Identifying Unseen Faults for Smart Buildings by Incorporating Expert Knowledge With Data

Dan Li, Yuxun Zhou, Guoqiang Hu, *Member, IEEE*, and Costas J. Spanos, *Fellow, IEEE*

***Abstract*— Thanks to the development of sensor networks and information technology, data-driven fault detection and diagnosis (FDD) is getting more and more popular with rich data. In the building FDD field, mature supervised learning algorithms and strategies have been applied to detect and diagnose known faults. However, it is out of the question to collect labeled training data for every possible fault. Thus, there is a necessity to study FDD when the training data for some faults are unavailable. To the authors' best knowledge, few works have reported how to identify "unseen faults." In this paper, authors propose a novel expert knowledge-based unseen fault identification (EK-UFI) method to identify unseen faults by employing the similarities between known faults and unknown faults. The similarity is captured by incorporating essential expert knowledge that is encoded in the fault gene matrix. The fault gene is integrated with a latent incorporation matrix that transfers knowledge from known faults to unseen faults. With application to a real system, the proposed method is proven to be effective in identifying various building unknown faults with a high accuracy.**

***Note to Practitioners*— FDD is of great importance for saving energy and improving occupancy comfort levels and building safety levels. Identifying unseen faults in real application is challenging since: 1) building faults are complicated and confusing while well-labeled fault data is rare; 2) experimental fault data collected in laboratory test beds cannot be directly used as judgment criteria for real buildings; and 3) it is impossible to measure every possible fault ahead of time. Although supervised learning methods have been successfully applied in existing works to solve the building FDD, they could not attack the UFI problem. In this paper, a novel EK-UFI method is proposed to identify unseen faults by employing the similarities between known faults and unknown faults. Experimental results show that the proposed method is essential.**

***Index Terms*— Air handling unit (AHU), heating, ventilation, and air conditioning (HVAC), smart building, unseen fault identification (UFI).**

Manuscript received August 15, 2018; accepted September 27, 2018. This paper was recommended for publication by Associate Editor Z. Xiong and Editor F.-T. Cheng upon evaluation of the reviewers' comments. This work was supported in part by the National Research Foundation, Prime Minister's Office, Singapore, through the Energy Innovation Research Programme for Building Energy Efficiency Grant Call administered by the Building and Construction Authority under Grant NRF2013EWT-EIRP004-051 and in part by the Republic of Singapore's National Research Foundation under its Campus for Research Excellence and Technological Enterprise programme through a grant to the Berkeley Education Alliance for Research in Singapore for the Singapore—Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. *(Corresponding author: Guoqiang Hu.)*

D. Li and G. Hu are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: dli006@e.ntu.edu.sg; gqhu@ntu.edu.sg).

Y. Zhou and C. J. Spanos are with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: yxzhou@berkeley.edu; spanos@berkeley.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2018.2876611

I. Introduction

BUILDING energy consumption contributes to more than 40% of the total energy usage worldwide, and most of that is due to heating, ventilation, and air conditioning (HVAC) systems [1], [2]. A large part of this energy is wasted because of poor maintenance, degradation, and improperly controlled equipment. Therefore, in terms of energy savings, automated building fault detection and diagnosis (FDD) is attracting more and more attention [3], [4].

Despite the miscellaneous FDD methods, faults in the real application are complicated and confusing [5], [6]. For example, faulty valves and dampers in air handling unit (AHU) are difficult to identify since the increased cooling coil entering air temperature caused by a leaky hot water valve would be compensated for by the increased mechanical cooling. An incorrect mixed air temperature due to excessive or inadequate outside air (damper fault) is usually compensated for by the increased cooling or heating load. Under those situations, faults could be unnoticed for years since the indoor air quality (IAQ) is generally not affected. To identify them, FDD strategies should be designed with capabilities to monitor the inner connections among different components. Compared with the traditional model-based methods [7]–[9], data-driven FDD method is becoming more and more popular due to its superiority in revealing the underlying patterns and relationships [10]–[12]. As a result, a wide range of pattern classification techniques have been explored as data-driven methods in the building FDD field, including multivariate regression models [13], Bayes classifiers [14], neural networks (NNs) [15], linear discriminant analysis (LDA) [16], Gaussian mixture models [17], support vector data description [18], [19], support vector machines [20], [21], and tree-structured learning method [22], [23].

It is shown that supervised data-driven FDD is quite popular and effective in terms of both diagnosis accuracy and efficiency. However, these approaches cannot tackle the challenge that new fault classes (unseen faults) may appear after the training stage. Generally speaking, supervised FDD fades when the real application considers the following factors.

1) The building management system (BMS) merely stores normal data (or unlabelled fault data) while well-labeled fault data are rare.
2) Experiments are possible in laboratory test-beds, however, available data sets are limited in terms of the sample size and a number of fault types.
3) As common sense, there does not exist two identical buildings in the world. The well-labeled fault data

obtained in one building cannot be directly used as judgment criteria for another one.

4) It is impossible to measure every possible fault ahead of time.

Taking those factors into consideration, in this paper, authors propose to identify unseen building faults whose labeled data are not available in the training process. Identifying unseen faults rises the problem about how to recognize new concepts on the basis of known categories [24]–[26]. With analogy to how human beings identify a new object, a completely new object could be recognised by a human via just reading a high-level description of it [27]. Similarities between the description of the new object and previously learned concepts are leveraged as the hidden knowledge that can help to deepen human beings' understanding of the new object. To achieve this, the following challenging points should be addressed. First, a new learning framework should be proposed since traditional supervised classification is not applicable due to lacking of labeled data, while pure unsupervised methods cannot identify fault categories. Second, high-level description about building faults needs to be explored to relate unseen faults with fault classes that are previously learned in the training process. Third, it requires careful mathematical design to transfer the extracted high-level knowledge and apply it to identify unseen categories.

In the literature, researchers have paid much attention to deal with the aforementioned situations (limitation of labeled data). Traditionally, when only normal data are available, unsupervised methods such as kernel principal component analysis [28] and Gaussian mixture model [29], [30] are the popular solutions for anomalous event detection while the causes and anomaly types are not well revealed. Keigo *et al.* [31] proposed a semisupervised algorithm to identify building energy faults with limited labeled data. Similar semisupervised solutions can be found in [32]–[34], where noise accumulation of the unlabelled data may hamper the identification performance. An unknown input observer was utilized in [35] to decouple unknown inputs from residuals which aims at improving the robustness of isolating known faults. Li *et al.* [16] proposed to use the fault range confirmation mechanism in their FDD strategy to make it aware of unknown faults. The aforementioned existing methods could detect new anomalous conditions with knowledge about the normal condition and known faults, however, they cannot tell corresponding fault types.

With the recent development in machine learning field, besides direct analysis of historical data, strategies that make use of the expert knowledge which can be transferred from one category/system/model to another are attracting more and more attention [27], [36]. Li *et al.* [22] reported that FDD performance can be greatly improved by incorporating system structural information, which is also called expert knowledge, into the classifier. In [37], hybrid minimal structurally overdetermined (HMSO) sets were designed to capture the system's fault feature and solve the fault diagnosis problem with HMSO selection. Those successful applications reveal that the system specific expert knowledge which conveys system structural information is of great importance for FDD research. In addition, to overcome the shortcoming that sensor measurements cannot be directly passed from building to building, universal expert rules are summarized by researchers and building experts on the basis of field tests in different buildings [38], [39]. Although rule-based FDD is intuitive and widely used in nowadays' HVAC systems, the high false alarm rate is an obvious defection. Moreover, the basic principal component analysis (PCA) has been widely utilized to extract the principal features, which could be leveraged to conduct fault detection, among multiple sensors [40], [41]. To combine benefits and avoid disadvantages, all the aforementioned knowledge[1] is considered as the expert knowledge in this paper.

Accordingly, a novel expert knowledge-based unseen fault identification (EK-UFI) method is proposed to identify unseen faults by levering information learned from known faults and expert knowledge and unseen faults. To be specific, the proposed EK-UFI method includes two parts: training and inference. In the training part, expert knowledge about known faults is first described in terms of a numerical matrix. Here, the matrix of expert knowledge is encoded as the fault gene matrix since those prior knowledge is always the same no matter where and when it occurs, just seems like the Gene of a fault. Then, a latent incorporation matrix describing the relationship between the known faults and their fault gene is learned. In the inference part, a similar relationship is assumed to exist between the unseen fault and its Gene, and it can be transferred from the known categories to unknown ones. Thus, the knowledge conveyed by the fault gene matrix is incorporated into the classifier to identify unseen fault classes. In contrast to the existing works, the contributions of this paper are summarized as follows: 1) a novel EK-FDD method is proposed to identify unseen faults for smart buildings; 2) the proposed method extends the building FDD research into a boarder field where well-labeled data are limited; and 3) the proposed method combines the advantages of rule-based methods and data-driven methods, which makes the strategy intuitive with mathematical guarantee.

The remaining part of this paper is arranged as follows. Section II introduces expert knowledge and formulates the unseen fault detection problem. Section III shows the mathematical formulation of the EK-UFI method. Next, UFI results by EK-UFI are presented with tests on real-time data, and the comparison is discussed in Section IV. Finally, Section V summarizes this paper and suggests possible future work.

II. Expert Knowledge and Unseen Fault Identification Framework

A. Unseen Fault Identification Framework

Consider the building FDD problem: given a set of sensor measurements, annotate it with one label to describe the building working condition (normal or faulty). As shown in Fig. 1 (top), the FDD problem is a prediction task whose

[1]Namely, expert rules, system-specific knowledge and data-driven principal information which are also described with details in Section II-B.

