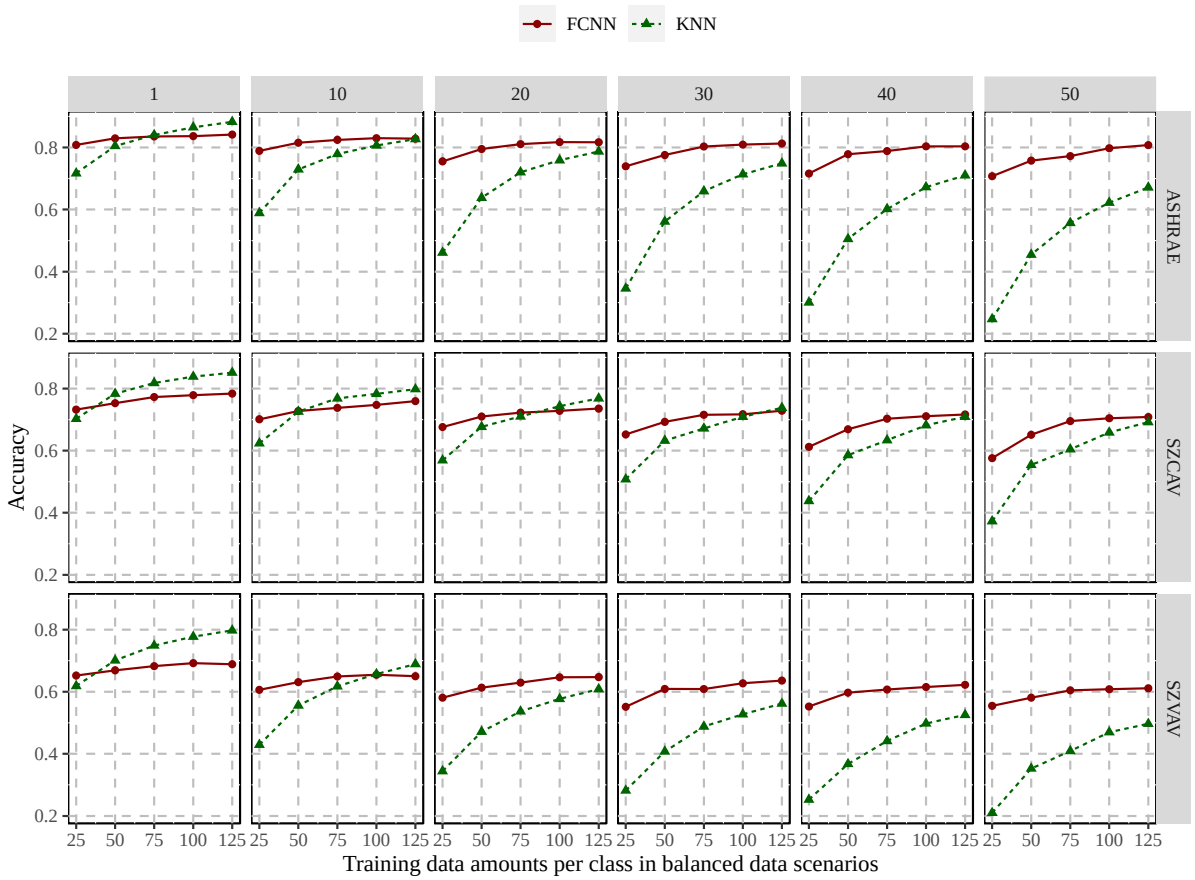

Fig. 14 Fault diagnosis performance comparison between FCNN and *k*NN methods

4.2 Fault diagnosis performance using graph neural networks

4.2.1 General behavioral patterns of graph neural networks

Figs. 15 to 17 describe general behavioral patterns of various GNNs with changes in training data amounts per class, imbalance ratios and hidden layer numbers. As shown in Fig. 15, the fault diagnosis accuracies increase with enlargements in training data amounts across different datasets. It is observed that Cheb and SAGE models typically result in the best performance, while GCN2 performs the worst in terms of fault classification accuracies. In addition, GAT models, which adopt rather complex attention mechanisms and have the largest number of model parameters, do not necessarily lead to better performance. Fig. 16 indicates that GNN models suffer from imbalance data problems, and higher imbalance ratios will lead to poorer generalization performance. Fig. 17 presents the general relationships between fault diagnosis accuracies and the number of graph convolution layers. It is shown that optimal performance can be obtained when setting the number graph convolution layers to be two for most graph convolution types. Further increases in graph convolution layers will not only deteriorate model performance, but also impose extra burden in computational costs. Such observations are in accordance with domain expertise, as the increase in graph convolution layers will cause the over-smoothing problem, i.e., the intrinsic information of a node is being diluted by its neighboring information through graph convolutions. Nevertheless, the results indicate that GCN2 models do benefit from increasing graph convolution layers. The reason behind is that GCN2 adopts two simple yet effective techniques, i.e., initial residual and identity mapping, to overcome the over-smoothing problem in graph modeling, sharing a similar concept of the famous ResNet architecture in the deep learning field (He et al. 2015; Chen et al. 2020). Higher classification accuracies may be possible given different k values for kNN-based graph generation. Such hyperparameter can be essential as it determines the graph layout for fault inferences. Further studies can be conducted to systematically investigate the optimal values considering the trade-off between graph complexity and fault diagnosis performance.

