Sustainable Cities and Society 70 (2021) 102874

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

A study on semi-supervised learning in enhancing performance of AHU unseen fault detection with limited labeled data

Cheng Fan [a,b], Yichen Liu [a,b], Xuyuan Liu [a,b], Yongjun Sun [c,*], Jiayuan Wang [a,b]

[a] *Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, China*
[b] *Department of Construction Management and Real Estate, Shenzhen University, Shenzhen, China*
[c] *Division of Building Science and Technology, City University of Hong Kong, Hong Kong, China*

A R T I C L E I N F O

Keywords:
Semi-supervised learning
Unseen fault detection
Air handling units
Artificial neural networks
Data science

A B S T R A C T

The fault detection and diagnosis (FDD) of air handling units (AHUs) serves as a major task in building operation management and energy savings. Data-driven classification methods have gained increasing popularities considering their flexibilities and effectiveness in practice. One essential challenge in developing accurate and reliable FDD classification models is the lack of sufficient labeled data. In practice, it can be highly time-consuming, labor-intensive and sometimes even infeasible to collected sufficient labeled data for all possible faulty operations. As a result, the fault detection models developed by limited and partially labeled data may not perform well in detecting any unknown or unseen faults in AHU operations. This study investigates the value of semi-supervised learning in detecting unseen faults during AHU operations. The main idea is to adopt a self-training strategy to gradually enhance the model capability by leveraging large amounts of unlabeled data. Data experiments have been designed to evaluate the unseen fault detection performance, the impacts of key semi-supervised learning parameters and the difficulties in detecting typical AHU faults. The insights obtained are valuable for the integration of data sciences with massive building operational data for smart building management.

1. Introduction

Air handling units (AHUs) are key components in the heating, ventilation and air-conditioning (HVAC) system considering their significant impacts on building energy efficiency and indoor comforts. The AHU fault detection and diagnosis (FDD) plays an essential role in ensuring building sustainability, as 15–30 % energy savings can be achieved by removing equipment, actuator and sensor faults in AHU operations (Gholamzadehmir, Pero, Buffa, Fedrizzi, & Aste, 2020; Sha et al., 2019). For examples, the leakage in air ducts is a typical equipment fault in AHU operations which may cause energy wastes from fans for air circulations. The stuck damper is a typical actuator fault which will lead to deviations between actual and desired system outputs. The sensor faults will lead to inaccurate data measurements and thereby, imposing negative impacts on feedback controls and indoor environment. To ensure the functionality and energy efficiency of AHUs, it is desired to develop accurate and reliable tools for real-time fault detection and diagnosis.

Existing literatures mainly adopted three approaches for AHU fault detection and diagnosis, i.e., the rule-based, physical model-based, and data-driven model-based approaches. The rule-based approach relies on engineering experiences and domain expertise to construct a set of expert rules for decision making (Schein, Bushby, Castro, & House, 2006; Wang, Chen, Chan, & Qin, 2012). Such approach is easy to use, yet the rules defined may be oversimplified, leading to less accurate generalization performance (Bruton et al., 2014; Capozzoli, Piscitelli, Brandi, Grassi, & Chicco, 2018). The physical model-based approach relies on physical principles to develop detailed process models and the differences between model outputs and actual measurements are used as fault indicators (Ren & Cao, 2019; Wang & Chen, 2016). Detailed information on building physics and system configurations is typically required to ensure the physical model validity, making it time-consuming and sometimes even infeasible for practical applications (Fan, Yan et al., 2021). Considering the increasing data availability in modern buildings, data-driven approaches have emerged as convenient and flexible solutions to achieve accurate, automated and real-time controls over various building services systems (Capozzoli, Piscitelli, Gorrino, Ballarini, & Gorrado, 2017; Fan, Xiao, Li, & Wang,

* Corresponding author.
E-mail address: yongjsun@cityu.edu.hk (Y. Sun).

https://doi.org/10.1016/j.scs.2021.102874
Received 18 December 2020; Received in revised form 10 February 2021; Accepted 20 March 2021
Available online 24 March 2021

2018; Yu, Haghighat, & Fung, 2016). One of the most straightforward methods is to develop supervised classification models using labeled data sets. A labeled data set consists of both model input and output variables. The output variable is also known as labels, which can be either binary for fault detection (i.e., whether a data sample is normal or faulty) or multi-class for fault diagnosis (i.e., whether a data sample corresponds to normal or a specific faulty operating condition) (Fan, Du, Jin, Yang, & Guo, 2010; Zhou et al., 2020). In theory, the labeled data should be balanced, i.e., the numbers of data samples corresponding to different labels are approximately the same. Otherwise, specific techniques, such as minority data oversampling, should be applied to ensure the validity of classification model developed.

A number of supervised classification algorithms have been successfully used for HVAC fault detection and diagnosis tasks, such as support vector machines, decision trees and Bayesian classifiers (Theissler, 2017; Xiao, Zhao, Wen, & Wang, 2014; Yan, Ma, Kokogiannakis, & Zhao, 2016). Among various data analytics, artificial neural networks have gained increasing popularity due to the rich variations in model architectures and compatibilities with different learning tasks, such as unsupervised learning (Fan, Xiao, Zhao, & Wang, 2018; Loy-Benitez, Li, Nam, & Yoo, 2020), supervised learning (Lee, Wu, & Peng, 2019; Zou, Yu, & Ergan, 2020) and generative modeling (Chen, Wawrzynski, & Lv, 2021; Kumar & Jayagopal, 2020). Previous studies have shown that given sufficient labeled data, artificial neural networks can achieve high accuracies for detecting and diagnosing building system operation faults (Mirnaghi & Haghighat, 2020). A variety of artificial neural networks have been utilized with the aim of enhancing FDD performance, such as fully connected neural networks (Bode, Thul, Baranski, & Muller, 2020; Yun, Hong, & Seo, 2021), deep belief networks (Shi & O'Brien, 2019; Zhu, Zhang, Jin, & Du, 2020), convolutional neural networks (Eom, Yoo, Hong, & Kim, 2010; Liu, Zhang, Wang, Zhao, & Liu, 2019) and auto-associative neural networks (Elnour, Meskin, & Al-Naemi, 2020).

However, the potential of powerful yet complicated data-driven classification models can be rather limited due to the lack of sufficient labeled data. In practice, it can be highly time-consuming and labor-intensive to collect sufficient labeled data for reliable model development. Consequently, individual buildings may have collected large amounts of building operational data, yet the majority of them are unlabeled and not directly applicable for predictive modeling. Previous studies mainly adopted two approaches to tackling the labeled data shortage problem in the building field. The first adopts different techniques to enrich or modify existing labeled data for model performance enhancement. Some studies assumed that only normal operational data were available for predictive modeling. In such case, one-class support vector models were utilized to describe normal data behaviors, based on which faulty operations were successfully detected (Li et al., 2016; Van Every, Rodriguez, Jones, Mammoli, & Martínez-Ramon, 2017; Yan, Ji, & Shen, 2017). Some studies adopted the concept of data augmentation to enrich the labeled data for reliable model development. Yan et al. relied on generative adversarial networks to overcome the imbalanced data problem for detecting and diagnosing faults in AHU and chiller operations (Yan, Huang, Shen, & Ji, 2020; Yan, Chong, & Mo, 2020). The generative models were developed to produce synthetic data of faulty operations, which were then utilized for classification model development. Fan et al. adopted the synthetic minority oversampling technology (i.e., SMOTE) to facilitate the reliable modeling of chiller faults under imbalanced data scenarios (Fan, Cui, Han, & Lu, 2019). The method was evaluated using the ASHRAE-1043 project data and indicated that an oversampling ratio of 100 % led to the best result.

The second resorts to massive amounts of unlabeled data for model performance enhancement. Semi-supervised learning, which aims to enhance the model performance by utilizing massive amounts of unlabeled data, has been successfully applied in various industries for this purpose (Jiang, Ge, & Song, 2017; Liu & Gryllias, 2020). Fig. 1 illustrates the potential value of semi-supervised learning for an example binary classification task. The two-class labeled data are represented as red crosses and green squares, while unlabeled data are shown as grey circles. Given limited labeled data, the classification model is typically designed with low complexity to avoid the overfitting problem. As a result, the decision boundary learnt (i.e., represented as the black dotted line) may be oversimplified for tackling real-world problems. Semi-supervised learning can modify or update the model decision boundary considering the intrinsic characteristics of unlabeled data. The resulting decision boundary (i.e., shown as the red dotted curve) is generally more reliable and robust for practical applications (Abaei, Selamat, & Fujita, 2015).