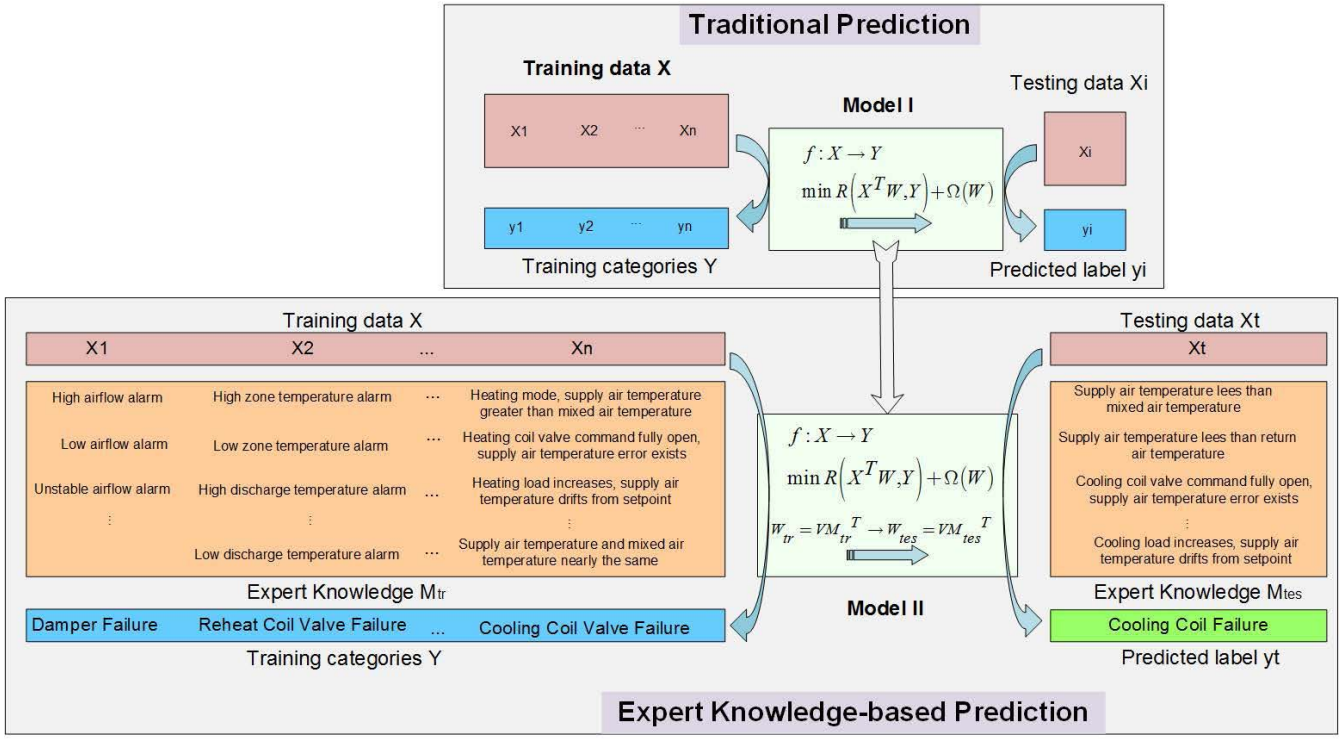


Fig. 1. Summary of the proposed EK-UFI framework. At the training stage, expert knowledge M_{tr} of known faults is incorporated with the training data to learn the transformation matrix V. At the prediction stage, expert knowledge M_{tes} of unseen fault is utilized to identify the testing fault.

goal is to learn a function $f : \mathcal{X} \longrightarrow \mathcal{Y}$ from the labeled training data $\mathcal{S} = \{(x_i, y_i), i = 1, \ldots, n\}$. With traditional prediction methods, the function maps an input x_i in the space of measurements $\mathcal{X}$ to an output in the space of training class labels $\mathcal{Y}$, namely, the predicted label y_i is included in the training label set $\mathcal{D} = \{y_i, i = 1, \ldots n\}$. However, identifying unseen faults means that we have no labeled samples for the testing fault classes but still wish to predict their labels ($y_j \nsubseteq \mathcal{D}$).

In this paper, authors propose a way to incorporate extracted expert knowledge (or high-level description about faults) with feature embedding, which consequently transfers knowledge from known faults to unseen ones. Summarized from the literature in Section I, expert knowledge includes but not limited to expert fuels, system-specific knowledge, and data-driven pricipal information. In Fig. 1 (bottom), as an example, expert knowledge for known faults and unseen faults are encoded by M_{tr} and M_{tes}, respectively. Note that expert rules in this figure are rules taken from the AHU performance assessment rule (APAR) diagnosis and variable air volume (VAV) box performance assessment control chart (VPACC) diagnosis. The APARs and the VPACCs are summarized by the National Institute of Standards and Technology of America [38]. For example, high/low/unstable airflow indicates a damper failure, and high/low zone temperature and high/low discharge temperature indicate a reheat coil valve failure, and so on. By incorporating M_{tr} as a hidden layer in the training process, a latent incorporation matrix can be learned. It is assumed that the latent incorporation matrix which relates known training faults with M_{tr} can be transferred to relate unseen testing faults with M_{tes} [24]–[26]. Thus, with the knowledge transferred by the expert knowledge M_{tes} of the unseen fault, a predicted label vector y'_t is allocated to the testing set x'_t. The proposed method is derived with details in Section III.

B. Expert Knowledge

In this section, the three kinds of expert knowledge considered in this paper are introduced.

1) ASHRAE Rule-Based Expert Knowledge: Rule-based detection and diagnosis is popular due to its simplicity and intuition [38], [42], [43]. Usually, rules are combinations of threshold values characterized by some representative variables. A fault will be detected and diagnosed if those threshold values are surpassed. Thus, FDD rules summarized by building experts are the first choice of the proposed expert knowledge-based UFI.

However, the rule-based method has obvious shortcomings. On one hand, it cannot distinguish similar system performances. According to [39], rules are summarized by ASHRAE experts to identify typical AHU faults under different seasonal cases. The faulty system performance for a typical AHU is described by 16 prime variables[2] and quantified by comparing the measurements under faulty and normal conditions. It can be seen in [39] that some faults are hardly evident merely by ASHRAE rules since they have almost the same performance

[2]Note that ASHRAE rules of [39] are summarized as reference fault characteristics for typical AHU systems, while in practice much more variables should be taken into consideration. Usually, hundreds to thousands of variables are monitored with threshold rules in active building HVAC systems.

as normal while they do cause damage to the system and lead to economical loss. On the other hand, a high false alarm rate is reported in previous works [38], [42], [43]. Faced with the frequent false alarms in BMS, managers tend to ignore all of them, which makes the rule-based FDD actually useless. What's worse, it takes lots of effort for experts to summarize useful rules that are suitable for all types of systems, while rules are still not perfect enough with individual bias. As a result, in the next step, other kinds of expert knowledge should also be incorporated in the fault gene matrix.

2) System-Specific Expert Knowledge: Li *et al.* [44] pointed out that not all the measured variables under laboratory environment are useful and available in real FDD applications. They proposed to select essential features that optimize FDD accuracy for each fault. It is reported that the subsets of features (sensor variables) should be selected to identify different faults. Also, the selected subsets of features are listed in order of how much they contribute to the FDD. Consequently, in this paper, authors follow the feature selection idea and regard the feature selection results as "system-specific expert knowledge." To be specific, the five most related features are screened for each fault based on the feature selection results of [44], and this type of expert knowledge is encoded accordingly.

3) Data-Driven Expert Knowledge: Since the PCA is a common tool for denoising and dimension reduction, Jolliffe [45] also apply it to extract the principal components of the data. This kind of information is regarded as data-driven expert knowledge, which can be easily obtained. In general, the projection by PCA is a subspace which is linearly combined from the original variables. Although it captures the major variant, no practical meaning is added to the principal components.[3] Besides PCA, other pattern extraction methods include locally linear embedding [46] and partial least squares according to the data characteristics.

Note that expert knowledge is case specific and not limited to the aforementioned types. For example, expert rules are summarized by building experts and ASHRAE researchers in building fields, while they could be general descriptions of objects in research domains. The system-specific knowledge is chosen as the sets of most related features in this paper, which could be other information such as system structure in other cases if it helps with the FDD results.

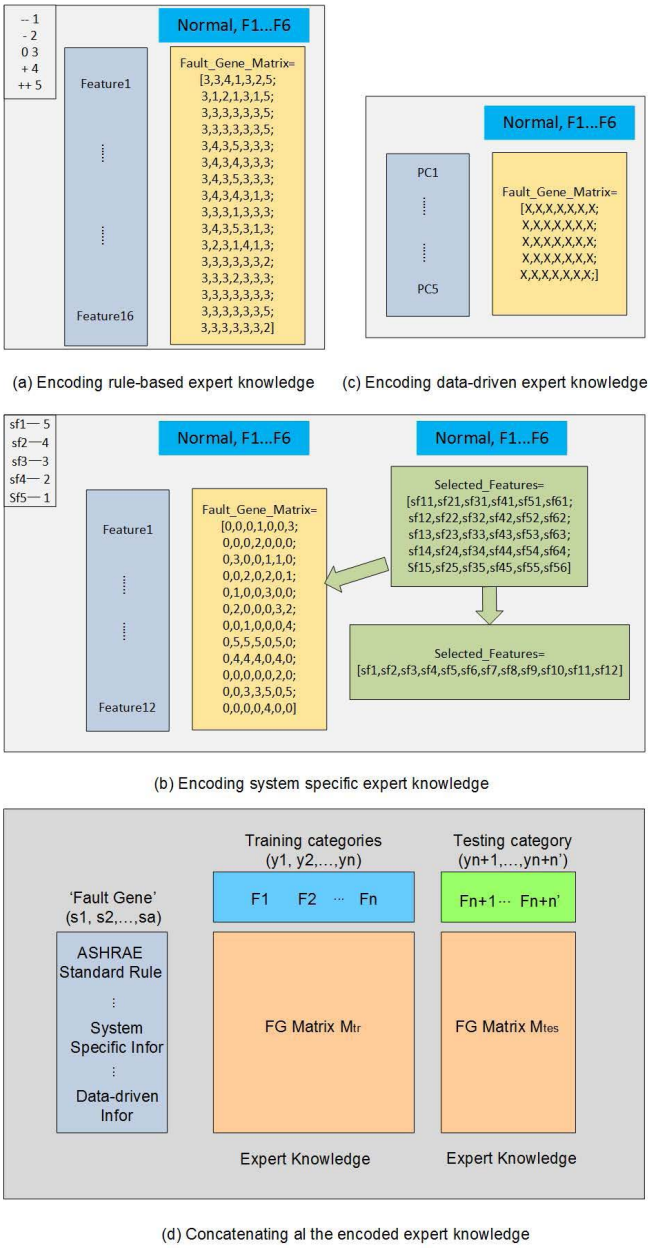


Fig. 2. Example for how to encode the expert knowledge.