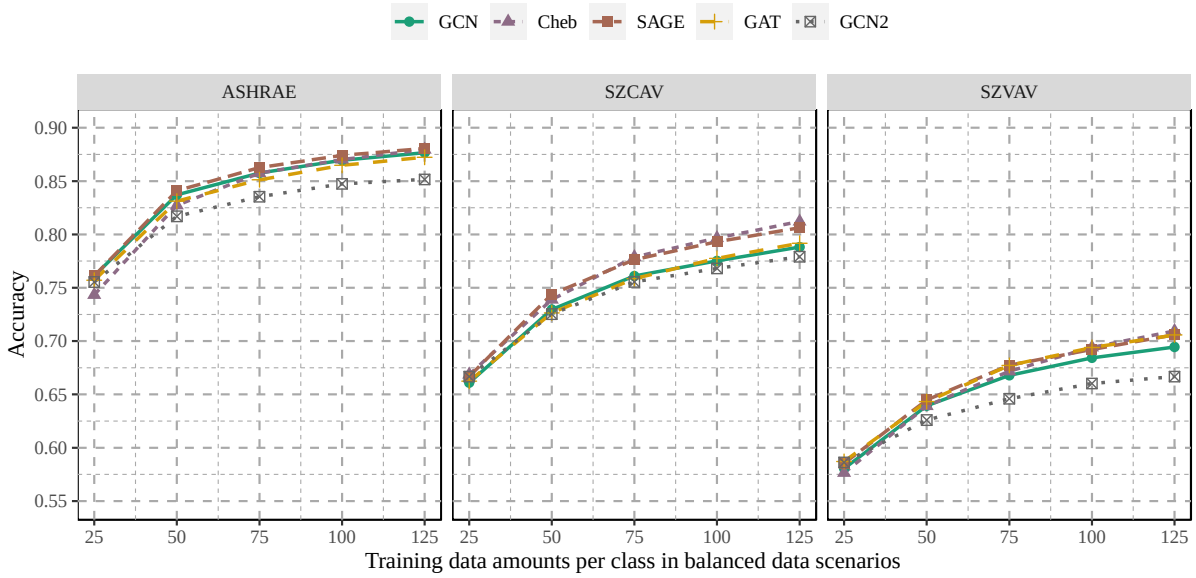

Fig. 15 The fault diagnosis performance of GNNs considering different training data amounts

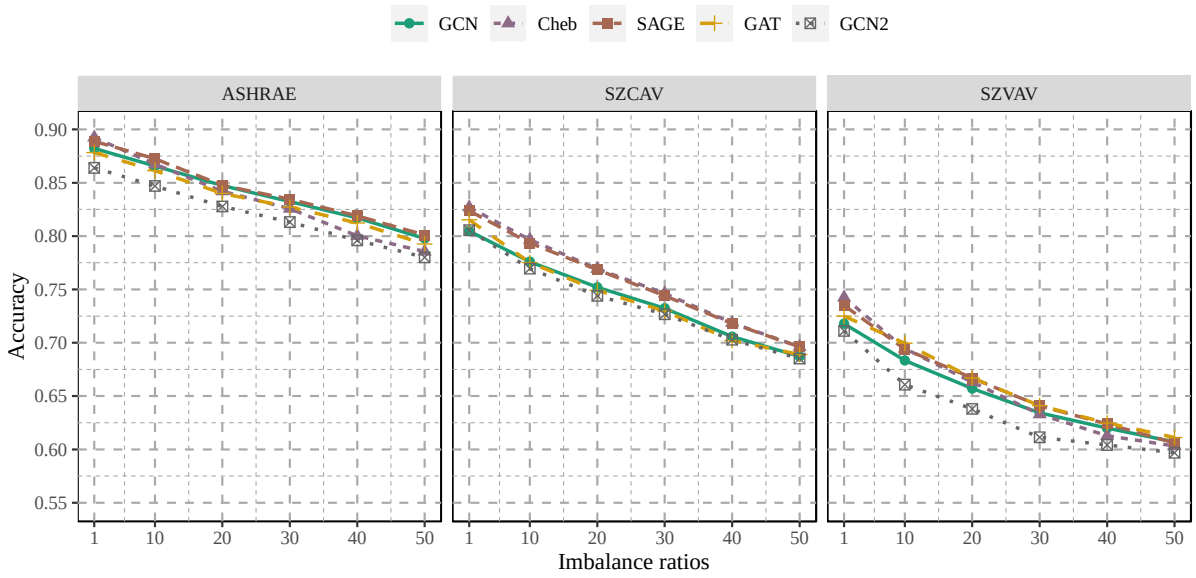

Fig. 16 The fault diagnosis performance of GNNs considering different imbalance ratios