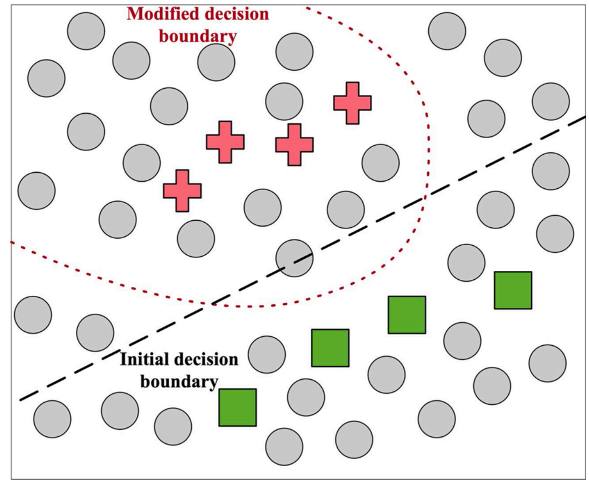

Fig. 1. The potential of semi-supervised learning in classification tasks.

Only a few studies have been performed to investigate the value of semi-supervised learning in building system FDD. Yan et al. explored the value of various semi-supervised learning algorithms for classifying AHU faults, such as semi-supervised support vector machines and random forests (Yan, Zhong, Ji, & Huang, 2018). The research results validated the value of semi-supervised learning in utilizing large amounts of unlabeled building operational data. Fan et al. investigated the value of semi-supervised neural networks in AHU fault diagnosis (Fan, Liu, Xue, & Wang, 2021). Data experiments were designed to quantify the performance of semi-supervised neural networks. Given suitable semi-supervised learning parameters, up to 27 % accuracy enhancement could be achieved in AHU fault diagnosis. Existing studies mainly assumed that all faulty operations were successfully identified and available in the labeled data. In practice, individual buildings may only possess labeled data with a subset of faults. As a result, the classification model developed cannot generalize well towards any unknown or unseen faults.

To tackle such challenge, this study proposes a semi-supervised learning method for unknown or unseen fault detection in AHU operations. Data experiments have been designed to quantitatively evaluate the unseen fault detection performance given different semi-supervised learning strategies. The insights obtained are valuable for the efficient utilization of massive unlabeled building operational data for smart building management. The paper is organized as follows. The research methodology is described in Section 2. Data experiments are described in Section 3. Results and discussions are presented in Section 4. Conclusions are drawn in Section 5.

2. Research methodology

2.1. Outline

In this study, the unknown or unseen faults are defined as faults which do not present in the labeled data, despite their presence in the unlabeled data. It simulates practical scenarios where building operation staffs do not possess sufficient domain expertise to correctly identify all possible fault types in AHU operations. As a result, the data-driven classification model developed using labeled data alone may not perform well in detecting unseen faults. Semi-supervised learning is of particular interests as it can exploit hidden knowledge in unlabeled data for performance enhancement.

As shown in Fig. 2, to evaluate the value of semi-supervised learning in unknown or unseen fault detection, data experiments on AHU operations have been designed with three major considerations. Firstly, data sampling techniques have been used to simulate practical scenarios where the labeling information of a certain fault is missing in the labeled data. Secondly, different labeled data availabilities have been designed to investigate the impacts of labeled data amounts on the reliability of the initial model development and the subsequent semi-supervised learning process. Thirdly, data experiments have been designed to investigate the performance of different semi-supervised learning strategies using different learning parameters. As we shall introduce in the following section, this study adopts the self-training approach to developing semi-supervised neural networks and the impacts of three key learning parameters, i.e., the confidence threshold for pseudo-labeled date selection, the learning rate for pseudo-labeled data utilization and the semi-supervised learning iteration, will be evaluated through the experiment results.

By comparing the unseen fault detection rates between the baseline model (i.e., which is developed based on limited and partly labeled data alone) and the semi-supervised model (i.e., which is developed based on the labeled and unlabeled data together), the experiment results can be used to quantify the general value of semi-supervised learning in AHU fault detection, the influences of different semi-supervised learning parameters on unseen fault detection, and the difficulty in detecting different AHU faults when their labeling information is missing.

2.2. The self-training strategy for semi-supervised neural network development

In general, there are four approaches for semi-supervised classifications, i.e., the generative model-based, low-density separation-based, graph-based and self-labeled approaches (Chapelle, Scholkopf, & Zien, 2006; Triguero, Garcia, & Herrera, 2015). The first two make specific assumptions on unlabeled data distributions and decision boundaries, which limits their potentials in analyzing real-world data. The third approach relies on the graph theories for semi-supervised learning. It is mainly used for analyzing unstructured data (e.g., image or texts) and typically requires more extensive computation resources (Yuan, Li, Wang, & Nie, 2021; Zhu, 2008). The self-labeled approach has gained great popularity due to its simplicity and flexibility for practical implementations. More importantly, the self-labeled approach can be easily adapted using various supervised learning algorithms, making it a promising candidate for analyzing building operational data (Lee, 2013; Ouali, Hudelot, & Tami, 2020).

To ensure the compatibility with building operational data and various supervised learning algorithms, this study adopts the self-training strategy to develop unseen fault detection models.

Fig. 3 illustrates the self-training process for developing semi-supervised neural networks. It is in essence an iterative training process consisting of five major steps. The first is to develop the initial classification model based on labeled data, where the input and output data are denoted as X_l and Y_l respectively. The second is to apply the classification model to unlabeled data (i.e., denoted as X_u) to generate pseudo labels. Given limited labeled data, the classification model may not be able to generate trustworthy and useful pseudo labels for model updates. Therefore, at the third step, a confidence threshold is introduced for pseudo label selection and only those with relatively high confidence scores are selected for further analysis. For instance, given a confidence threshold of 0.7, only predictions with class probabilities no less than 0.7 will be selected for model updates in the next iteration. At the fourth step, another learning parameter, i.e., learning rate, is introduced to specify the training weights of pseudo-labeled data. For instance, given a learning rate of 0.5, a pseudo-labeled data sample with a confidence score of 0.8 will be weighted as 0.4 (i.e., $0.8 \times 0.5 = 0.4$) in the next model training iteration, while the actual labeled data are utilized with training weights of 1.0. In this study, the maximum iteration is fixed as five, i.e., the initial classification model will be updated five times.

2.3. Artificial neural network configurations for AHU unseen fault detection

This study adopts artificial neural networks for predictive modeling. Previous studies have evaluated the usefulness of different operational variables as inputs for AHU fault detection and diagnosis (Fan, Du, Jin, Yang, & Guo, 2010; Roger, Guo, & Rasmussen, 2019). In this study, 13 variables have been selected as inputs considering their importance in AHU fault detection and availability in practice. As shown in Fig. 4, the model is a three-layer fully connected neural network. It is designed with relatively low complexity considering that the amount of labeled data can be very limited for initial model development. More specifically, the input layer considers 13 input variables, including the supplied, returned, outdoor and mixed air temperatures, the supplied, returned and mixed air flow rates, the operating status of supply and return air fans, the differential pressures of supply and return air fans, and the water temperatures at the heating and cooling coils. The hidden layer consists 50 hidden neurons using the *Rectified Linear Unit* as the activation function. To reduce the overfitting risk, a dropout of 0.3 is applied between hidden and output layers for model regularization. The model is trained using an early stopping scheme, indicating that the training process will terminate if the model performance over the

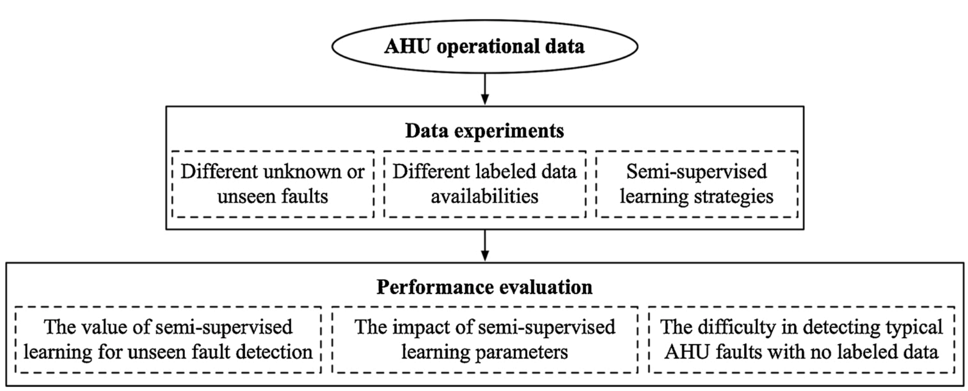

Fig. 2. Outline of semi-supervised neural networks for unseen fault detection.