C. Encoding

All the aforementioned types of expert knowledge are numerically encoded and concatenated in the fault gene matrix. A simple example for how to encode the expert knowledge is shown in Fig. 2, where 16 terms of rules, 12 items of system-specific knowledge, and 5 principal components are concatenated as the expert knowledge for six fault class as well as one normal class. According to [39], the expert rules were summarized according to how much the variable measurements exit the normal range. Consequently, as shown in Fig. 2(a), rules for values that are below the normal range are encoded as "1" and "2," rules for variables within the normal range are encoded as "3," and rules for variables that are above the normal range are encoded as "4" and "5." Similarly, as shown in Fig. 2(b), the most related feature is encoded as "5" and the less related features are encoded with smaller numbers. If one feature is not included in the five most related features, it is encoded as "0" in the fault gene matrix. Note that the selected subset of features (five most related) might be different but with nonempty intersections. Thus, in the simple example shown in Fig. 2(b), 12 features are assumed to be the final selected features for six faults. As for the data-driven expert knowledge, the five most rated principal components are regarded as the expert knowledge items in this paper. Finally, all the aforementioned fault

[3]Since the PCA-processed data are already numerical, there is no need to encode this type of expert knowledge.

gene encodings are concatenated in the fault gene matrices M_{tr} and M_{tes} for training and testing, respectively.

III. Fault Gene Embedding for EK-UFI Method

A. Fault Gene Embedding

In this section, we introduce how to incorporate the expert knowledge from the mathematical aspect. To begin with, given a set of sensor measurements $\mathcal{S} = \{(x_i, y_i), i = 1, \ldots, n\}$ with $x_i \in \mathcal{X} = \Re^{d\times n}$ and $y_i \in \mathcal{Y} = \Re^{n\times 1}$.[4] Here, n is the number of input measurements and d is the measurement dimension. The prediction objective is to learn a function $f : \mathcal{X} \longrightarrow \mathcal{Y}$ by minimizing an empirical risk function $\frac{1}{n}\sum_{i=1}^{n}\Delta(y_i, z_i = f(x_i))$ where we consider 0/1 loss: $\Delta(y, z) = 0$ if if $y = z$, and $\Delta(y, z) = 1$ otherwise. We consider a general version of discriminant functions $f : \mathcal{X} \times \mathcal{Y}$ and define the prediction function f as

$$f(x; w) = \arg\max_{y\in\mathcal{Y}} F(x, y; w) \tag{1}$$

where w is the model parameter vector of F. As a straight consequence of the linearity[5] of (1), we obtain $F(x, y; w) = \langle w, \Phi(x, y)\rangle$. Since the joint embedding function Φ can be represented by the tensor product between the input embedding (training measurements) θ and output embedding (labels) φ, F can be rewritten as

$$F(x, y; w) = \langle w, \Phi(x, y)\rangle = \langle w, \theta^T(x) \otimes \varphi(y)\rangle \tag{2}$$

where θ and φ are the embedding functions,[6] $\theta : \mathcal{X} \to \tilde{\mathcal{X}} = \Re^{d\times n}$ for inputs, and $\varphi : \mathcal{Y} \to \tilde{\mathcal{Y}} = \Re^{n\times N}$ for outputs (N is the number of training categories). Note that instead of being written as a label vector, output embedding φ is enlarged to the dimension of N for expert knowledge incorporation, which is described in (5). Since $\Phi(x, y)$ is defined as $\mathcal{X} \times \mathcal{Y} : \Re^{d\times n} \times \Re^{n\times N} \to \Re^{d\times N}$, vector w in (1) can be reshaped into a $d \times N$ matrix $\mathbf{w}$. Consequently, F can be formulated as a bilinear form

$$F(x, y; \mathbf{w}) = \langle \mathbf{w}, \Phi(x, y)\rangle = \theta^T(x)\mathbf{w}\varphi(y). \tag{3}$$

Next, assume that we have N training classes, i.e., $\mathcal{C} = \{c_i,\ i = 1, \ldots, N\}$, and a items of expert knowledge, i.e., $\mathcal{K} = \{k_j,\ j = 1, \ldots, a\}$. Those expert knowledge can be used to describe the fault classes with association measures M_{k_j,c_i}, which associates the j^{th} expert knowledge item with the i^{th} training class. Thus, those associations could be encoded into a real-valued matrix (refer to Sections II-B and II-C for more details), which is called the fault gene matrix $M \in \Re^{a\times N}$.[7] The incorporation transformation between the expert knowledge and the class labels is

$$\varphi^{\mathcal{K}}(y) = \varphi(y)M^T. \tag{4}$$

In accordance with the definition of joint embedding $\Phi(x, y)$

$$\Phi^{\mathcal{K}}(x, y) = \theta^T(x) \otimes \varphi^{\mathcal{K}}(y) \tag{5}$$

where $\theta : \mathcal{X} \to \tilde{\mathcal{X}} = \Re^{d\times n}$, and $\varphi^{\mathcal{K}} : \mathcal{Y} \to \tilde{\mathcal{Y}} = \Re^{n\times a}$. Consequently, in line with (3) we obtain

$$\begin{aligned} &F^{\mathcal{K}}(x, y; M, \mathbf{v}) \\ &\quad= \langle \mathbf{v}, \Phi^{\mathcal{K}}(x, y)\rangle = \theta^T(x)\mathbf{v}\varphi^{\mathcal{K}}(y) = \theta^T(x)\mathbf{v}M^T\varphi(y) \end{aligned} \tag{6}$$

where the weight vector $\mathbf{v}$ of $F^{\mathcal{K}}$ can be defined as the latent incorporation matrix that relates the expert knowledge with the fault type predictor.

Now, let us make a slight detour and take a look at the supervised FDD with traditional pattern classification algorithm. Usually, $\mathbf{w}$ in (3) is a weight matrix, which can be learned from the training samples of known faults and corresponding labels. In the prediction stage, the predicted label (fault type) y_t is within the given label set $\{y_i, i = 1, \ldots, n\}$ for any testing set $x_t \in \Re^{d\times n_t}$. Taken in this sense, an unseen fault ($x'_t \nsubseteq \mathcal{S}$) will not be identified correctly by the traditional classifier. Back to our work, with the help of fault gene matrix M associations between labels and expert knowledge are embedded to the discrimination function in (6). In addition, although formulated with different embedding forms, the discrimination with $\mathcal{X}$ and $\mathcal{Y}$ should be the same, namely,

$$F(x, y; \mathbf{w}) = F^{\mathcal{K}}(x, y; M, \mathbf{v}). \tag{7}$$

Thus

$$\mathbf{w} = \mathbf{v}M^T. \tag{8}$$

In such a case, we have

$$F(x, y; M, \mathbf{v}) = \left(\mathbf{v}^T\theta(x)\right)^T(M^T\varphi(y)). \tag{9}$$

In case that the fault gene matrix is redundant, we make use of the compatibility function (9). The latent incorporation matrix $\mathbf{v}$ is learned from training examples $x_i \in \Re^{d\times n}$ with labels $y_i \in \Re^{n\times N}$ jointly with given $M_{\text{tr}} \in \Re^{a\times N}$. As a simpler alternative, it is possible to first try to learn by performing a singular value decomposition on the $\mathbf{w}$ matrix, and then learn $\mathbf{v}$.

B. Offline Model Training

Based on the formulation in last section, our learning objective is to find the optimal latent incorporation matrix $\mathbf{v}$ which is calculated from the optimal weight matrix $\mathbf{w}$ and the given fault gene matrix M. Thus, we can still consider $\mathbf{w}$ as the optimization aim for two classes

$$f(x; \mathbf{w}) = \arg\max_{y\in\mathcal{Y}} F(x, y; \mathbf{w}). \tag{10}$$

Derived from the formulation of [22] and [23], we define the multiclass margin of a training data sample (x_i, y_i) with respect to $\mathbf{w}$ as

$$m_i(\mathbf{w}) = F(x_i, y_i; \mathbf{w}) - \max_{y\neq y_i} F(x_i, y; \mathbf{w}). \tag{11}$$

[4] Usually, normal condition is included in the training classes [44].

[5] F is generally assumed to be linear in the combined feature embedding of inputs and outputs.

[6] Here, embedding functions of $\mathcal{X}$ and $\mathcal{Y}$ represent the data preprocessing (normalization and outlier removing [22]) of the training data.

[7] $M \in \Re^{a\times N}$ indicates that the fault gene matrix includes "a" items of expert knowledge for N training classes.

Note that the correct classification of (x_i, y_i) requires a positive margin. Thus, the maximum margin principle for classification can be applied to determine the optimal vector $\mathbf{w}^*$ which leads to the maximum class separation

$$\mathbf{w}^* = \underset{\mathbf{w}:\|\mathbf{w}\|=1}{\arg\max} \ \min \sum_{i=1}^{n} m_i(\mathbf{w}). \tag{12}$$

Consider a category dependent cost $\Delta(y_i, y)$ for misclassifying y_i as y, (12) can be equivalently written as a norm minimization problem, which is augmented by slack variables for possible margin violations by outliers. Thus, we obtain the following L_2 regularized soft-margin objective[8]

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}\xi_i + \lambda\|\mathbf{v}\|^2 \tag{13}$$

$$\text{s. t.} \begin{cases} \xi_i \geq 0 \\ \|\mathbf{v}^T x_i\|^2 \leq B \\ m_i(\mathbf{w}) \geq 1 - \dfrac{\xi_i}{\Delta(y_i, y)} \end{cases} \quad \forall\, i \tag{14}$$

where C is a hyperparameter that tunes the margin loss penalty, B is used to bound the latent incorporation matrix $\mathbf{v}$ to make it invariant enough along different training distributions. Note that $\mathbf{v}$ is just a partial representative of the universal connection between faults and fault gene. The more training sets (known fault types) we have, the more realistic $\mathbf{v}$ we can obtain. More discussion about this point is presented in Section IV-D.