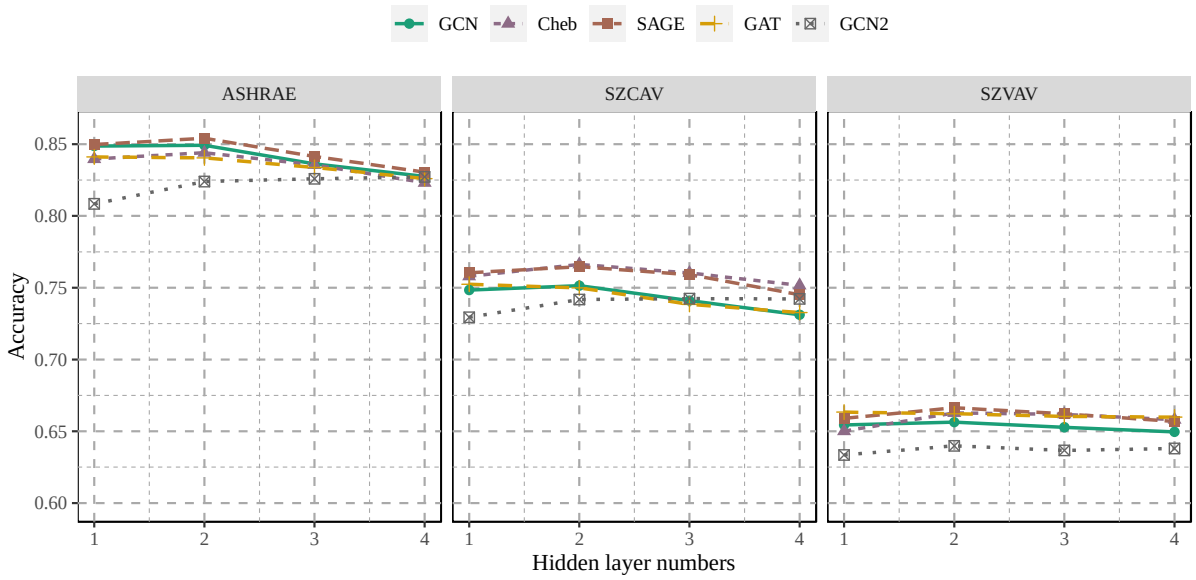

Fig. 17 The fault diagnosis performance of GNNs considering different hidden layer numbers

4.2.2 Shapley value-based methods for in-depth GNN model interpretations

Similar to Section 4.1.1, either multiple linear regression or Shapley values can be used to interpret possible impacts of training data amounts, imbalance ratios, graph convolution layers and graph convolution types on GNN model performance. In this study, initial attempts using multiple linear regression yield contradicting interpretation results across different datasets, indicating that nonlinear relationships do exist between GNN model performance and different experiment settings. Therefore, the Shapley value-based interpretation method is used for more accurate and reliable interpretations.

More specifically, surrogate LightGBM models are firstly developed, where the GNN fault diagnosis accuracies are used as the model output, and the training data amounts, imbalance ratios, graph convolution layers and graph convolution types are used as model inputs. Local Shapley values can then be obtained using the KernelSHAP algorithm to describe the influences of training data amount N, imbalance ratio IR, graph convolution layer number L and types on fault classification accuracies. The results obtained across different datasets are illustrated using boxplots and violin plots as shown in Figs. 18-20, where positive and negative Shapley values indicate marginal effects on fault classification accuracies compared with the average performance. As summarized in Table-4, the global importance of different settings can be calculated based on the sum of their absolute Shapley values. The results indicate that the training data amounts N and imbalance ratio IR have the largest and comparable effect

strengths on GNN model performance, while the effect strengths of graph convolution layers and graph convolution types are less significant. Clear performance trends can be observed among different settings on training data amounts and imbalance ratios, indicating that the increase in training data amounts and decrease in imbalance ratios typically lead to higher fault diagnosis accuracies. Most graph convolutional networks perform the best with no more than two graph convolution layers, and additional increases in graph convolution layers will lead to decreases in model performance. The performance of different graph convolution types is less consistent across different datasets. To summarize, SAGE models generally result in the best performance, while GCN2 models perform the worst in all three datasets. GAT models, which have larger parameter numbers due to the adoption of the multi-head attention mechanism, do not present clear performance edges over conventional GCN and Cheb models. The results indicate that SAGE, Cheb and GCN models are sufficient and more cost-effective to describe faulty operations in building operational data.