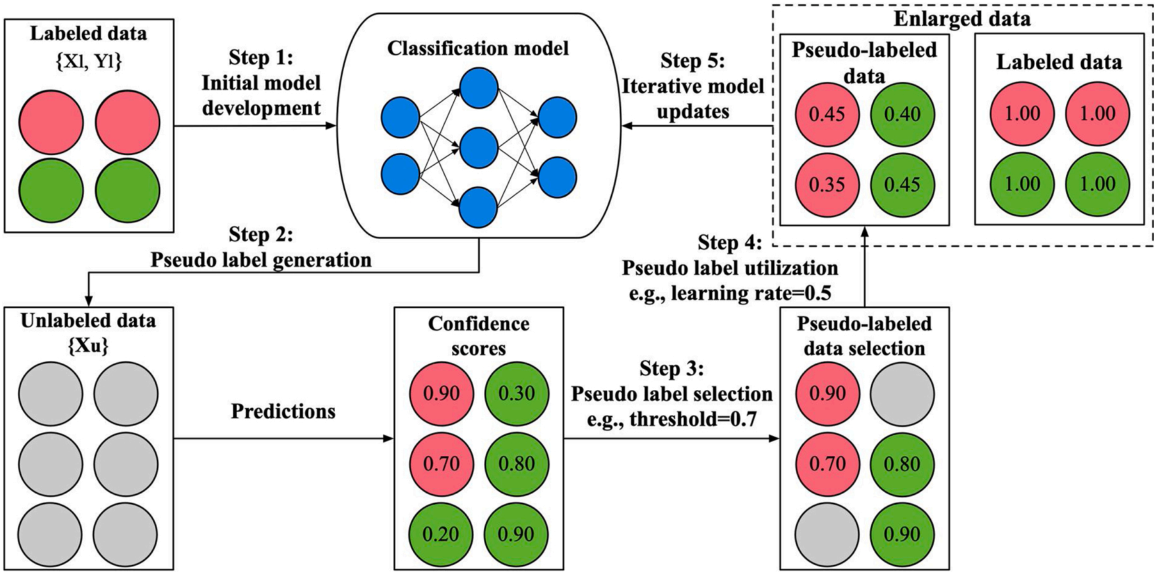

Fig. 3. The self-training strategy for semi-supervised neural network development.

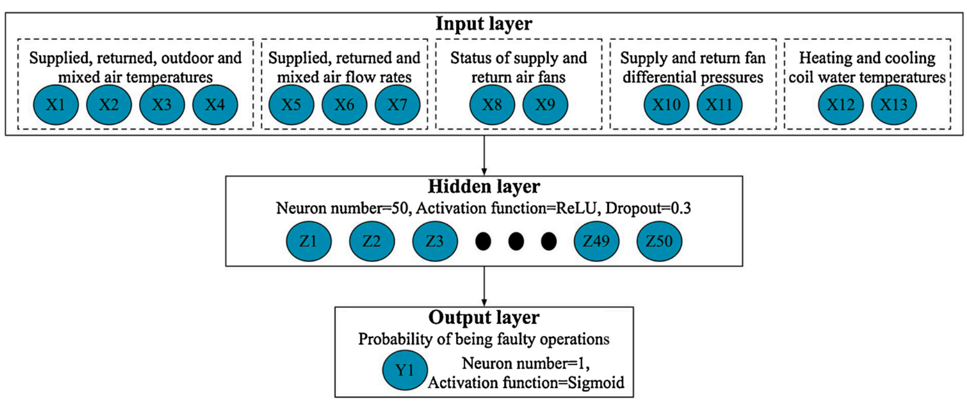

Fig. 4. The artificial neural network configuration for AHU fault detection.

validation data does not improve after certain training iterations. It is worth mentioning that the primary goal of this study is to quantify the relative performance enhancement of semi-supervised learning and therefore, the model architecture optimization is excluded from analysis. The model is utilized to predict the operation status of a data sample (i.e., faulty or normal), which is in essence a binary classification task. Therefore, the model output layer is designed with one neuron using the *Sigmoid* activation function. The model output can be interpreted as the probability of being faulty operations.

3. Data experiments

3.1. Data descriptions

The AHU operational data collected by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) research project RP-1312 have been adopted for data experiments (Wen & Li, 2011). The operational data were collected in three testing periods, i.e., the summer of 2007, the spring and winter of 2008, each lasting for two to three weeks. A variety of AHU operating faults were manually introduced, such as the air damper and water valve faults. The AHU operation variables, ranging from temperature to equipment status, have been collected under different faulty and normal conditions at one-minute collection interval. We refer the interested readers to reference (Wen & Li, 2011; Zhao, Wen, & Wang, 2015; Zhao, Wen, & Wang, 2017) for more detailed descriptions on the ASHRAE RP-1312 data.

In this study, five major AHU fault types, each with three subtypes, have been selected for experiments. As shown in Table 1, the experiment data have 34,560 samples and contain 16 unique labels, i.e., one is

Table 1
A summary of AHU faults selected for analysis.

Fault type	Fault subtype	Data sample numbers
	Fully open	2160
EA damper stuck	40% open	2160
	Fully closed	720
	Fully closed	2160
OA damper stuck	45% open	720
	55% open	720
	Fully open	2160
Cooling coil valve stuck	50% open	1440
	Fully closed	720
	0.4 gallon per minute	720
Heating coil valve leaking	1.0 gallon per minute	720
	2.0 gallon per minute	720
	Total failure	720
Return fan at fixed speed	30% speed	720
	80% speed	720
Normal operations	NA	16560

Normal and the other 15 correspond to different faulty operations. As introduced in Section 2.3, 13 variables have been selected as model inputs considering their importance in fault detection and availability in practice. The statistical characteristics of these input variables are summarized in Table 2.

3.2. Experiment setups

Fig. 5 illustrates the general logics on data experiments. As introduced in Section 3.1, five major AHU faults, each with three subtypes, have been adopted for analysis, resulting in fifteen faulty operations in total, i.e., $P = 15$. The first step is to randomly partition the whole data set into two parts, i.e., 70 % as the existing data set for simulating different semi-supervised learning scenarios, and 30 % as the testing data for performance evaluation.

The second step is to simulate practical scenarios by varying labeled data availabilities and missing labels. In total, fifteen sets of experiments have been conducted, each assuming the lack of one certain fault in the labeled data. For each set of the experiment, eight labeled data sets have been created assuming different labeled data availabilities, i.e., 10, 20, 30, 40, 50, 60, 70 and 80 data samples per fault. Considering that fault detection is in essence a binary classification task trying to distinguish between *Normal* and *Faulty*, a balanced labeled data is then created by sampling the same amount of the faulty data, e.g., given 10 data samples per fault, the number of *Normal* data will be 140 (i.e., $10 \times 14 = 140$) assuming one fault absent from the labeled data. Meanwhile, data selection is conducted on the testing data to generate a balanced evaluation data set for performance evaluation. Assuming *Fault A* is absent from the labeled data, the evaluation data set will consist of both *Normal* and *Fault A* data with a data sample size of $2m$, where m is the data sample number of *Fault A* in the testing data. The sizes of labeled and unlabeled data sets under different labeled data availabilities are summarized in Table 3. It is worth mentioning that random data downsampling should be conducted considering that the number of *Normal* data is much larger than that of a certain fault. To minimize the potential negative impacts of data randomness, each set of experiment has been repeated for ten times and the averaged results are used to quantify the value of semi-supervised neural networks in detecting unknown or previously unseen faults.