Next, according to (3) and (11), $m_i(\mathbf{w})$ is derived as

$$\begin{aligned} m_i(\mathbf{w}) &= \langle \mathbf{w}, \Phi(x_i, y_i)\rangle - \max_{y\neq y_i}\langle \mathbf{w}, \Phi(x_i, y)\rangle \\ &\leq \langle \mathbf{w}, \Phi(x_i, y_i)\rangle - \langle \mathbf{w}, \Phi(x_i, y)\rangle (\forall y \neq y_i). \end{aligned} \tag{15}$$

Letting $\delta\Phi_i(y) \triangleq \Phi(x_i, y_i) - \Phi(x_i, y)$, for $\forall i, \forall y \neq y_i$, we can get

$$\langle \mathbf{w}, \delta\Phi_i(y)\rangle \geq m_i(\mathbf{w}) \geq 1 - \frac{\xi_i}{\Delta(y_i, y)}. \tag{16}$$

Thus, the second constraint of (14) can be rewritten as

$$\langle \mathbf{w}, \delta\Phi_i(y)\rangle - 1 + \frac{\xi_i}{\Delta(y_i, y)} \geq 0. \tag{17}$$

The dual formulation of the above-mentioned primal problem with the Lagrangian multiplier method is

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \xi, \alpha, \eta) = &\frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}\xi_i + \lambda\|\mathbf{v}\|^2 \\ &- \sum_{i=1}^{n}\beta_i(B - \|\mathbf{v}^T x_i\|^2) - \sum_{i=1}^{n}\eta_i\xi_i \\ &- \sum_{i=1}^{n}\sum_{y\neq y_i}\alpha_{iy}\left(\langle \mathbf{w}, \delta\phi_i(y)\rangle - 1 + \frac{\xi_i}{\Delta(y_i, y)}\right). \end{aligned} \tag{18}$$

[8]Note that in this section the formulation of the soft-margin problem [(11) and (12)] and the constraint derivation [(15), (16), and (17)] follow the formulation and derivation in [22] and [23]. One of the novelties of the proposed EK-UFI method is to integrate the expert knowledge incorporation matrix $\mathbf{v}$ into the optimization functions [(13) and (18)].

As for the optimization, we follow [47] and use stochastic gradient descent to train embedding models from (9), which is fast by sampling both training samples and classes.

C. Online Fault Identification

Once the latent incorporation matrix $\mathbf{v}$ has been learned from the training class $\mathcal{S}$, the labels of unseen classes can be predicted based on $\mathbf{v}$ and the corresponding fault gene matrix M_{tes}. Thus, the following discrimination function for the unseen classes $\mathcal{S}' = \{x'_j, j = 1, \ldots, m\}$ can be obtained:

$$F(x', y; M_{\text{tes}}, \mathbf{v}) = \theta^T(x')\mathbf{v}M_{\text{tes}}^T\varphi(y). \tag{19}$$

Since x', M_{tes} and $\mathbf{v}$ in (19) are known, the label vector $y' = \arg\max_y \ F\left(x', y; M_{\text{tes}}, \mathbf{v}\right)$ can be accordingly predicted. The online update for the exert knowledge-based UFI method is summarized in Algorithm 1.

Algorithm 1 Online Update Algorithm for EK-UFI

Input $\mathcal{S} = \{(x_i, y_i), i = 1, \ldots, n\}$
$\mathcal{S}' = \{x'_j, j = 1, \ldots, t, t+1\}$
$M_{\text{tr}}, M_{\text{tes}}$
for $i = 1, \ldots, n$ **do**
$m_i(\mathbf{w}) = F(x_i, y_i; \mathbf{w}) - \max_{y\neq y_i} F(x_i, y; \mathbf{w})$
$\mathbf{w}^* = \underset{\mathbf{w}:\|\mathbf{w}\|=1}{\arg\max} \ \min \sum_{i=1}^{n} m_i(\mathbf{w})$
$\mathbf{v} = \mathbf{w}^* M_{\text{tr}}$
end for
$L_{t+1} \Leftarrow \emptyset$
while $L_1 \ldots L_{t+1}$ still change **do**
for $j = t+1 : -1 : 1$ **do**
$F\left(x'_j, y; M_{\text{tes}}, \mathbf{v}\right) = \left(\mathbf{v}^T\theta\left(x'_j\right)\right)^T\left(M_{\text{tes}}^T\varphi(y)\right)$
$y' = \arg\max_y \ F\left(x'_j, y; M_{\text{tes}}, \mathbf{v}\right)$
$L_i \leftarrow L_i \cup \{y'\}$
end for
end while
Output $L_1 \ldots L_t, L_{t+1}$

IV. Experimental Results and Discussion

A. AHU and Faults

AHU is the key component for maintaining comfortable and healthy IAQ through connecting outdoor and indoor environment with heating and cooling plants [48], [49]. A typical single-duct VAV AHU system with monitoring and control devices is shown in Fig. 3. Major components and devices in this system are: supply/return fans (RFs), heating/cooling/preheat coils, heating/cooling/preheat coil valves, recirculated/exhaust/outdoor air dampers, deployed sensor instrumentation, and ducts that transfer the air to and from the conditioned spaces [50].

The outside air first passes through coils with an amount of heat that is added or removed according to heating/cooling requirements. This air is then drawn through the supply fan (SF), which is equipped with a variable frequency drive,

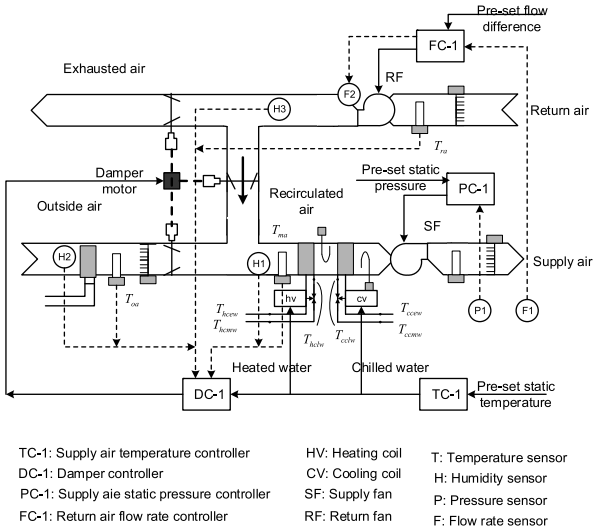


Fig. 3. Typical single-duct VAV AHU system. The VAV system maintains the supply air temperature (T_{sa}) to the terminals for air conditioning. T_{sa} is measured and compared with preset temperature of TC-1. The control is linked to dc-1 in order to automatically operate Outside Air damper and Return Air damper for appropriately mixing temperature (T_{ma}) before entering the coil.

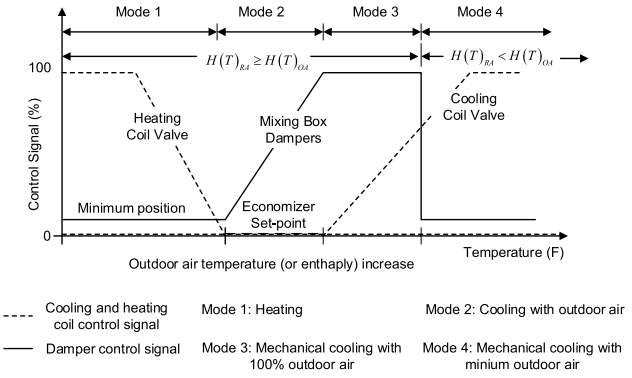


Fig. 4. Operating modes of AHU. An economizer set point can be an outdoor temperature set point, a combination of outdoor temperature and humidity set points, or an outdoor enthalpy set point. When the outdoor temperature (and humidity) is above the economizer set point, the outdoor air intake will be a minimum quantity just to satisfy the ventilation requirement.

and distributed to the zones according to different building load conditions. Then, the return air from the rooms is drawn by the RF, which is also equipped with a VFD, and partially recirculated to mix with the outside intake air and partially exhausted to the outside environment. The exhaust air, recirculated air, and outdoor air dampers are used to regulate the air flow in the AHU.

On the other hand, AHU operating modes change in accordance with the seasonal outdoor air temperature and humidity and supply air temperature (heating or cooling) set points. There are four different modes as shown in Fig. 4. In the mechanical heating mode (Mode 1), the outdoor air damper is maintained at its minimum position. The heating coil valve is controlled to keep the supply air temperature at the heating set point, and the cooling coil valve is closed. In the free cooling mode (Mode 2), both heating and cooling coil valves are closed. The outdoor air dampers are modulated to maintain the

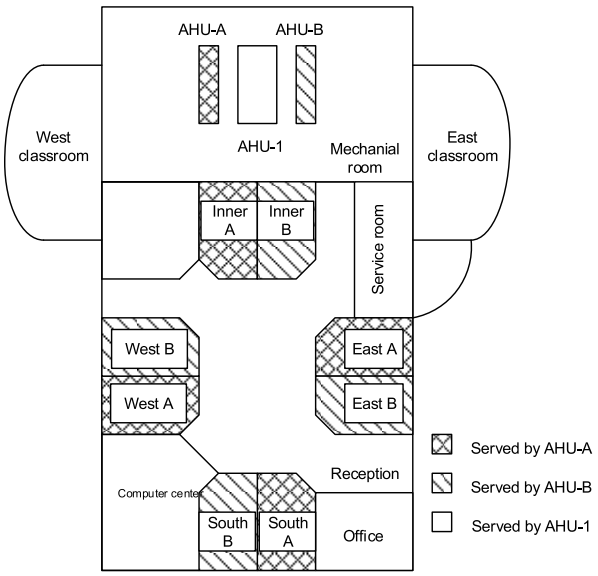


Fig. 5. Layout of ERS. AHU-A and AHU-B are identical, and each AHU serves four zones. Three of the four zones have external exposures and one only gets internal conditions. The A and B zones are the mirror images with identical external thermal loads.

supply air temperature at its set point with the outdoor air only. In the mechanical and economizer cooling mode (Mode 3), the outdoor air damper is fully open. The cooling coil valve is modulated to maintain the supply air temperature at the cooling set point. In the mechanical cooling mode (Mode 4), the outdoor air damper is fixed at the minimum position since the outdoor air temperature cannot meet the economizer set point. The cooling coil valve is modulated to maintain the supply air temperature at the cooling set point.