As shown above, Shapely values have been used in this study to investigate the influence of training data, graph convolutional neural network types and architectures on fault classification accuracies. In practice, it is desired to implement model-specific or model-agnostic methods to provide either local or global explanations for interpreting data-driven fault diagnosis models (Machlev et al. 2022; Chen et al. 2023). Such interpretations are helpful for building professionals to understand the inference mechanisms learnt from data, while inspecting model validity with subjective domain expertise. Graph neural networks can be interpreted using either conventional model-agnostic methods (e.g., local interpretable model-agnostic explanations or LIME (Ribeiro et al. 2016) and Shapley additive explanation or SHAP (Lundberg and Lee 2017)) or model-specific methods (e.g., gradient-based and class activation-based interpretations (Pope et al. 2019)).

Table-4 Shapley value-based global importance of different settings on GNN performance

Shapley importance	**ASHRAE**	**SZCAV**	**SZVAV**
Training data amounts	20.36	23.94	20.70
Imbalance ratios	15.67	21.84	22.00
Hidden layer numbers	3.92	3.69	2.69

Graph convolution types	4.66	5.13	5.18

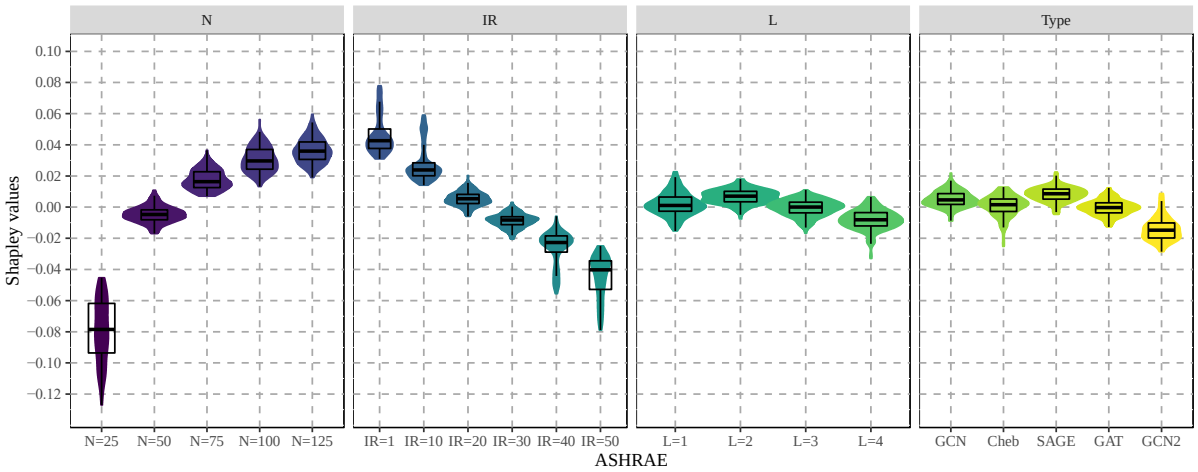

Fig. 18 Shapley values for GNN model interpretation using ASHRAE data

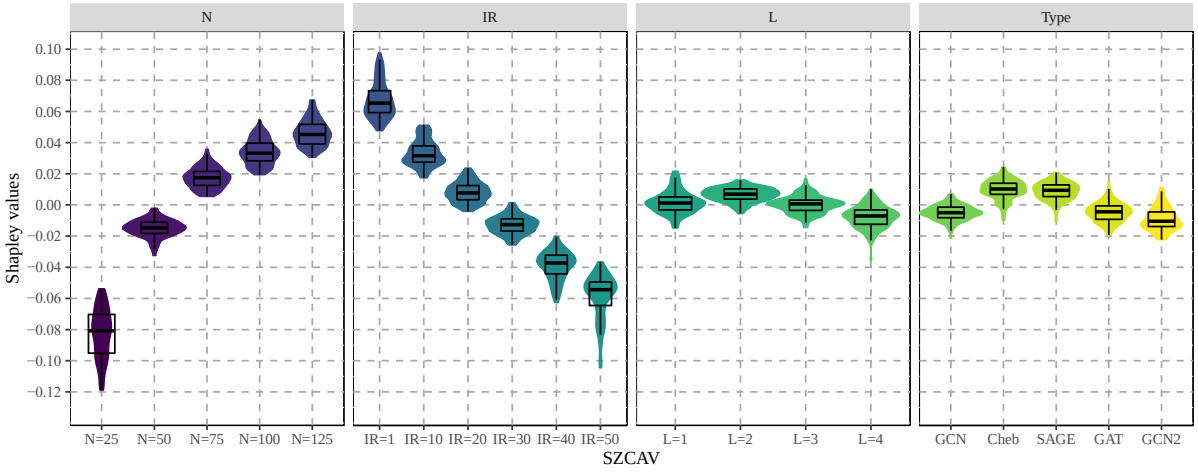

Fig. 19 Shapley values for GNN model interpretation using SZCAV data

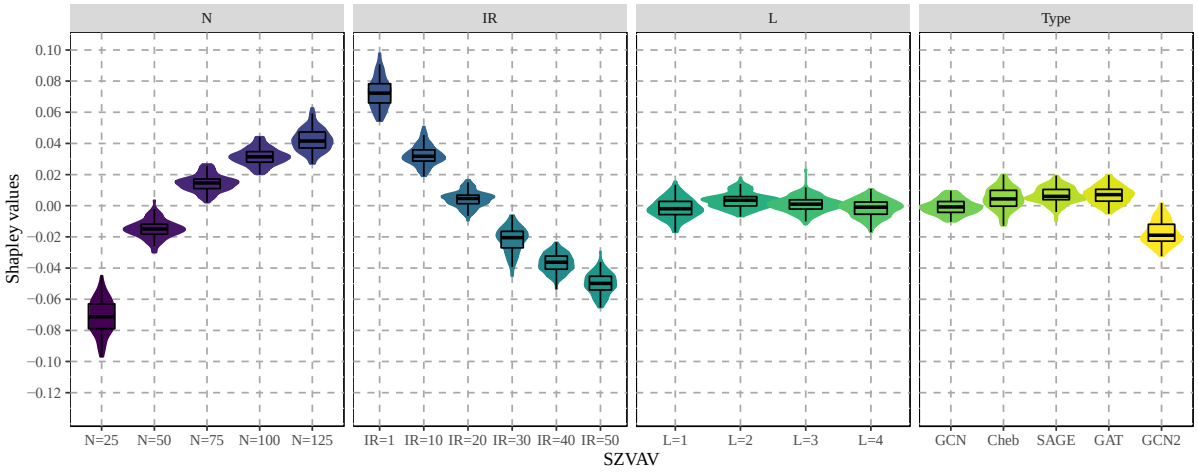