As a third step, a label transformation step is conducted for three data sets, i.e., (1) transforming labels to be *Normal* and *Faulty* in the labeled data; (2) removing labels of the remaining existing data as the unlabeled data; (3) transforming labels in the evaluation data set to be *Normal* and *Faulty*. Afterwards, two fault detection models are developed using artificial neural networks. To ensure the comparison fairness, the model configurations are kept the same as described in Section 2.3. The *Baseline Model* is developed based on the labeled data alone, while the *Semi-supervised Model* is developed using both labeled and unlabeled data. As introduced in Section 2.2, different semi-supervised learning parameters have been tested and the detailed settings are shown in Table 4. In total, sixteen semi-supervised learning processes will be conducted considering four confidence thresholds and four learning rates for pseudo-labeled data selection and utilization.

Table 2
Statistical characteristics of input variables.

Variable	Min	1^{st} Q	Mean	3^{rd} Q	Max
Supplied air temperature (°F)	45.3	54.9	58.6	64.0	93.6
Returned air temperature (°F)	65.7	71.0	72.7	73.2	92.0
Outdoor air temperature (°F)	−2.8	0.0	28.5	55.0	80.3
Mixed air temperature (°F)	33.6	55.8	64.4	73.3	90.9
Supplied air flow rates (CFM)	2.25	1255.2	1696.8	2119.7	3372.1
Returned air flow rates (CFM)	0.5	1533.8	1698.5	2006.8	2837.1
Mixed air flow rates (CFM)	−5.0	381.3	824.3	1105.8	3457.7
Supply air fan status (On = 1, Off = 0)	0	1	1	1	1
Return air fan status (On = 1, Off = 0)	0	1	1	1	1
Differential pressure of supply air fan (Pa)	−0.1	2.0	2.4	2.8	4.5
Differential pressure of return air fan (Pa)	−0.3	0.3	0.3	0.4	0.8
Heating coil water temperature (°F)	37.8	60.4	68.3	74.2	121.8
Cooling coil water temperature (°F)	32.0	51.7	55.0	55.0	90.6

3.3. Performance evaluation

This study adopts recall to quantify the value of semi-supervised learning in detecting unknown faults. It is an evaluation metric which can be used to specify the proportions of correctly identified faulty operations over all faulty operations. It ranges from zero to one. In this study, a recall of one indicates that all the unknown or unseen fault operations have been correctly labeled as *Faulty* by the classification model. The *Baseline Model* is expected to have lower recalls, as it is developed based on labeled data with no labeling information of the unknown or unseen fault. By contrast, the *Semi-supervised Model* is expected to have better recalls due to the existence of unknown faults in the unlabeled data. To quantify the value of semi-supervised neural networks in detecting unknown faults, a novel metric called the Performance Improvement in Recalls (i.e., PIR) is defined as $PIR = Recall_{semi} - Recall_{base}$, where $Recall_{semi}$ and $Recall_{base}$ represent recalls of the *Semi-supervised Model* and *Baseline Model* respectively. A positive PIR justifies the value of semi-supervised learning and vice versa. All the research work was conducted using the R programming language (R Development Core Team, 2008) and semi-supervised neural networks were developed using the *Keras* package (Chollet, 2015).

4. Results and discussions

4.1. Overall performance of semi-supervised neural networks in unseen fault detection

Fig. 6 illustrates the averaged recalls of baseline and semi-supervised models given different labeled data amounts together with the resulting performance improvement in recalls (PIRs). The averaged recalls of semi-supervised neural networks are always higher than those of baseline models, indicating that semi-supervised learning is helpful for enhancing the classification performance on unseen fault detection. The maximum PIR is 10.06 % when the number of labeled data is the smallest, i.e., 280.

The averaged recalls of semi-supervised neural networks decrease significantly with the increase in labeled data amounts, while the averaged recalls of baseline models do no present significant changes across different labeled data availabilities. It has two major indications. Firstly, it indicates that the marginal benefit of semi-supervised learning in detecting unseen faults decreases with the increase in labeled data amounts. It is expected as the larger the labeled data set is, the more confined the baseline model will be, leading to smaller marginal benefits in utilizing unlabeled data for unseen fault detection. Secondly, it indicates that the ability of semi-supervised models in detecting unseen faults decreases when more labeled data are available. Fig. 7 serves as an example to explain the impact of labeled data amounts on unseen fault detection. With the increase in labeled data amounts, the decision boundary of the initial classification model will become more confined and tailored to distinguish between *Normal* and *Faulty* operations seen. As there is no labeling information for the unseen fault, the model is more likely to classify the unseen fault in the unlabeled data to be *Normal*, and such pseudo labels will not be helpful to enhance the unseen fault detection rate.

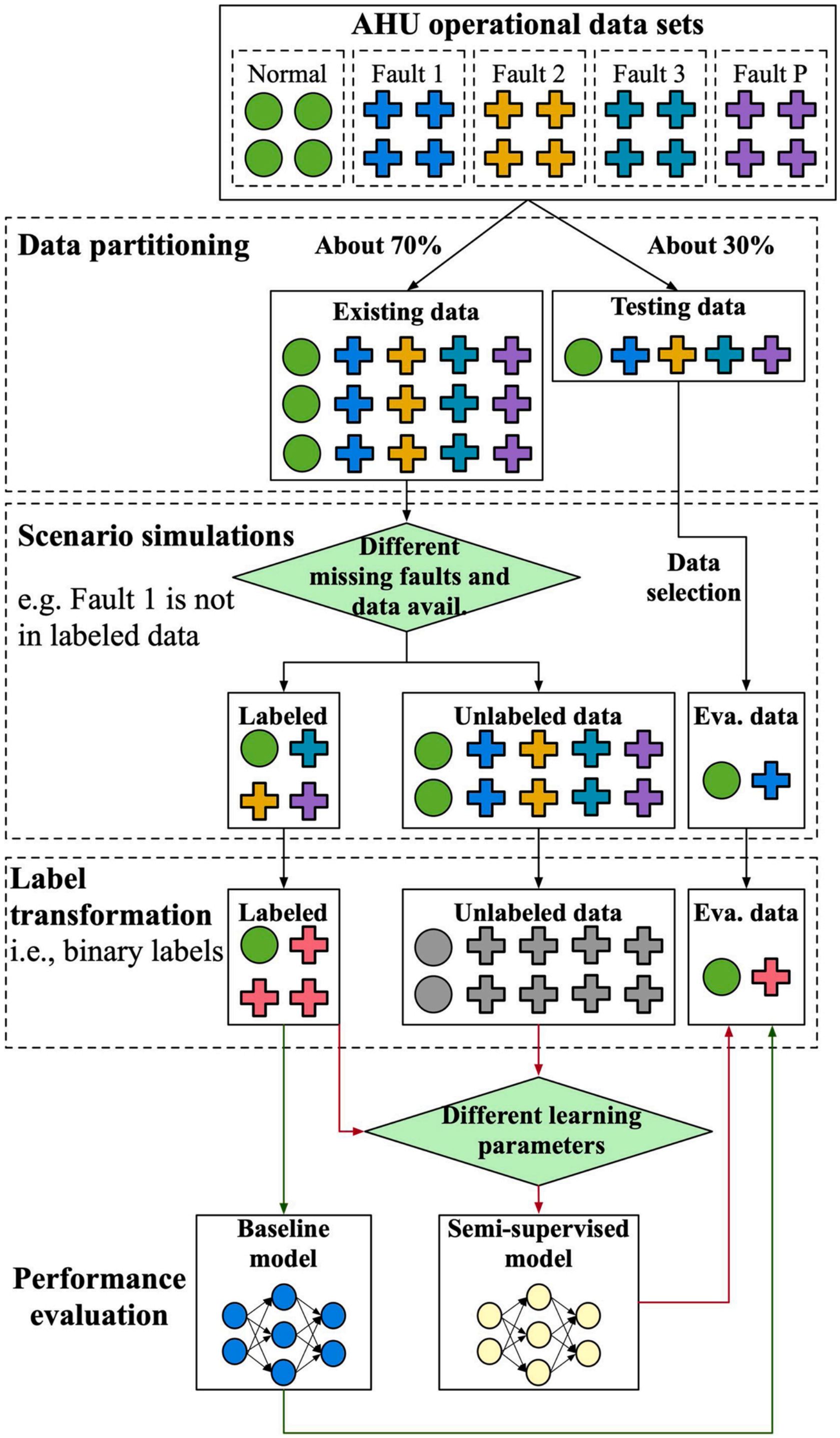

Fig. 5. An illustration on data experiments.