Due to the heavy workload, there is a high chance for AHU to encounter hardware failures and controller errors due to improper system design, configuration, and operation. According to their causes and locations, faults are mainly divided into four categories, i.e., faults in AHU equipment, actuators, sensors, and feedback controllers [10]. As listed in the third columns of Tables I–III, 25 typical AHU faults that are commonly encountered in three seasons (11 typical faults occur in Spring, 8 typical faults occur in Summer, and 6 typical faults occur in Winter) are studied in this paper.

B. Data Description

The real-time data of ASHRAE RP-1312 are adopted as the experimental data in this paper. ASHRAE RP-1312 was implemented in the test facility at the energy resource station (ERS) [51], [52]. As a brief introduction, RP-1312 conducted several on-site experiments to emulate the dynamic behaviors of a single duct dual fan VAV AHU system serving four building zones under various seasonal conditions. The experimental data under normal and typical faulty status is archived so as to be used in the future research. Interested readers can refer to [53] for the details about the test facility provided by Price and Smith. As shown in Fig. 5, the experiment involved two identical AHUs, i.e., AHU-A and AHU-B, which served as treatment and control groups, respectively.

The testing space included Inner A & B, West A & B, South A & B, and East A & B. Faults were manually introduced into the air-mixing box, coils, and fan sections of AHU-A, while AHU-B was operated at nominal states. During each experiment, the system operation was scheduled "ON" during the occupied period from 6:00 to 18:00 and "OFF" during the unoccupied period from 18:00 to 6:00. Sensor measurements were recorded every 2 min. All the experiments were conducted under the real weather and building load conditions during spring, summer, and winter, respectively. Details about how those faults are implemented can be found in [39].

C. Experimental Setup

With the RP-1312 data, EK-UFI method is applied to identify unseen faults for building AHU systems. During each test round, the unseen fault (the expert knowledge about that fault is known) is identified by incorporating different sources of expert knowledge to both the known and unseen faults in the training and testing processes, respectively. Authors compare the UFI accuracy in terms of different sources of expert knowledge, namely, ASHRAE rule-based, system specific, and data-driven expert knowledge, and the combinations of them, which will be introduced in Section II-B.

In this paper, the accuracy of identifying an unseen fault equals to the prediction accuracy between fault and normal. To be specific, the UFI goal is to estimate the chance that the predictor $f(x)$ is correct on unseen data, i.e., the generalization performance of the predictor. In this paper, the empirical accuracy on a batch testing data set is used as an unbiased estimator. Let sign[., .] be 1 if the predicted label of one testing data point accorded with its original label and 0 otherwise, the testing accuracy is

$$\text{Accu}(f) = \frac{1}{n}\sum_{i=1}^{n} \text{sign}[f(x_j), y_j] \tag{20}$$

$$\text{sign}[f(x_j), y_i] = \begin{cases} 1, & f(x_j) = y_j \\ 0, & f(x_j) \neq y_j \end{cases} \tag{21}$$

where $f(x_j)$ is the predicted label for testing data point x_j, which represents that the data point is recognized as a certain severity level of one fault type, and y_j is the true label that records the real experiment condition.

In the remaining part, authors consider the one versus all situation, namely, only one available typical fault is regarded as unknown while the others are known. The training data set and testing data set are randomly picked from the RP-1312 data files. In each UFI round, expert knowledge of known faults is first incorporated with the training data to learn the latent incorporation matrix $\mathbf{v}$. Then, the learned incorporation matrix is applied to predict the testing data set based on the expert knowledge of the unseen fault. Note that in this paper, "unseen fault" means that information (including data and expert knowledge) about that fault is not included in the training stage. The training-testing round is repeated 20 times with randomly chosen samples. The final UFI accuracy is the average value of all the training-testing rounds.

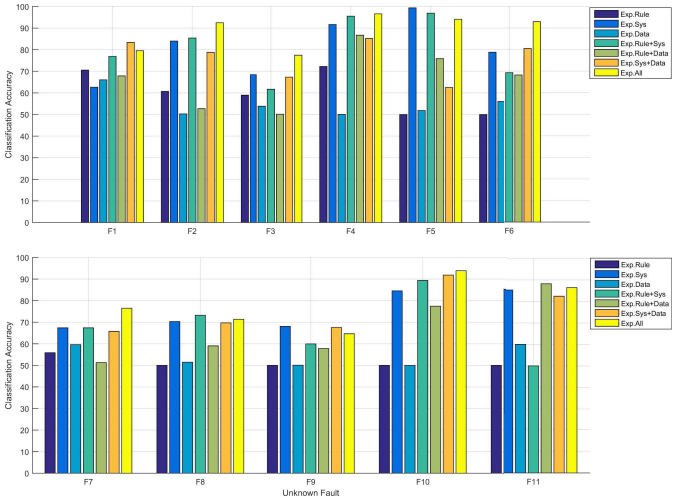


Fig. 6. Spring case: eleven typical faults. UFI accuracy with different incorporated EK. "Exp.Rule" means only the ASHRAE rules are incorporated as fault gene. "Exp.Sys" means only the system specific knowledge is incorporated as fault gene. "Exp.Data" means only the data-driven information is incorporated as fault gene. "Exp.All" means all the above-mentioned sources of expert knowledge are incorporated.

D. Unseen Fault Identification

Based on the real-time data collected by ASHRAE RP-1312 in three seasons (spring, summer, and winter), the proposed EK-UFI is tested in the one versus all situation (namely, only one fault is regarded as unknown while the others are known), and different combinations of expert knowledge are tested separately. Based on the testing accuracy of all the training-testing rounds, it is seen that unseen faults could be successfully identified with satisfactory accuracy (about 50%–99%). Note that the accuracy of random guess is 8.33%, 11.11%, and 14.29% for the spring, summer, and winter, respectively. In the remaining part, the results are analyzed in terms of sources of expert knowledge and the number of known faults.

1) Comparison Among Sources of Expert Knowledge: As shown in Figs. 6–8, UFI accuracy values with different incorporated expert knowledge are displayed by bar plots for the three seasonal cases, respectively. Each fault is regarded as an unseen fault once while the others are known. We compare the UFI accuracy for each fault with three different sources of expert knowledge and their combinations. Generally speaking, the UFI accuracy with system-specific expert knowledge is higher than that with rule-based expert knowledge. The proposed EK-UFI method performs the worst when incorporating the data-driven expert knowledge alone. This can be explained by the ambiguity of PCA. Usually, there is no theoretical guarantee for the PCA-projected space to be the optimal representation of the original data. Next, the differences among the three sources of expert knowledge are compared with details.

1) In the spring case, as shown in Fig. 6, the proposed EK-UFI method performs so by incorporating rule-based expert knowledge with the average identification accuracy less than 75% for each fault. Comparing the three sources of expert knowledge alone (shown by the first three bars of each fault), the average UFI accuracy

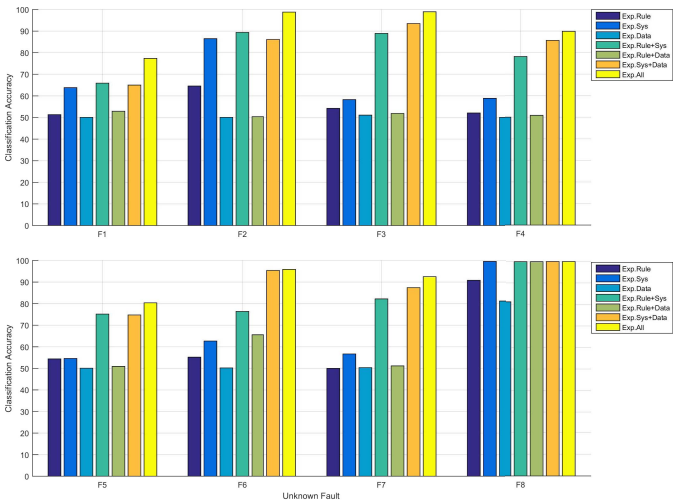


Fig. 7. Summer case: eight typical faults. UFI accuracy with different incorporated EK. Four cases of different expert knowledge sources are considered. "Exp.Rule" means only the ASHRAE rules are incorporated as fault gene. "Exp.Sys" means only the system specific knowledge is incorporated as fault gene. "Exp.Data" means only the data-driven information is incorporated as fault gene. "Exp.All" means all the above-mentioned sources of expert knowledge are incorporated.

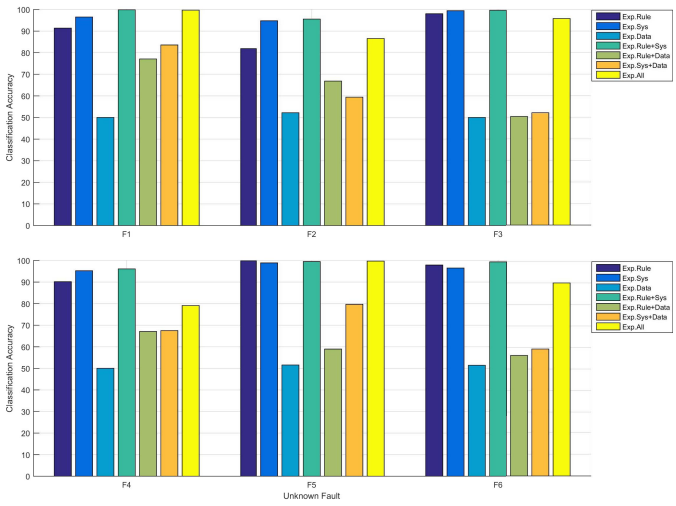


Fig. 8. Winter case: six typical faults. UFI accuracy with different incorporated EK. Four cases of different expert knowledge sources are considered. "Exp.Rule" means only the ASHRAE rules are incorporated as fault gene. "Exp.Sys" means only the system specific knowledge is incorporated as fault gene. "Exp.Data" means only the data-driven information is incorporated as fault gene. "Exp.All" means all the above-mentioned sources of expert knowledge are incorporated.

with system-specific expert knowledge is the highest for all except for Fault 1, where UFI accuracy with rule-based expert knowledge is 70.85% while those with the other two sources are 62.66% and 66.66%, respectively. Especially, the average UFI accuracy with system-specific expert knowledge for faults 4 and 5 is as high 95.55% and 96.93%. We also see in Fig. 6 that for most of the faults (Fault 2–10), the UFI accuracy with data-driven expert knowledge is lower than 60%, which is relatively poor compared with other expert knowledge sources.