Fig. 20 Shapley values for GNN model interpretation using SZVAV data

4.2.3 Performance comparison between GNN and FCNN models

In this study, the performance improvement ratio (i.e., PIR) is formulated to quantify the performance gap between GNN and FCNN models. Fig. 21 plots the trend of average PIRs of different GNNs considering different training data amounts and imbalance ratios. The average PIRs are all positive when the imbalance ratio is no more than 20, even though the values can be rather small given limited training data amounts (e.g., N=25). Negative values of average PIRs can be observed given extreme data scenarios, where training data amount is small and imbalance ratio is larger than 20. Table-5 summarizes the mean PIR values across data scenarios with different training data amounts and imbalance ratios. The results show that GNN are helpful for enhancing fault diagnosis accuracies with potential performance boosts of at least 4% in most data scenarios. Taking GCN2 as an example, even though it performs the worst among various graph convolution types, the resulting PIRs are still positive with mean values of 4.51%, 4.13% and 2.86% on ASHRAE, SZCAV and SZVAV datasets respectively. Meanwhile, the better GNNs, such as SAGE and Cheb models, can lead to more than 6.0% performance improvement ratios across three datasets.

Table-5 A summary of PIR means across different data scenarios

PIRs	GCN	Cheb	SAGE	GAT	GCN2
ASHRAE	6.87%	6.17%	7.30%	6.23%	4.51%
SZCAV	4.66%	6.85%	6.62%	4.71%	4.13%
SZVAV	5.42%	6.08%	6.63%	6.72%	2.86%