4.2. Impacts of semi-supervised learning parameters

The semi-supervised learning process is controlled by three key learning parameters, i.e., the confidence thresholds for pseudo label selection, the learning rate for pseudo label utilization, and the learning iterations. In practice, it can be very challenging to determine the optimal settings for these three learning parameters due to their contradicting impacts on labeled and unlabeled data utilizations. For instance, a larger confidence threshold will select more trustworthy pseudo-labeled data for model updates, yet at the cost of smaller unlabeled data utilization rates. Fig. 8 presents the averaged unseen fault recalls given different confidence thresholds, i.e., 0.6, 0.7, 0.8 and 0.9. The purple dotted line represents averaged recalls of baseline models. The averaged PIR is summarized in Table 5. In general, higher confidence thresholds should be used for unseen fault detection, as average PIRs are the highest when the confidence threshold is 0.8 or 0.9. It indicates that more stringent criteria should be adopted for pseudo-labeled data selection.

The learning rate also presents contradicting impacts on the semi-supervised learning process. A smaller learning rate refers to a more conservative approach to pseudo-labeled data utilization, which may result in a more stable learning process. Nevertheless, it also limits the

Table 3
A summary on labeled and unlabeled data sizes.

Labeled data availabilities per fault	Labeled data size (*Normal* and *Faulty*)	Unlabeled data size
10	280	23,912
20	560	23,632
30	840	23,352
40	1,120	23,072
50	1,400	22,792
60	1,680	22,512
70	1,960	22,232
80	2,240	21,952

Table 4
Candidate values for semi-supervised learning parameters.

Semi-supervised learning parameters	Candidate values
Confidence thresholds	{0.6, 0.7, 0.8, 0.9}
Learning rate	{0.1, 0.2, 0.3, 0.4}
Maximum iteration	5

influence of unlabeled data to modify or update the model decision boundary. As shown in Fig. 9 and Table 6, the averaged recalls and PIRs become larger when the learning rate is set relatively high, i.e., 0.3 or 0.4. Such finding is not the same with our previous studies, where all the labeling information is present in the labeled data (Fan, Liu et al., 2021). When the aim is to enhance the classification model performance on existing faults in the labeled data, smaller learning rates (e.g., 0.1 or 0.2) should be used. By contrast, given the aim of detecting unknown or unseen faults, the semi-supervised learning process should put more weights on pseudo-labeled data.

As described in Section 3.2, the maximum learning iteration was set as five for data experiments. The averaged recalls and PIRs are summarized in Fig. 10 and Table 7 respectively. It is evident that the larger the semi-supervised learning iteration is, the better the performance for unseen fault detection. Further studies can be conducted to identify the most suitable learning iterations or other stopping criteria for semi-supervised unseen fault detection.

4.3. Evaluation of difficulties in detecting typical yet unseen AHU faults

In this study, five major fault types (i.e., exhaust air damper stuck, outdoor air damper stuck, cooling coil valve stuck, heating coil valve leaking and return air fan with stagnant speed), each with three subtypes, have been utilized for data experiments. The intrinsic differences in fault natures impose different challenges for fault detection when their labeling information is absent in the labeled data. To evaluate the difficulties in detecting such typical yet unseen faults, the averaged recalls and PIRs have been calculated for in-depth analysis.

Fig. 11 presents the averaged recalls and PIRs for the exhaust air damper stuck fault considering different numbers of labeled data. The green and purple lines represent unseen fault recalls of baseline and semi-supervised models respectively. The red bars represent the PIR or the value of semi-supervised learning. The results validate the value of semi-supervised learning in improving the unseen fault detection rate with a maximum PIR of 6.40 %. The maximum recall achieved is 21.89 % when the number of labeled data is 280. The recalls decrease with the increase in labeled data amounts, indicating that the fault nature of exhaust air damper stuck deviate significantly from the other faults and therefore, the increase in labeled data amounts will negatively affect the unseen fault detection performance. The averaged recalls across all

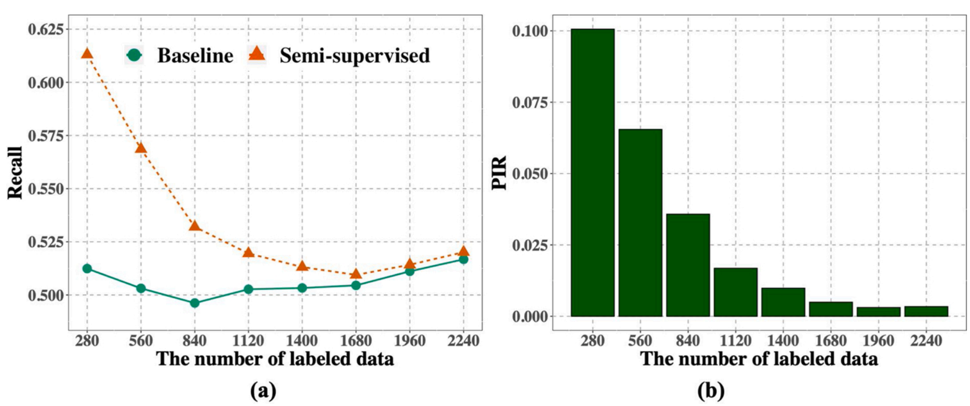

Fig. 6. The general performance of semi-supervised learning for unseen fault detection.

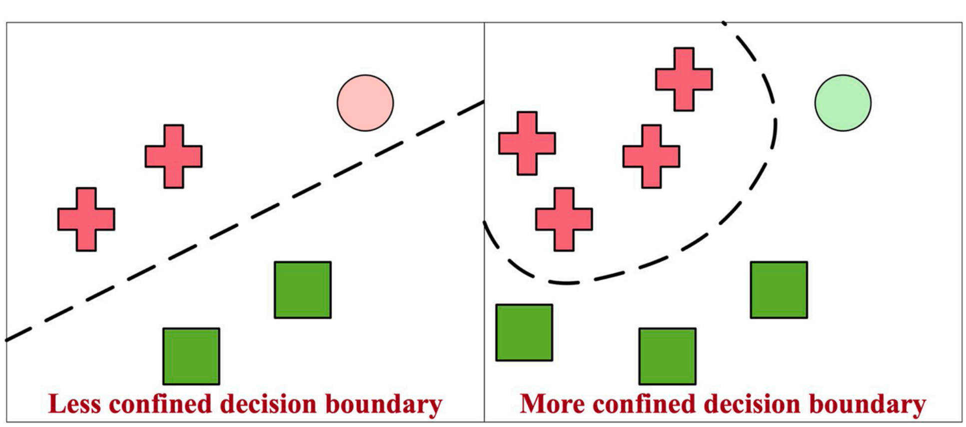

Fig. 7. A graphical illustration on the impact of labeled data amounts on unseen fault detection.

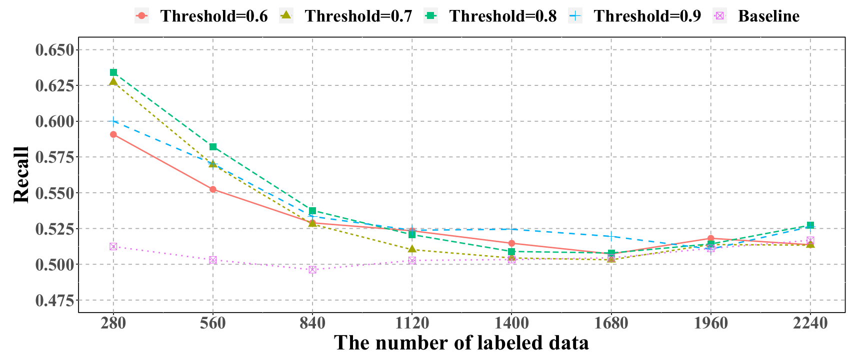

Fig. 8. Averaged unseen fault recalls given different confidence thresholds.