2) In the summer case, as shown in Fig. 7, the identification performance for unseen Fault 8 with all types of expert knowledge combinations are impressive (near 100% with most of the combinations, and the accuracy with data-driven expert knowledge is 81.22%). With more scrutiny, we observe that, for all the faults, by combining rule-based expert knowledge and system specific expert knowledge, the UFI accuracy can be improved (2.08%–36.70% improvement). Similar improvements can also be achieved by combining rule-based expert knowledge/system specific expert knowledge with the data-driven expert knowledge. To be specific, the improvement of adding rule-based expert knowledge to data-driven expert knowledge is 0.02%–18.64%. The improvement of adding system-specific expert knowledge to data-driven expert knowledge is 14.95%–42.40%. The improvement of adding both rule-based and system-specific expert knowledge to data-driven expert knowledge is 18.71%–48.75%. Moreover, under all the unseen fault situations, UFI accuracy with system-specific expert knowledge is higher than that with rule-based expert knowledge by 0.21%–21.91%. As a result, we can conclude that system-specific expert knowledge and the combinations with it are more suitable for identifying unseen faults in the summer case.

3) In the winter case, as shown in Fig. 8, the proposed EK-UFI method presents an ordinary performance with data-driven expert knowledge and combinations with it (50.03%–51.57%, 50.38%–77.17%, and 52.12%–83.59%, respectively) under all the six unseen fault situations. The UFI performance with rule-based and system-specific expert knowledge is obviously higher (81.91%–98.05% and 95.25%–99.44%, respectively). Also, by combining rule-based expert knowledge and system-specific expert knowledge together, 0.35%–8.33% accuracy improvement can be achieved. However, by adding all the three types of expert knowledge together, the UFI accuracy is slightly worse (0.14%–17.00%) than that with rule-based and system specific expert knowledge. This is because data-driven expert knowledge performs poorly compared with other two in the winter case. Note that an accuracy around 50% is still theoretically acceptable since random guess for the unseen fault to be right is 16.67% in winter case, and even lower in summer and spring cases (12.50% and 9.09%, respectively).

The sensitivity of the EK-UFI method to different combinations of expert knowledge is summarized in Tables I–III. In those tables, "+" represent 50%–60% UFI accuracy, "++" represent 60%–70% UFI accuracy, "+++" represent 70%–80% UFI accuracy, "++++" represent 80%–90% UFI accuracy, and "+++++" represent 90%–100% UFI accuracy. It can be observed that in the spring case, the EK-UFI method is more sensitive to the system-specific expert knowledge and the combinations with it. In the summer case, the EK-UFI method is more sensitive to the combinations without data-driven expert knowledge, and unseen Fault 8 is sensitive to all the three types of expert knowledge and

TABLE I

SUMMARY OF UFI RATE IN TERMS OF DIFFERENT SOURCES OF EXPERT KNOWLEDGE IN SPRING

Category	No.	Fault Description	Rule	Sys	Data	Rule+Sys	Rule+Data	Sys+Data	All
EF	F1	Air filter blockage (25%)	+++	++	++	+++	++	++++	+++
AF	F2	OA damper stuck (fully close)	++	++++	+	++++	+	+++	+++++
	F3	EA damper stuck (fully close)	+	++	+	++	+	++	+++
	F4	Cooling coil valve stuck (fully open)	+++	+++++	+	+++++	++++	++++	+++++
SF	F5	OA temperature sensor bias +3F	++	+++++	+	+++++	+++	++	+++++
	F6	OA temperature sensor bias -3F	+	+++	+	+	++	++++	+++++
CF	F7	Return fan at fixed speed (80%)	+	++	+	++	+	++	+++
	F8	Mixed air damper unstable	+	+++	+	+++	+	++	+++
	F9	Cooling Coil Control Unstable	+	++	+	+	+	++	++
	F10	Heating and cooling unstable	+	++++	+	++++	+++	+++++	+++++
	F11	Supply fan control unstable	+	++++	++	+	++++	++++	++++

*Note: EF-equipment faults, AF-actuator faults, SF-sensor faults, CF-controller faults.

*Note: Rule-rule-based expert knowledge, Sys-system specific expert knowledge, Data-data-driven expert knowledge.

TABLE II

SUMMARY OF UFI RATE IN TERMS OF DIFFERENT SOURCES OF EXPERT KNOWLEDGE IN SUMMER

Category	No.	Fault Description	Rule	Sys	Data	Rule+Sys	Rule+Data	Sys+Data	All
EF	F1	AHU duct leak after supply fan	+	++	+	+	+	++	+++
	F2	AHU duct leak before supply fan	++	++++	+	++++	+	++++	+++++
AF	F3	OA damper stuck (fully close)	+	+	+	++++	+	+++++	+++++
	F4	OA damper leak (55%)	+	+	+	+++	+	++++	++++
	F5	EA damper stuck (fully close)	+	+	+	+++	+	+++	++++
	F6	Cooling coil valve stuck (fully open)	+	++	+	+++	++	+++	+++++
CF	F7	Cooling coil valve control unstable	+	+	+	++++	+	++++	+++++
	F8	Cooling coil valve reverse action	+++++	+++++	++++	+++++	++++++	+++++	+++++

*Note: EF-equipment faults, AF-actuator faults, CF-controller faults.

*Note: Rule-rule-based expert knowledge, Sys-system specific expert knowledge, Data-data-driven expert knowledge.

TABLE III

SUMMARY OF UFI RATE IN TERMS OF DIFFERENT SOURCES OF EXPERT KNOWLEDGE IN WINTER

Category	No.	Fault Description	Rule	Sys	Data	Rule+Sys	Rule+Data	Sys+Data	All
EF	F1	Heating coil fouling	+++++	+++++	+	+++++	+++	++++	+++++
	F2	Heating coil reduced capacity	++++	+++++	+	+++++	++	+	++++
AF	F3	OA damper stuck (fully close)	+++++	+++++	+	+++++	+	+	+++++
	F4	OA damper leak (62%)	+++++	+++++	+	+++++	++	++	+++
	F5	EA damper stuck (fully close)	+++++	+++++	+	+++++	+	+++	+++++
	F6	Cooling coil valve stuck (fully open)	+++++	+++++	+	+++++	+	+	++++

*Note: EF-equipment faults, AF-actuator faults.

*Note: Rule-rule-based expert knowledge, Sys-system specific expert knowledge, Data-data-driven expert knowledge.

their combinations. In the winter case, the EK-UFI method is more sensitive to both rule-based and system-specific expert knowledge, while the unseen fault performance with data-driven expert knowledge and combinations with it is not satisfactory.

2) Comparison via Number of Known Faults: As mentioned in Section II, the latent incorporation matrix $\mathbf{v}$ is just a partial representation of the universal connection between faults and fault gene. Here, we first test the "in-variance" of the latent incorporation matrix (take the Spring case as an example). In Fig. 9, singular values of latent incorporation matrices under different faulty conditions were plotted as a function of the singular value sequence number. Since 11 typical AHU faults are studied in the Spring case, we trained eleven latent incorporation matrices according to different known data sets (10 faults as known and 1 fault as unknown). In Fig. 9(a), the first 11 singular values[9] are plotted for all the 11 matrices (there are 11 dots in each column). Fig. 9(b) is the corresponding box plot for each singular value sequence. For better understanding, we also plotted the log values for all the singular values in Figs. 9(c) and (d). We could see that the largest difference is less than 0.15. Since singular values are scales for the original matrix [54], we can claim the difference among the tested latent incorporation matrices is small. Recall that in (14) $\mathbf{v}$ is bounded by a constraint to make the latent incorporation matrix invariant enough along different training distributions. This is verified in Fig. 9.

Fig. 10 shows the identification accuracy as a function of the number of known faults. Based on the given faults emulated by ASHRAE researchers, the combinations of known faults with different numbers would be complicated. For example, in the spring case (where 11 typical faults are emulated in the ASHRAE library), if Fault 1 is treated as the unseen fault and two faults from the fault library would be chosen as known faults, there could be $C_{10}^2 = 45$ possible combinations of known faults. For simplicity, in this paper, we only consider randomly chosen cases. We can generally observe an

[9]There are 35 singular values for each latent incorporation matrix $\mathbf{v}$, since each $\mathbf{v}$ is a 35×51 matrix in the calculation process for the Spring case. Except for the 11 plotted singular values, others are near to 0 (less than 10^{-10}).

TABLE IV
COMPARISON BETWEEN EXISTING BUILDING FDD METHODS AND THE PROPOSED EK-UFI

Category	Examples	KFD	KFI	UFD	UFI	Propagation
Quantitative Model-based Methods	Fault detection with chiller models [55]	✓		✓		
	Condition monitoring in HVAC subsystems [56]	✓		✓		
Qualitative Model-based Methods	Rule-based FDD for HVAC systems [42], [43]	✓	✓			
Statistical Methods	CUSUM [57]	✓		✓		✓
	EWMA [58]	✓		✓		✓
	SPC [2]	✓		✓		✓
Unsupervised Learning Methods	PCA [40], [41]	✓		✓		✓
Supervised Learning Methods	Multivariate regression [13]	✓	✓			✓
	Bayes classifiers [14]	✓	✓			✓
	NN [15]	✓	✓			✓
	LDA+Range monitoring [16]	✓	✓	✓		✓
	SVM [20]	✓	✓			✓
Novel Methods	EK-UFI	✓	✓	✓	✓	✓

*Note: KFD-Known Fault Detection, KFI-Known Fault Identification, UFD-Unseen Fault Detection, UFI-Unseen Fault Identification.