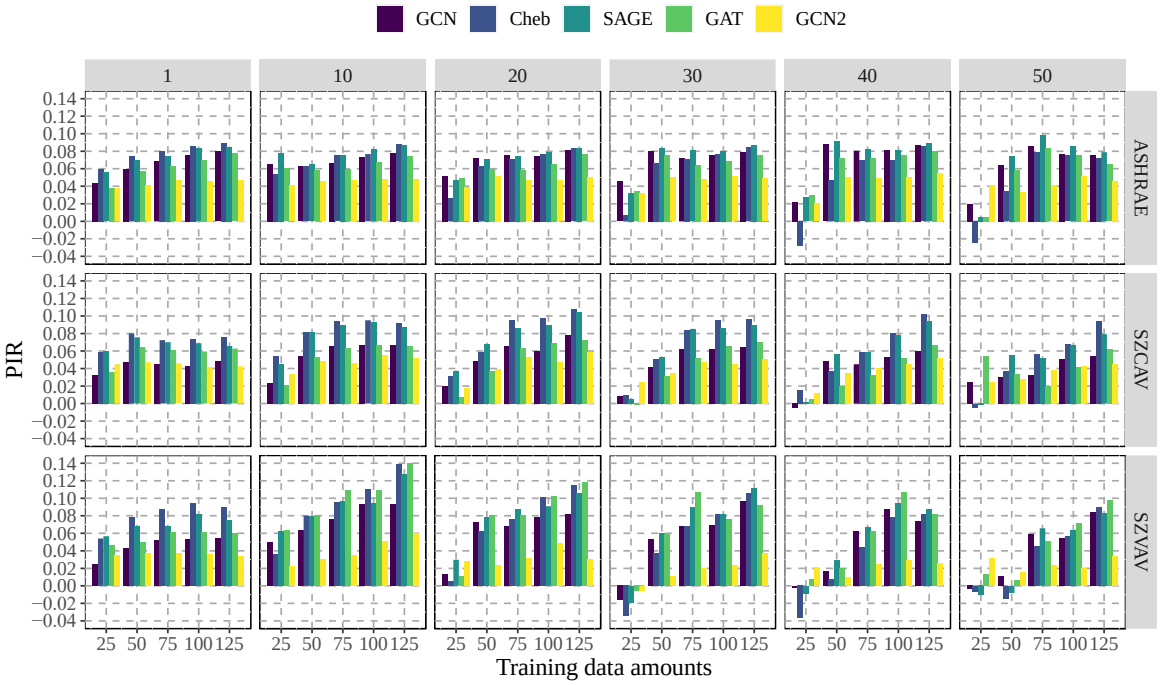

Fig. 21 Visualization of average performance improvement ratios of GNN models

As summarized in Table-6, to make fair performance comparisons, model complexities have been expressed in terms of total parameter numbers across different datasets. Assuming each hidden layer has the same number of hidden neurons (i.e., denoted as H), the total parameter number of FCNN models can be expressed as $H(I+1)+H(H+1)(L-1)+O(H+1)$, where I and O denote the input and output neuron numbers, the first (i.e., $H(I+1)$) and third (i.e., $O(H+1)$) terms represent the weight and bias numbers at input and output layers, and the second term $H(H+1)(L-1)$ represents the parameter number of L hidden layers. GCNs have the same parameter number as FCNNs as they do not introduce additional parameters neighboring information aggregation. The parameter number of Cheb models is $H(KI+1)+(H^2K+H)(L-1)+O(H+1)$, where K represents the Chebyshev filter size which was set as two in this study. SAGE models use two different groups of parameters for updating self- and neighboring-information except for the output layer, resulting in a total parameter number of $H(2I+1)+H(2H+1)(L-1)+O(H+1)$. GCN2 models operate with initial residual connections and identity mapping, and the resulting parameter number is $H(I+1)+H^2L+O(H+1)$. The parameter number of GAT can be very large due to the use of multi-head attentions. It is worth mentioning that except for GAT, all the other GNN models have similar levels of complexity with FCNNs in terms of parameter numbers. The essential difference is

that FCNN treats each data sample as an independent observation, while GNN also considers its neighboring data samples for predictive modeling. In other words, GNNs enable a novel approach to effectively integrating both "self-information" and "neighboring-information" in tabular data for fault diagnosis. Such graph convolution operations can be readily performed without requiring the neighboring nodes to be "labeled", making it more suitable to analyze both labeled and unlabeled data for semi-supervised learning tasks in the building field.

Table- A summary of model complexity in terms of total parameter numbers

Parameter numbers	**Hidden layers**	**FCNN**	**GCN**	**Cheb**	**SAGE**	**GAT**	**GCN2**
ASHRAE	1	856	856	1,186	1,186	2,536	1,756
	2	1,789	1,789	3,016	3,016	9,976	2,656
	3	2,716	2,716	4,846	4,846	17,416	3,556
	4	3,646	3,646	6,676	6,676	24,856	4,456
SZCAV	1	825	825	1,155	1,155	2,475	1,725
	2	1,755	1,755	2,985	2,985	9,915	2,625
	3	2,685	2,685	4,815	4,815	17,355	3,525
	4	3,615	3,615	6,645	6,645	24,795	4,425
SZVAV	1	608	608	938	938	2,048	1,508
	2	1,538	1,538	2,768	2,768	9,488	2,408
	3	2,468	2,468	4,598	4,598	16,928	3,308
	4	3,398	3,398	6,428	6,428	24,368	4,208

5. Conclusions

HVAC fault detection and diagnosis is an essential task in building energy management, as it is closely related to the efficiency and safety of system operations. Reliable data-driven fault diagnosis methods are of great significance to facilitate decision makings in smart building operations. To tackle the labeled data shortage challenge in the building field, semi-supervised solutions have been proposed to enhance the quality of data-driven models. This study explores the potential of graph convolutional networks in HVAC fault diagnosis tasks, where the fault diagnosis problem is formulated as a transductive node-level classification problem. More

specifically, a graph generation method has been proposed to transform tabular building operational data into graphs using *k*-nearest neighbors and thereby, enabling fault inferences based on both "self-information" and "neighboring-information". Data experiments have been conducted to obtained behavioral patterns of graph neural networks considering different data availabilities and model architectures. The performance of graph convolutional networks has been evaluated using three HVAC operational datasets and compared with conventional fully connected neural networks and nonparametric *k*-nearest neighbor methods. The main findings are summarized as follows.

(1) Graph convolutional networks provide a novel approach to semi-supervised fault diagnosis, as it can effectively utilize neighboring information, either labeled or unlabeled, for fault inferences. The results indicate that graph convolutional network-based methods can achieve higher fault diagnosis accuracies than conventional machine learning models with similar complexities. The average performance improvement ratios are mostly positive and range between 2.86% to 7.30% in different data scenarios. It is worth mentioning that negative performance improvement ratios can be observed in extreme training data contexts with limited labeled data samples and high imbalance ratios.