Table 5
Averaged PIRs given different confidence thresholds.

Averaged PIRs Labeled data amounts	Threshold = 0.6 (%)	Threshold = 0.7 (%)	Threshold = 0.8 (%)	Threshold = 0.9 (%)
280	8.04	10.22	12.40	9.57
560	4.41	7.60	7.48	6.70
840	2.40	3.36	3.73	4.82
1120	1.22	1.08	2.00	2.42
1400	0.71	0.60	1.11	1.52
1680	0.07	0.23	0.58	1.10
1960	0.09	0.14	0.11	0.88
2240	−0.18	−0.37	0.78	1.11

labeled data availabilities are 9.77 % and 12.82 % for baseline and semi-supervised models respectively.

As shown in Fig. 12, the averaged recalls for the outdoor air damper stuck fault are much higher, i.e., 26.23 % and 32.49 % for baseline and semi-supervised models respectively. The highest recall achieved is 63.81 % when semi-supervised learning is used given 280 labeled data. Similarly, both recalls and PIRs decrease with the increase in labeled data amounts, indicating a rather different fault nature of the outdoor air damper stuck fault.

Figs. 13 and 14 present the averaged recalls and PIRs for cooling coil and heating coil valve faults respectively. Compared with exhaust and outdoor air damper faults, the recalls are maintained at much higher levels, indicating that such faults are easier for identification. In addition, there is no significant trend between recalls and the labeled data amounts. The averaged recalls for the cooling coil valve fault detection are 70.4 % and 73.64 % for baseline and semi-supervised models respectively. Meanwhile, the averaged recalls are even higher, i.e., 92.42 % and 92.74 %, for the heating coil valve leaking fault detection. One possible explanation is that the heating coil leaking fault data were obtained in winter seasons and their working conditions are much different from the others. Therefore, it is much easier to correctly label such data as faulty operations. Fig. 15 presents the recalls and PIRs for the return air fan with the stagnant speed fault. The averaged recalls are 54.29 % and 56.43 % for baseline and semi-supervised models respectively. A slightly increasing trend in recalls is observed for the return air fan with stagnant speed. It indicates that the fault nature of return air fan shares similarities with other faults and therefore, the increase in labeled data amounts helps to enhance the fault detection rate.

To summarize, based on the averaged recalls across all labeled data availabilities, it is possible to evaluate the difficulties in detecting typical AHU faults when their labeling information is missing. The heating coil valve leaking fault is the easiest to identify due to the great variations between heating and cooling working conditions. By contrast, air damper faults are much harder to identify without their labeling information.

5. Conclusions

The accurate and reliable fault detection and diagnosis of AHUs plays an essential role in building sustainability considering its significant

Table 6
Averaged PIRs given different learning rates.

Averaged PIRs Labeled data amounts	Learning rate = 0.1 (%)	Learning rate = 0.2 (%)	Learning rate = 0.3 (%)	Learning rate = 0.4 (%)
280	7.67	9.41	11.56	11.60
560	4.24	6.20	7.89	7.85
840	2.03	3.48	3.90	4.90
1120	1.35	1.01	1.95	2.41
1400	0.94	0.66	1.01	1.32
1680	0.26	0.28	0.63	0.80
1960	0.54	0.04	0.33	0.31
2240	0.70	0.34	0.05	0.26

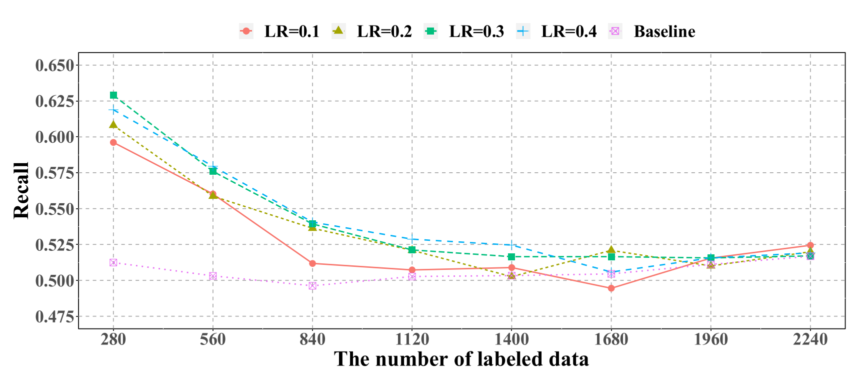

Fig. 9. Averaged unseen fault recalls given different learning rates.

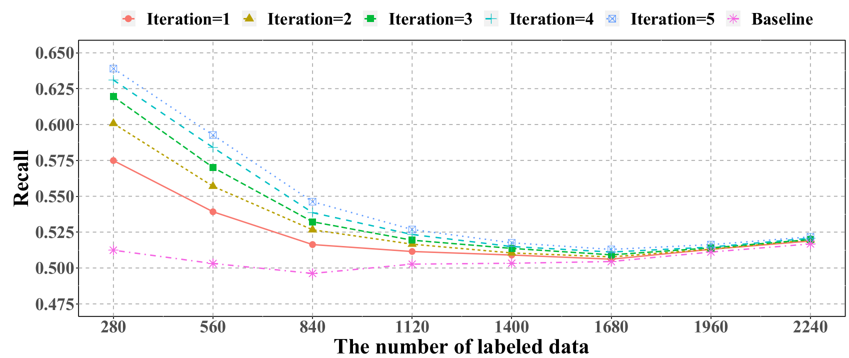

Fig. 10. Averaged unseen fault recalls given different learning iterations.

Table 7
Averaged PIRs given different learning iterations.

Averaged PIRs Labeled data amounts	Iteration = 1 (%)	Iteration = 2 (%)	Iteration = 3 (%)	Iteration = 4 (%)	Iteration = 5 (%)
280	6.25	8.83	10.70	11.86	12.66
560	3.61	5.38	6.70	8.09	8.96
840	2.01	3.04	3.60	4.23	5.00
1120	0.88	1.39	1.67	2.06	2.41
1400	0.57	0.72	1.04	1.16	1.42
1680	0.17	0.32	0.47	0.66	0.85
1960	0.19	0.20	0.29	0.35	0.52
2240	0.20	0.28	0.30	0.40	0.49

impacts on building energy efficiency and indoor environment controls. Data-driven classification methods are effective and promising solutions for automated fault detection and diagnosis. To ensure the reliability of data-driven models, it is essential to prepare sufficient labeled data for model training and validation. In practice, individual buildings may have collected large amounts of building operational data, yet the majority of them are unlabeled and not directly applicable for predictive modeling. To tackle such challenges, this study proposes a novel semi-supervised learning method to fully realize the value in both labeled and unlabeled data. Data experiments have been conducted to quantitatively assess the value of semi-supervised neural networks in detecting unseen faults during AHU operations. The major research findings are summarized as follows:

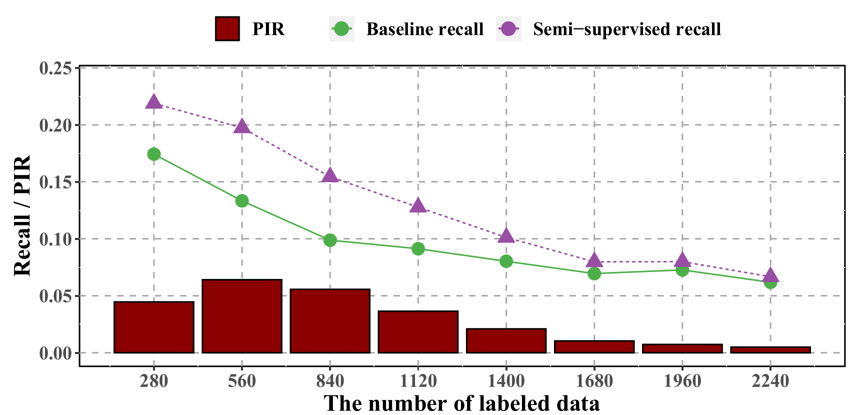

Fig. 11. The recalls and PIRs for the exhaust air damper stuck fault.