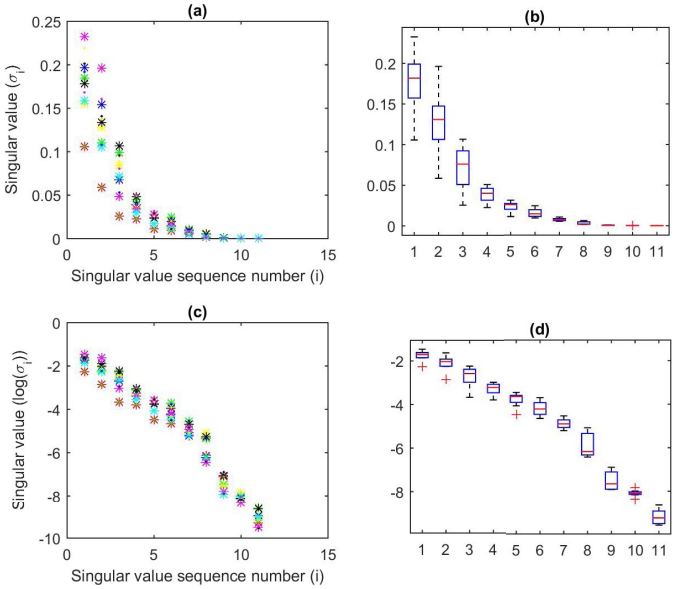


Fig. 9. Testing of the "in-variance" of the latent incorporation matrix, as an example in the Spring case. Singular values of latent incorporation matrices under different conditions were plotted as a function of singular value sequence number. (a) Singular values of different latent incorporation matrices with stacked points. (b) Distributions of singular values for different latent incorporation matrices with box-plots. (c) and (d) Log singular values with stacked points and box-plots, respectively.

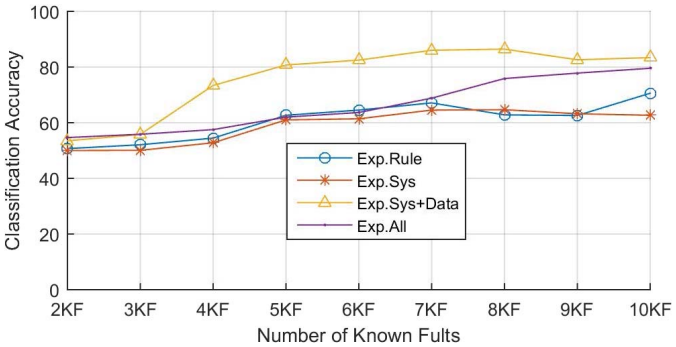


Fig. 10. Prediction accuracy of unseen fault as a function of number of known faults in the Spring case.

increasing trend of UFI accuracy along with the accumulation of known faults in Fig. 10.

Those observations reveal that in most of the situations more known faults means better EK-UFI performance. However, after careful scrutiny, we understand that the number of known faults is not the single factor that influences the EK-UFI performance. One similar and the related known fault will be more helpful for recognizing the unseen fault than several totally irrelevant known faults. For example, the UFI accuracy is higher when controller faults are taken as the known faults than that with equipment faults as known faults. Moreover, as discussed in Section II-B, EK-UFI performance is also related to different sources of expert knowledge. How to choose known faults and how to decide the incorporated expert knowledge are an interesting topic which deserves effort in the future work.

3) Comparison With Existing Building FDD Methods: As given in Table IV, the proposed novel EK-UFI method is compared with the existing fault detection and identification (FDI)[10] methods in smart building field. Traditionally, building FDI methods can be summarized as quantitative model-based methods (physical model-based methods), qualitative model-based methods (rule-based methods), statistical methods, and machine learning methods [59]. The proposed EK-UFI method is a novel machine learning method that combines the advantages of recent machine learning and traditional *a priori* knowledge-based detection techniques (rule-based methods). The key advantages of the proposed EK-UFI methods and its main difference from the existing FDD methods are summarized as follows.

1) Physical model-based methods can provide accurate estimation results if they are well formulated and the estimation function enables unseen fault detection. However, it is hard for physical model-based methods to identify fault types since modeling all the faulty operations based on first principles is not easy. Traditional statistical methods, such as cumulative sum control chart (CUSUM), exponentially weighted moving average (EWMA) method, and statistical process control (SPC) method, can find abnormal processing

[10]Also known as FDD

points based on control ranges estimated from the normal historical data. Notwithstanding, statistical methods cannot identify corresponding fault types since information about faults is not included in the calculation/modeling/training process. A similar situation can be observed for unsupervised machine learning methods, whose model is trained with normal data.

2) With bounds of monitoring thresholds defined by a prior knowledge, rule-based methods can detect known faults and identify fault types simultaneously. However, pure rule-based methods are not suitable for the identification of unseen faults. Through learning from both normal and labeled abnormal data, supervised machine learning methods can detect and diagnose known faults with decent accuracy. Although they have been adequately applied in building FDI, supervised learning methods cannot distinguish unseen faults due to a lack of training data. However, by combining supervised learning methods with detection ranges, one could recognize unseen faults without identifying their types.
3) Since the physical models and expert-summarized rules are system specific and require fine development and definition, it is energy wasting and time consuming to extend the related FDI methods to analogous equipment and systems. On the contrary, with reliable training data, the data-driven FDI framework (statistical and machine learning methods) can be easily applied to similar systems.
4) One of the advantages of the proposed UK-FDI method is its ability to identify unseen faults with acceptable accuracy. However, UK-FDI is dependent on how well the expert knowledge can describe the fault categories in a high level. The method would not be useful without reliable expert knowledge.

V. Conclusion

In this paper, the EK-UFI method is proposed to identify unseen building faults by incorporating various sources of expert knowledge which is encoded in the fault gene matrix. The EK-UFI method is designed to recognize unseen faults by incorporating expert knowledge, provided with its description in terms of numerical matrices. By applying EK-UFI to identify unseen AHU faults with real data measured by ASHRAE RP-1312, it is proven to be effective with high identification accuracy. Besides building communities, the proposed framework can also been extend to identify unseen faults, anomalous working conditions or malfunctions for other mechanical systems and cyberphysical systems, such as smart vehicles, smart water treatment system, smart power system, and so on.

Through comparing the identification accuracy values with rule-based expert knowledge, system-specific expert knowledge, and data-driven expert knowledge, authors show that different sources of expert knowledge lead to different UFI performances. It is interesting to study how to select the essential expert knowledge in the future work. Moreover, by accumulating the number of known faults, authors also found the EK-UFI performance varies as a function of the number of known faults. This reveals the value of studying how to make the best use of known faults in the ongoing work.

References

[1] K. Weekly, M. Jin, H. Zou, C. Hsu, A. Bayen, C. Spanos. (2014). "Building-in-briefcase (BiB)." [Online]. Available: https://arxiv.org/abs/1409.1660

[2] B. Sun, P. B. Luh, Q.-S. Jia, Z. O'Neill, and F. Song, "Building energy doctors: An SPC and Kalman filter-based method for system-level fault detection in HVAC systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 1, pp. 215–229, Jan. 2014.

[3] Y. Yan, P. B. Luh, and K. R. Pattipati, "Fault diagnosis of HVAC air-handling systems considering fault propagation impacts among components," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 705–717, Apr. 2017.

[4] S. M. Namburu, M. S. Azam, J. Luo, K. Choi, and K. R. Pattipati, "Data-driven modeling, fault diagnosis and optimal sensor selection for HVAC chillers," *IEEE Trans. Autom. Sci. Eng.*, vol. 4, no. 3, pp. 469–473, Jul. 2007.

[5] S. Katipamula and M. R. Brambley, "Review article: Methods for fault detection, diagnostics, and prognostics for building systems—A review, part II," *HVAC R Res.*, vol. 11, no. 2, pp. 169–187, 2005.

[6] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2226–2238, Nov. 2013.

[7] S. X. Ding, *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms and Tools*. New York, NY, USA: Springer, 2008.

[8] B. R. Vernon, "Failure accomodation in linear systems through self-reorganization," Ph.D. dissertation, Dept. Aeronaut. Astronaut., Massachusetts Inst. Technol., Cambridge, MA, USA, 1971.

[9] C. Jie and R. J. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*, vol. 3. New York, NY, USA: Springer, 2012.

[10] Y. Yu, D. Woradechjumroen, and D. Yu, "A review of fault detection and diagnosis methodologies on air-handling units," *Energy Buildings*, vol. 82, pp. 550–562, Oct. 2014.

[11] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.

[12] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015.

[13] G. Mustafaraj, J. Chen, and G. Lowry, "Development of room temperature and relative humidity linear parametric models for an open office using BMS data," *Energy Buildings*, vol. 42, no. 3, pp. 348–356, Aug. 2010.

[14] F. Xiao, Y. Zhao, J. Wen, and S. Wang, "Bayesian network based FDD strategy for variable air volume terminals," *Automat. Construct.*, vol. 41, pp. 106–118, May 2014.

[15] Z. Du, B. Fan, X. Jin, and J. Chi, "Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis," *Building Environ.*, vol. 73, pp. 1–11, Mar. 2014.

[16] D. Li, G. Hu, and C. J. Spanos, "A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis," *Energy Buildings*, vol. 128, pp. 519–529, Sep. 2016.

[17] P. Jaikumar, A. Gacic, B. Andrews, and M. Dambier, "Detection of anomalous events from unlabeled sensor data in smart building environments," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 2268–2271.

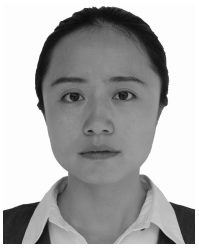
[18] Y. Zhao, S. Wang, and F. Xiao, "Pattern recognition-based chillers fault detection method using support vector data description (SVDD)," *Appl. Energy.*, vol. 112, pp. 1041–1048, Dec. 2013.

[19] Y. Zhao, F. Xiao, J. Wen, Y. Lu, and S. Wang, "A robust pattern recognition-based fault detection and diagnosis (FDD) method for chillers," *HVAC R Res.*, vol. 20, no. 7, pp. 798–809, 2014.

[20] T. Mulumba, A. Afshari, K. Yan, W. Shen, and L. K. Norford, "Robust model-based fault diagnosis for air handling units," *Energy Buildings*, vol. 86, pp. 698–707, Jan. 2015.

[21] Y. Zhou, J. Y. Baek, D. Li, and C. J. Spanos, "Optimal training and efficient model selection for parameterized large margin learning," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Cham, Switzerland: Springer, 2016, pp. 52–64.