(2) The general behavioral patterns of graph convolutional networks in fault diagnosis tasks have been obtained through data experiments. To summarize, the fault diagnosis accuracies are higher given larger labeled training data amounts and lower imbalance ratios. In addition, the increase in graph convolution layers does not necessarily enhance model performance, and most graph neural networks perform the best using one or two graph convolution layers only. Such findings are in accordance with results reported in other field, as excessive graph convolutions may cause the over-smoothing problem, i.e., the intrinsic information of a node is being diluted by its neighboring information.

(3) This study compares the performance of five graph convolution operations. The results show that SAGE models perform the best, while GCN2 models with residual connections performs the worst across all three HVAC datasets. GAT models, which have the highest model complexity due to the use of multi-head attention mechanisms, do not present clear performance edge over the other two graph convolution types (i.e., GCN and Cheb). In practice, it is suggested to use SAGE, GCN and Cheb models with less than two graph

convolution layers to ensure the cost-effectiveness of HVAC fault diagnosis tasks.

This study serves an exploration to leverage graph neural networks for building operational data analysis. The methods proposed are helpful for the effective utilization of heterogenous or multi-relational information in building operations. Further studies can be conducted to investigate the influences of different graph generation methods on transductive or inductive graph learning, and the performance of graph neural networks for other building management tasks.

Acknowledgements

The authors gratefully acknowledge the support of this research by the National Natural Science Foundation of China (No. 52278117), the Philosophical and Social Science Program of Guangdong Province, China (GD22XGL20) and the Shenzhen Science and Technology Program (No. 20220531101800001 and 20220810160221001).

References

Chakraborty D., Elzarka H. Early detection of faults in HVAC systems using an XGBoost model with a dynamic threshold. Energy and Buildings, 2019, 185:326-344.

Chen J.L., Zhang L., Li Y.F., Shi Y.F., Gao X.H., Hu Y.Q. A review of computing-based automated fault detection and diagnosis of heating, ventilation and air conditioning systems. Renewable and Sustainable Energy Reviews, 2022, 161:112395.

Chen M, Wei Z, Huang Z, Ding B, Li Y. Simple and deep graph convolution networks. In Proceedings of the 37th International Conference on Machine Learning, 2020, Vienna, Austria.

Chen Z., Xiao F., Guo F.Z., Yan J.Y. Interpretable machine learning for building energy management: A state-of-the-art review. Advances in Applied Energy, 2023, 9:100123.

Daigavane A. Ravindran B., Aggarwal G. Understanding convolutions on graphs. Distill, 2021, DOI: 10.23915/distill.00032.

Fan C., He W.L., Liu Y.C., Xue P., Zhao Y.P. A novel image-based transfer learning framework for cross-domain HVAC fault diagnosis: From multi-source data integration to knowledge sharing strategies. Energy and Buildings, 2022, 262:111995.

Fan C., Li X.Q., Zhao Y., Wang J.Y. Quantitative assessments on advanced data synthesis strategies for enhancing imbalanced AHU fault diagnosis performance. Energy and Buildings,

2021d, 252:111423.

Fan C., Liu X.Y., Xue P., Wang J.Y. Statistical characterization of semi-supervised neural networks for fault detection and diagnosis of air handling units. Energy and Buildings, 2021c, 234:110733.

Fan C., Liu Y.C., Liu X.Y., Sun Y.J., Wang J.Y. A study on semi-supervised learning in enhancing performance of AHU unseen fault detection with limited labeled data. Sustainable Cities and Society, 2021b, 70:102874.

Fan C., Yan D., Xiao F., Li A., An J.J., Kang X.Y. Advanced data analytics for enhancing building performances: Form data-driven to big data-driven approaches. Building Simulation, 2021a, 14:3-24.

Fan Y.Q., Cui X.Y., Han H., Lu X.L. Chiller fault diagnosis with field sensors using technology of imbalanced data. Applied Thermal Engineering, 2019, 159:113933.

Fey M., Lenssen J.E. Fast graph representation learning with PyTorch Geometric. URL: https://github.com/pyg-team/pytorch_geometric. 2019, arXiv:1903.02428.

Gao Y., Han H., Ren Z.X., Gao J.Q., Jiang S.X., Yang Y.T. Comprehensive study on sensitive parameters for chiller fault diagnosis. Energy and Buildings, 2022, 251:111318.

Granderson J., Lin G.J. Inventory of data sets for AFDD evaluation. Lawrence Berkeley National Laboratory, Feb 2019, USA.

Grover A., Leskovec J. Node2vec: Scalable Feature Learning for Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, Association for Computing Machinery.

Han H., Cui X, Fan Y., Qing H. Least squares support vector machine-based chiller fault diagnosis using fault indicative features. Applied Thermal Engineering, 2019, 154:540-547.