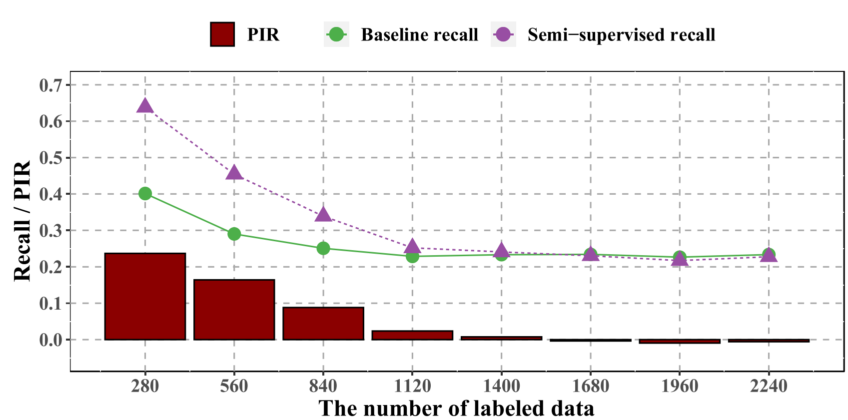

Fig. 12. The recalls and PIRs for the outdoor air damper stuck fault.

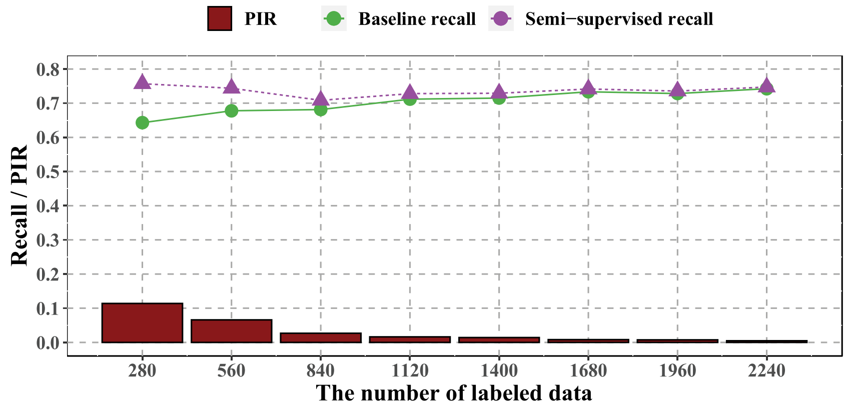

Fig. 13. The recalls and PIRs for the cooling coil valve stuck fault.

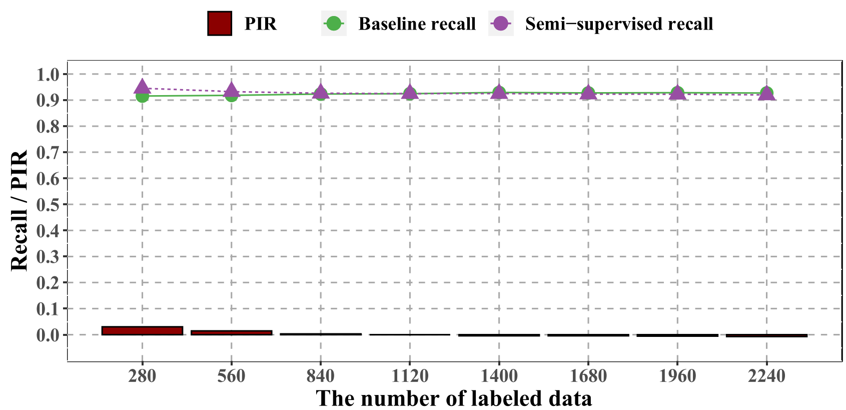

Fig. 14. The recalls and PIRs for the heating coil valve leaking fault.

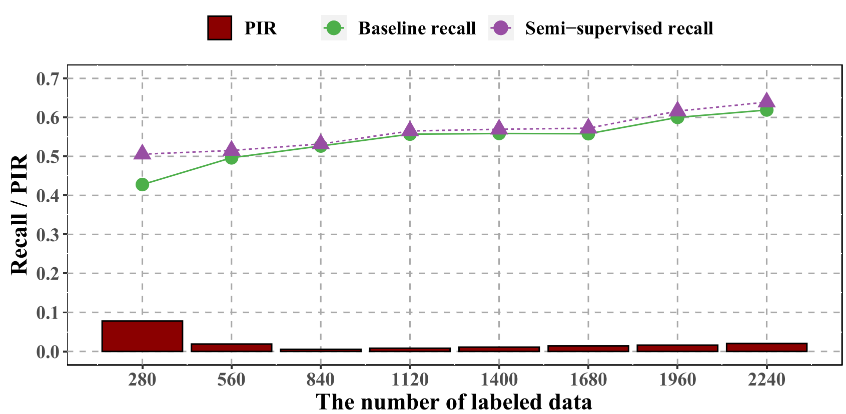

Fig. 15. The recalls and PIRs for the return air fan with the stagnant speed fault.

(1) The semi-supervised learning method is effective in enhancing the unseen fault detection rate in AHU operations. The research results show that the averaged performance improvement in recall (PIR) is maximal (i.e., 10.06 %) given the smallest labeled data set (i.e., 280 samples). The value of semi-supervised learning generally decreases with the increase in labeled data amounts, which is expected as the initial classification models will become more confined or tailored to distinguish between normal operations and known faults. Nevertheless, the research results still justify the value of unlabeled data as semi-supervised models generally have higher recalls of unseen faults.

(2) The impacts of key learning parameters on semi-supervised neural networks have been quantified through data experiments. To enhance the performance in detecting unknown or unseen faults, larger confidence thresholds (e.g., 0.8 or 0.9) and learning rates (e.g., 0.3 or 0.4) are recommended for pseudo-labeled data selection and utilization.

(3) The difficulties in detecting typical AHU faults have been quantitatively assessed when their labeling information is absent from the labeled data. The heating and cooling coil valve faults are relatively easier for fault identification. The averaged recalls achieved by semi-supervised neural networks are 92.74 % and 73.64 % respectively. By contrast, the other three typical faults, i.e., the returned air fan with stagnant speed, the exhaust and outdoor air damper stuck fault, are more difficult to identify without their labeling information. With the aid of semi-

supervised learning, the averaged recalls achieved are 56.44 %, 32.49 %, and 12.82 % respectively.

This study systematically explores the value of semi-supervised neural networks in detecting various AHU faults when their labeling information is absent from labeled data. The research results are valuable for the efficient utilization of both labeled and unlabeled building operational data, and the development of advanced data analytics for smart and sustainable building management. Further studies can be conducted from two aspects. The first is to investigate the performance of semi-supervised neural networks for detecting and diagnosing co-occurrence faults. The second is to assess the potential of other semi-supervised learning methods for unseen fault detection, e.g., semi-supervised support vector machines and random forests.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

The authors gratefully acknowledge the support of this research by the National Natural Science Foundation of China (No. 51908365 and 71772125) and the Philosophical and Social Science Program of Guangdong Province, China (GD18YGL07).

References

Abaei, G., Selamat, A., & Fujita, H. (2015). An empirical study based on semi-supervised hybrid self-organizing map for software fault prediction. *Knowledge-Based Systems, 74*, 28–39.

Bode, G., Thul, S., Baranski, M., & Muller, D. (2020). Real-world application of machine-learning-based fault detection trained with experimental data. *Energy, 198*, Article 117323.

Bruton, K., Raftery, P., O'Donovan, P., Aughney, N., Keane, M. M., & O'Sullivan, D. T. J. (2014). Development and alpha testing of a cloud based automated fault detection and diagnosis tool for air handling units. *Automation in Construction, 39*, 70–83.

Capozzoli, A., Piscitelli, M. S., Gorrino, A., Ballarini, I., & Gorrado, V. (2017). Data analytics for occupancy pattern learning to reduce the energy consumption of HVAC systems in office buildings. *Sustainable Cities and Society, 35*, 191–208.

Capozzoli, A., Piscitelli, M. S., Brandi, S., Grassi, D., & Chicco, G. (2018). Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings. *Energy, 157*, 336–352.

Chapelle, O., Scholkopf, B., & Zien, A. (2006). *Semi-supervised learning*. Cambridge, London, England: The MIT Press.

Chen, D. L., Wawrzynski, P., & Lv, Z. H. (2021). Cyber security in smart cities: A review of deep learning-based applications and case studies. *Sustainable Cities and Society, 66*, Article 102655.