[22] D. Li, Y. Zhou, G. Hu, and C. J. Spanos, "Fault detection and diagnosis for building cooling system with a tree-structured learning method," *Energy Buildings*, vol. 127, pp. 540–551, Sep. 2016.
[23] D. Li, Y. Zhou, G. Hu, and C. J. Spanos, "Fusing system configuration information for building cooling plant fault detection and severity level identification," in *Proc. IEEE Int. Conf. Automat. Sci. Eng. (CASE)*, Aug. 2016, pp. 1319–1325.
[24] B. Hariharan, S. V. N. Vishwanathan, and M. Varma, "Efficient max-margin multi-label classification with applications to zero-shot learning," *Mach. Learn.*, vol. 88, nos. 1–2, pp. 127–155, 2012.
[25] Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid, "Label-embedding for attribute-based classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 819–826.
[26] B. Romera-Paredes and P. H. S. Torr, "An embarrassingly simple approach to zero-shot learning," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2152–2161.
[27] C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 453–465, Mar. 2014.
[28] Y. Zhou, R. Arghandeh, and C. Spanos, "Distribution network event detection with ensembles of bundle classifiers," in *Proc. IEEE PES Gen. Meeting*, 2016, pp. 1–5.
[29] D. J. Hill, B. S. Minsker, and E. Amir, "Real-time bayesian anomaly detection for environmental sensor data," in *Proc. Congr.-Int. Assoc. Hydraulic Res.*, vol. 32, 2007, p. 503.
[30] J. P. Liu, O. F. Beyca, P. K. Rao, Z. J. Kong, and S. T. Bukkapatnam, "Dirichlet process Gaussian mixture models for real-time monitoring and their application to chemical mechanical planarization," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 1, pp. 208–221, Jan. 2017.
[31] K. Yoshida, M. Inui, T. Yairi, K. Machida, M. Shioya, and Y. Masukawa, "Identification of causal variables for building energy fault detection by semi-supervised LDA and decision boundary analysis," in *Proc. IEEE Int. Conf. Data Mining Workshop (ICDMW)*, Dec. 2008, pp. 164–173.
[32] M. Jin, R. Jia, Z. Kang, I. C. Konstantakopoulos, and C. J. Spanos, "PresenceSense: Zero-training algorithm for individual presence detection based on power monitoring," in *Proc. BuildSys*, Memphis, TN, USA, Nov. 2014, pp. 1–10.
[33] Y. Zhou, R. Arghandeh, I. Konstantakopoulos, S. Abdullah, and C. J. Spanos, "Data-driven event detection with partial knowledge: A hidden structure semi-supervised learning method," in *Proc. IEEE Amer. Control Conf. (ACC)*, Jul. 2016, pp. 5962–5968.
[34] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Partial knowledge data-driven event detection for power distribution networks," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5152–5162, Aug. 2018.
[35] F. Xu, J. Tan, X. Wang, V. Puig, B. Liang, and B. Yuan, "Mixed active/passive robust fault detection and isolation using set-theoretic unknown input observers," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 2, pp. 863–871, Apr. 2018.
[36] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[37] H. Khorasgani and G. Biswas, "Structural fault detection and isolation in hybrid systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 4, pp. 1585–1599, Oct. 2018.
[38] J. Schein, S. T. Bushby, N. S. Castro, and J. M. House, "Results from field testing of air handling unit and variable air volume box diagnostic tools," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. NISTIR 7365, 2003.
[39] J. Wen and S. Li, *Tools for Evaluating Fault Detection and Diagnostic Methods for air-Handling Units*. Philadelphia, PA, USA: Drexel Univ., 2011.
[40] C. Tong and X. Yan, "A novel decentralized process monitoring scheme using a modified multiblock PCA algorithm," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 1129–1138, Apr. 2007.
[41] T. J. Rato, J. Blue, J. Pinaton, and M. S. Reis, "Translation-invariant multiscale energy-based PCA for monitoring batch processes in semiconductor manufacturing," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 894–904, Apr. 2017.
[42] J. Schein, S. T. Bushby, N. S. Castro, and J. M. House, "A rule-based fault detection method for air handling units," *Energy Buildings*, vol. 38, no. 12, pp. 1485–1492, 2006.
[43] J. Schein and S. T. Bushby, "A hierarchical rule-based fault detection and diagnostic method for HVAC systems," *HVAC R Res.*, vol. 12, no. 1, pp. 111–125, 2006.
[44] D. Li, Y. Zhou, G. Hu, and C. J. Spanos, "Optimal sensor configuration and feature selection for AHU fault detection and diagnosis," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1369–1380, Jun. 2017.
[45] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2002.
[46] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[47] J. Weston, S. Bengio, and N. Usunier, "Large scale image annotation: Learning to rank with joint word-image embeddings," *Mach. Learn.*, vol. 81, no. 1, pp. 21–35, 2010.
[48] Y. Zhao, J. Wen, and S. Wang, "Diagnostic Bayesian networks for diagnosing air handling units faults—Part II: Faults in coils and sensors," *Appl. Therm. Eng.*, vol. 90, pp. 145–157, Nov. 2015.
[49] S. Li and J. Wen, "Application of pattern matching method for detecting faults in air handling unit system," *Autom. Construct.*, vol. 43, pp. 49–58, Jul. 2014.
[50] S. Li and J. Wen, "A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform," *Energy Buildings*, vol. 68, pp. 63–71, Jan. 2014.
[51] S. Li, J. Wen, X. Zhou, and C. J. Klaassen, "Development and validation of a dynamic air handling unit model, Part I," *ASHRAE Trans.*, vol. 116, no. 1, p. 45, 2010.
[52] S. Li, J. Wen, X. Zhou, and C. J. Klaassen, "Development and validation of a dynamic air handling unit model, Part II," *ASHRAE Trans.*, vol. 116, no. 1, pp. 57–74, 2010.
[53] B. Price and T. Smith, "Development and validation of optimal strategies for building HVAC systems," Dept. Mech. Eng., Univ. Iowa, Iowa City, IA, USA, Tech. Rep. ME-TEF-03-001, 2003.
[54] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numer. Math.*, vol. 14, no. 5, pp. 403–420, Apr. 1970.
[55] P. Sreedhara and P. Haves. (2001). Comparison of Chiller Models for use in Modelbased Fault Detection. Energy Systems Laboratory. [Online]. Available: http://esl.tamu.edu and http://hdl.handle.net/1969.1/5118
[56] H. Philip, T. I. Salsbury, and J. A. Wright. (1996). Condition Monitoring in HVAC Subsystems Using First Principles Models. Loughborough Univ. Institutional Repository. [Online]. Available: https://dspace.lboro.ac.uk/2134/5028
[57] H. Wang, Y. Chen, C. W. H. Chan, J. Qin, "A robust fault detection and diagnosis strategy for pressure-independent VAV terminals of real office buildings," *Energy Buildings*, vol. 43, no. 7, pp. 1774–1783, 2011.
[58] H. Wang, Y. Chen, C. W. H. Chan, J. Qin, and J. Wang, "Online model-based fault detection and diagnosis strategy for vav air handling units," *Energy Buildings*, vol. 55, pp. 252–263, Dec. 2012.
[59] S. Katipamula and M. R. Brambley, "Review article: Methods for fault detection, diagnostics, and prognostics for building systems—A review, Part I," *HVAC R Res.*, vol. 11, no. 1, pp. 3–25, 2005.



Dan Li received the B.S. degree in automation engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2012, and the Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2017.

She is currently a Research Fellow with the Institute of Data Science, National University of Singapore, Singapore. Her research interests include statistical learning theory, deep learning, data analytics, anomaly detection, and fault detection and diagnosis with applications to smart buildings and cyberphysical systems.

Yuxun Zhou received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, Shaanxi, China, in 2009, the Diplome d'Ingenieur degree in applied mathematics from Ecole Centrale Paris, Paris, France, in 2012, and the Ph.D. degree from the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA, USA, in 2017.

His research interests include statistical learning theory and paradigms for modern information rich, large scale, and human-involved systems.

Guoqiang Hu (M'08) received the B.Eng. degree in automation from the University of Science and Technology of China, Hefei, China, in 2002, the M.Phil. degree in automation and computer-aided engineering from the Chinese University of Hong Kong, Hong Kong, in 2004, and the Ph.D. degree in mechanical engineering from the University of Florida, Gainesville, FL, USA, in 2007.

From 2008 to 2011, he was an Assistant Professor with Kansas State University, Manhattan, KS, USA. In 2011, he joined the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a Tenured Associate Professor and the Director of the Centre for System Intelligence and Efficiency. His research focuses on analysis, control, design, and optimization of distributed intelligent systems with applications to cooperative robotics and smart city systems.

Dr. Hu was a recipient of the Best Paper in Automation Award in the 14th IEEE International Conference on Information and Automation and the Best Paper Award (Guan Zhao-Zhi Award) in the 36th Chinese Control Conference. He serves as an Associate Editor for the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, a Technical Editor for the IEEE/ASME TRANSACTIONS ON MECHATRONICS, an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, and a Subject Editor for the *International Journal of Robust and Nonlinear Control*.

Costas J. Spanos (M'77–SM'92–F'98) received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, and the M.S. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA.

In 1988, he joined the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley (UC Berkeley), Berkeley, CA, USA, where he was the Chair of EECS, the Associate Dean for Research with the College of Engineering, and the Director of the UC Berkeley Microfabrication Laboratory. He has co-founded two successful companies in USA in the Semiconductor Manufacturing space. He is currently an Andrew S. Grove Distinguished Professor and the Director of the Center for Information Technology Research in the Interest of Society, an Information Technology Research Institute across UC Berkeley, UC Davis, Davis, CA, USA, UC Merced, Merced, CA, USA, and UC Santa Cruz, Santa Cruz, CA, USA. He is also the Founding Director and the CEO of the Berkeley Education Alliance for Research in Singapore and the Lead Investigator of a large research program on smart buildings based in California and Singapore. He has authored or co-authored more than 300 papers and a textbook. He holds 15 patents. He has supervised more than 40 Ph.D. recipients who now hold key positions in industry and academia. His research focuses on sensing, data analytics, modeling and machine learning, with broad applications in semiconductor technologies, and cyberphysical systems.

Dr. Spanos was elected as a Fellow of the IEEE for contributions and leadership in semiconductor manufacturing in 2000.