He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2015, arXiv:1512.03385.

Jiang W., Luo J. Graph neural network for traffic forecasting: A survey. Expert Systems with Applications, 2022, 207:117921.

Keramatfar A., Rafiee M., Amirkhani H. Graph neural networks: A bibliometrics overview. Machine Learning with Applications, 2022, 10:100401.

Li B.X., Cheng F.Y., Cai H., Zhang X., Cai W.J. A semi-supervised approach to fault

detection and diagnosis for building HVAC systems based on the modified generative adversarial network. Energy and Buildings, 2021, 246:111044.

Li G., Chen H., Hu Y., Wang J., Guo Y., Liu J., Li R., Huang R., Lv H., Li J. An improved decision tree-based fault diagnosis method for practical variable refrigerant flow system using virtual sensor-based fault indicators. Applied Thermal Engineering, 2018, 129:1292-1303.

Li S., Wen J. A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform. Energy and Buildings, 2014, 68:63-71.

Li T., Zhou Z., Li S., Sun C., Yan R., Chen X. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. Mechanical Systems and Signal Processing, 2022, 168:108653.

Li Y.F., O'Neill Z. A critical review of fault modeling of HVAC systems in buildings. Building Simulation, 2018, 11:953-975.

Liu J.Y., Zhang Q., Li X., Li G.N., Liu Z.M., Xie Y., Li K.N., Liu B. Transfer learning-based strategies for fault diagnosis in building energy systems. Energy and Buildings, 2021, 250:111256.

Lundberg S., Lee S.I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, 4768-4777.

Machlev R., Heistrene L., Perl M., Levy K.Y., Belikov J., Mannor S., Levron Y. Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. Energy and AI, 2022, 9:100169.

Mirnaghi M.S., Haghighat F. Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: A comprehensive review. Energy and Buildings, 2020, 229:110492.

Piscitelli, M.S., Brandi, S., Capozzoli, A., Xiao, F. A data analytics-based tool for the detection and diagnosis of anomalous daily energy patterns in buildings. Building Simulation, 2021, 14(1), pp. 131–147

Piscitelli M.S., Mazzarelli D.M., Capozzoli A. Enhancing operational performance of AHUs through an advanced fault detection and diagnosis process based on temporal association and decision rules. Energy and Buildings, 2020, 226:110369.

Pope P.E., Kolouri S., Rostami M., Martin C.E., Hoffmann H. Explanability methods for graph convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, 10764-10773, USA.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, ISBN 3-900051-070; 2008. URL: http://www.R-project.org.

Ribeiro M.T., Singh S., Guestrin C. Why should I trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, 1135-1144.

Sanchez-Lengeling B., Reif E., Pearce A., Wiltschko A.B. A gentle introduction to graph neural networks, Distill, 2021, DOI: 10.23915/distill.00033.

Vishwanathan S., Schraudolph N.N., Kondor R., Borgwardt K.M. Graph kernels. Journal of Machine Learning Research, 2010, 11(40):1201-1242.

Wen J., Li S. Tools for evaluating fault detection and diagnostic methods for air-handling units. ASHRAE RP-1312 Final Report, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, USA, 2011.

Wu Z.H., Pan S.R., Chen F.W., Long G.D., Zhang C.Q., Yu S. A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1):4-24.

Xia Y.D., Ding Q., Li Z., Jiang A. Fault detection for centrifugal chillers using a kernel entropy component analysis (KECA) method. Building Simulation, 2021, 14:53-61.

Yan K., Chong A., Mo Y. Generative adversarial network for fault detection diagnosis of chillers. Building and Environment, 2020a, 172:106698.

Yan K., Huang J., Shen W., Ji Z. Unsupervised learning for fault detection and diagnosis of air handling units. Energy and Buildings, 2020b, 210:109689.

Yan K., Zhong C.W., Ji Z.W., Huang J. Semi-supervised learning for early detection and diagnosis of various air handling unit faults. Energy and Buildings, 2018, 181: 75-83.

Yao W., Li D., Gao L. Fault detection and diagnosis using tree-based ensemble learning methods and multivariate control charts for centrifugal chillers. Journal of Building

Engineering, 2022, 51:104243.

Zhang Y., Yu J. Pruning graph convolution network-based feature learning for fault diagnosis of industrial processes. Journal of Process Control, 2022, 113:101-113.

Zhao Y., Wang S.W., Xiao F. Pattern recognition-based chillers fault detection method using support vector data description (SVDD). Applied Energy, 2013, 112:1041-1048.

Zhou J., Cui G, Hu S., Zhang Z., Yang C., Liu Z., Wang L., Li C., Sun M. Graph neural networks: A review of methods and applications. AI Open, 2020, 1:57-81.

Zhu X., Chen K., Anduv B., Jin X.Q., Du Z.M. Transfer learning based methodology for migration and application of fault detection and diagnosis between building chillers for improving energy efficiency. Building and Environment, 2021, 200:107957.