Chollet, F. (2015). *Keras, Github*. https://github.com/fchollet/keras.

Elnour, M., Meskin, N., & Al-Naemi, M. (2020). Sensor data validation and fault diagnosis using auto-associative neural network for HVAC systeems. *J. Build. Eng., 27*, Article 100935.

Eom, Y. H., Yoo, J. W., Hong, S. B., & Kim, M. S. (2010). Refrigerant charge fault detection method of air source heat pump system using convolutional neural network for energy saving. *Energy, 187*, Article 115877.

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

Fan, B., Du, Z. M., Jin, X. Q., Yang, X. B., & Guo, Y. B. (2010a). A hybrid FDD strategy for local system of AHU based on artificial neural network and wavelet analysis. *Building and Environment, 45*, 2698–2708.

Fan, B., Du, Z. M., Jin, X. Q., Yang, X. B., & Guo, Y. B. (2010b). A hybrid FDD strategy for local system of AHU based on artificial neural network and wavelet analysis. *Building and Environment, 45*, 2698–2708.

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

Fan, C., Xiao, F., Li, Z., & Wang, J. (2018). Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review. *Energy and Buildings, 159*, 296–308.

Fan, C., Xiao, F., Zhao, Y., & Wang, J. Y. (2018). Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. *Applied Energy, 211*, 1123–1135.

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

Gholamzadehmir, M., Pero, C. D., Buffa, S., Fedrizzi, R., & Aste, N. (2020). Adaptive-preditive control strategy for HVAV systems in smart buildings-A review. *Sustainable Cities and Society, 63*, Article 102480.

Jiang, L., Ge, Z., & Song, Z. H. (2017). Semi-supervised fault classification based on dynamic sparse stacked auto-encoders model. *Chemometrics and Intelligent Laboratory, 168*, 72–83.

Kumar, M. R. P., & Jayagopal, P. (2020). Generative adversarial networks: A survey on applications and challenges. *The International Journal of Multimedia Information Retrieval, 139*.

Lee, D. H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *ICML 2013 Workshop: Challenges in Representation Learning*.

Lee, K. P., Wu, B. H., & Peng, S. L. (2019). Deep-learning-based fault detection and diagnosis of air-handling units. *Building and Environment, 157*, 24–33.

Li, G. N., Hu, Y. P., Chen, H. X., Shen, L. M., Li, H. R., Hu, M., et al. (2016). An improved fault detection method for incipient centrifugal chiller faults using the PCA-R-SVDD algorithm. *Energy and Buildings, 116*, 104–113.

Liu, C. Y., & Gryllias, K. (2020). A semi-supervised support vector data description-based fault detection method for rolling element bearings based on cyclic spectral analysis. *Mechanical Systems and Signal Processing, 140*, Article 106682.

Liu, J. J., Zhang, M., Wang, H., Zhao, W., & Liu, Y. (2019). Sensor fault detection and diagnosis method for AHU using 1-D CNN and clustering analysis. *Computational Intelligence and Neuroscience, 2019*, 1–20.

Loy-Benitez, J., Li, Q., Nam, K., & Yoo, C. K. (2020). Sustainable subway indoor air quality monitoring and fault-tolerant ventilation control using a sparse autoencoder-driven sensor self-validation. *Sustainable Cities and Society, 52*, Article 101847.

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

Ouali, Y., Hudelot, C., & Tami, M. (2020). An overview of deep semi-supervised learning. *arXiv: 2006.05278v2*.

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

Ren, C., & Cao, S. J. (2019). Development and application of linear ventilation and temperature models for indoor environmental prediction and HVAC systems control. *Sustainable Cities and Society, 51*, Article 101673.

Roger, A. P., Guo, F., & Rasmussen, B. P. (2019). A review of fault detection and diagnosis methods for residential air conditioning system. *Building and Environment, 161*, Article 106236.

Schein, J., Bushby, S. T., Castro, N. S., & House, J. M. (2006). A rule-based fault detection method for air handling units. *Energy and Buildings, 38*, 1485–1492.

Sha, H. J., Xu, P., Hu, C. H., Li, Z. L., Chen, Y. B., & Chen, Z. (2019). A simplified HVAC energy prediction method based on degree-day. *Sustainable Cities and Society, 51*, Article 101698.

Shi, Z. X., & O'Brien, W. (2019). Development and implementation of automated fault detection and diagnostics for building systems: A review. *Automation in Construction, 104*, 215–229.

Theissler, A. (2017). Detecting known and unknown faults in automotive systems using ensemble-based anomaly detection. *Knowledge-Based Systems, 123*, 163–173.

Triguero, I., Garcia, S., & Herrera, F. (2015). Self-labeled techniques for semi-supervised learning: Taxonomy, software and empirical study. *Knowledge and Information Systems, 42*, 245–283.

Van Every, P. M., Rodriguez, M., Jones, C. B., Mammoli, A. A., & Martínez-Ramon, M. (2017). Advanced detection of HVAC faults using unsupervised SVM novelty detection and Gaussian process models. *Energy and Buildings, 149*, 216–224.

Wang, H. T., & Chen, Y. M. (2016). A robust fault detection and diagnosis strategy for multiple faults of VAV air handling units. *Energy and Buildings, 127*, 442–451.

Wang, H. T., Chen, Y. M., Chan, C. W. H., & Qin, J. Y. (2012). An online fault diagnosis tool of VAV terminals for building management and control systems. *Automation in Construction, 22*, 203–211.

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

Xiao, F., Zhao, Y., Wen, J., & Wang, S. W. (2014). Bayesian network based FDD strategy for variable air volume terminals. *Automation in Construction, 41*, 106–118.

Yan, R., Ma, Z. J., Kokogiannakis, G., & Zhao, Y. (2016). A sensor fault detection strategy for air handling units using cluster analysis. *Automation in Construction, 70*, 77–88.

Yan, K., Ji, Z. W., & Shen, W. (2017). Online fault detection methods for chillers combining extended Kalman filter and recursive one-class SVM. *Neurocomputing, 228*, 205–212.

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

Yan, K., Chong, A., & Mo, Y. C. (2020). Generative adversarial network for fault detection diagnosis of chillers. *Building and Environment, 172*, Article 106698.

Yan, K., Huang, J., Shen, W., & Ji, Z. W. (2020). Unsupervised learning for fault detection and diagnosis of air handling units. *Energy and Buildings, 210*, Article 109689.

Yu, Z., Haghighat, F., & Fung, B. C. M. (2016). Advances and challenges in building engineering and data mining applications for energy-efficient communities. *Sustainable Cities and Society, 25*, 33–38.

Yuan, Y., Li, X., Wang, Q., & Nie, F. P. (2021). A semi-supervised learning algorithm via adaptive Laplacian graph. *Neurocomputing, 426*, 162–173.

Yun, W. S., Hong, W. H., & Seo, H. (2021). A data-driven fault detection and diagnosis scheme for air handling units in building HVAC systems considering undefined states. *The Journal of Building Engineering, 35*, Article 102111.

Zhao, Y., Wen, J., & Wang, S. W. (2015). Diagnostic Bayesian networks for diagnosing air handling units faults – Part II: Faults in coils and sensors. *Applied Thermal Engineering, 90*, 145–157.

Zhao, Y., Wen, J., & Wang, S. W. (2017). Diagnostic Bayesian networks for diagnosing air handling units faults – Part I: Faults in dampers, fans, filters and sensors. *Applied Thermal Engineering, 111*, 1272–1286.

Zhou, Z. X., Li, G. N., Wang, J. Y., Chen, H. X., Zhong, H. L., & Cao, Z. H. (2020). A comparison study of basic data-driven fault diagnosis methods for variable refrigerant flow system. *Energy and Buildings, 224*, Article 110232.

Zhu, X. J. (2008). *Semi-supervised learning literature survey*. USA: University of Wisconsin-Madison.

Zhu, X., Zhang, S., Jin, X. Q., & Du, Z. M. (2020). Deep learning based reference model for operaional isk evaluation of screw chillers for energy efficeincy. *Energy, 213*, Article 118833.

Zou, Z., Yu, X., & Ergan, S. (2020). Towards optimal control of air handling units using deep reinforcement learning and recurrent neural network. *Building and Environment, 168*, Article 106535.