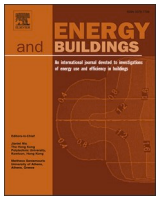
Energy & Buildings 317 (2024) 114378

Contents lists available at ScienceDirect

Energy & Buildings

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



An entropy-based causality framework for cross-level faults diagnosis and isolation in building HVAC systems

Jiajing Huang [a,b], Naghmeh Ghalamsiah [c], Abhidnya Patharkar [a,b], Ojas Pradhan [c], Mengyuan Chu [d], Teresa Wu [a,b,*], Jin Wen [c], Zheng O'Neill [d], Kasim Selcuk Candan [a,e]

[a] School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, USA
[b] ASU-Mayo Center for Innovative Imaging, Arizona State University, Tempe, AZ, USA
[c] Department of Civil, Architectural and Environmental Engineering, Drexel University, Philadelphia, PA, USA
[d] J. Mike Walker'66 Department of Mechanical Engineering, Texas A&M University, College Station, TX, USA
[e] Center for Assured and Scalable Data Engineering, Arizona State University, Tempe AZ, USA

ARTICLE INFO

Keywords:
Causal learning
Information entropy
Bayesian network
Cross-level fault
Root cause fault diagnosis
Building HVAC system

ABSTRACT

Faults, such as malfunctioning sensors, equipment, and control systems, significantly affect a building's performance. Automatic fault detection and diagnosis (AFDD) tools have shown great potential in improving building performances, including both energy efficiency and indoor environment quality. Since modern buildings have integrated systems where multiple subsystems and equipment are coupled, many faults in a building are cross-level faults, i.e., faults occurring in one component that trigger operational abnormalities in other subsystems. Compared with non-cross-level faults, it is more challenging to isolate the root cause of a cross-level faults due to the system coupling effects. Bayesian networks (BNs) have been studied for the root cause isolation for building faults. While promising, existing BN-based diagnosis methods highly rely on expert domain knowledge, which is time-consuming and labor expensive, especially for cross-level faults. To address this challenge, we propose an entropy-based causality learning framework, termed Eigen-Entropy Causal Learning (EECL), to learn BN structures. The proposed method is data-driven without the use of expert domain knowledge; it utilizes causal inference to determine the causal mechanisms between faults status and symptoms to construct the BN model. To demonstrate the effectiveness of the proposed framework, three fault test cases are used for evaluation in this study. Experimental results show that the BN constructed by the proposed framework is able to conduct building cross-level faults diagnosis with a comparable isolation accuracy to those by domain knowledge while maintaining less complexed BN structure.

1. Introduction

Building Heating, Ventilation and Air conditioning (HVAC) systems are complex with a variety of sensors, subsystems and automatically controlled components. According to the United Nations Environment Programme, approximately 135 EJ operational energy demand and 10 Gt energy-related carbon dioxide emission were attributed to the building systems in 2021 [1]. It is also reported that 30 % of this energy usage [2] was wasted due to malfunctioning sensors and components in the HVAC systems [3,4]. Automatic fault detection and diagnosis (AFDD) technologies thus are vital to ensure satisfactory building performances, especially from the aspect of energy efficiency [5]. Field studies and practices indicate that AFDD technologies cannot only achieve up to 20 % building energy savings [6,7] but also improve equipment lifecycles and indoor comforts [8–10].

Modern building HVAC systems typically include a set of multiple, highly coupled subsystems such as cooling/heating plant, primary air distribution, and terminal air distribution subsystems. Due to the coupling effect among building components, a fault occurring in one equipment or subsystem may propagate and influence other equipment or subsystems [11,12]. Hence, component-level AFDD methods may not be efficient and suitable solutions to the root cause analysis for cross-level faults, i.e., faults causing adverse effects across multiple components and subsystems [13]. Chen et al. [13] has provided an example of a chiller supply water temperature sensor bias fault (e.g., sensor reading higher than actual temperature) in the chiller plant which would cause the cooling valve open position in a downstream air handling unit

* Corresponding author at: School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, USA.
E-mail address: teresa.wu@asu.edu (T. Wu).

https://doi.org/10.1016/j.enbuild.2024.114378
Received 22 February 2024; Received in revised form 17 May 2024; Accepted 2 June 2024
Available online 3 June 2024

Nomenclature

AFDD	Automatic fault detection and diagnosis
AHU	Air handling unit
AIA	Average isolation accuracy
BAS	Building automation system
BN	Bayesian network
CL	Causal learning
EE	Eigen-Entropy
EECL	Eigen-Entropy-based causality learning
HVAC	Heating, ventilation and air conditioning
IA	Isolation accuracy
PN	Probability of necessity
SIA	Sufficient isolation accuracy
VAV	Variable air volume

(AHU) to be lower than normal. In this case, a component-level AFDD tool that only monitors the AHU might result in false alarms such as a cooling coil valve fault or a supply air temperature sensor fault. Hence, a root cause analysis is necessary to ensure the correct diagnosis of cross-level faults.

Compared with fault detection studies, much fewer fault diagnosis/root cause analysis studies exist [14]. A root cause analysis has shown its importance to improve quality assurance, reliability and performance in many fields, such as power systems [15], manufacturing [16], just to name a few. Bayesian networks (BNs) have been extensively studied as a root cause analysis technique. For example, Wang et al. [15] conducted an analysis on root causes of occurring alarms in thermal power plants based on posterior probability from BN. In their research, the BN was constructed by one child node and multiple parent nodes that describe the relationship between an alarm variable and root-cause variables using the process knowledge. Lokrantz et al. [16] proposed a BN-based graphic probabilistic models using the expert knowledge to identify the causality of failure and quality deviation among multiple manufacturing stages, where network parameters were trained by historical data, and root cause was inferred according to defect types and measurements. Liu et al. [17] developed a strong relevant mechanism BN combining process mechanism analysis with historical data mining for unmonitored root cause variables in chemical plants fault diagnosis, which showed great practicability and satisfactory performances in fault propagation recognitions. Amin et al. [18] presented a hybrid data-driven method integrating principal component analysis with the Bayesian networks for fault detection and diagnosis in process plants, which demonstrated a strong efficacy of diagnosis performance while maintaining lesser false diagnosis. In [19], the same authors developed a dynamic Bayesian network-based fault detection and root cause diagnosis, which had an ability to convert the continuous process data into meaningful evidence instead of a probabilistic domain. There are also some BN studies emphasizing cross-level fault diagnosis in an HVAC system from the data-driven perspective. For example, Wang et al. [20] introduced a practical, efficient discretized BN-based diagnosis method for chiller faults; Chen et al. [21] proposed a whole building fault diagnosis method based on Discrete BN to isolate faults causing significant abnormalities in multiple subsystems/equipment during system operation, and further designed a weather and schedule-based pattern matching Discrete BN to diagnose cross-level faults in building HVAC systems for real-time fault diagnosis and isolations [13]. Wang et al. [22] combined a reference model-based approach using normal data with a BN-based approach using faulty data to improve the effectiveness of fault diagnosis. Taal and Itard [23] proposed an automated fault identification (AFI) process for HVAC systems using a diagnostic BN, successfully isolating control faults in a thermal energy plant. Pradhan et al. [24] developed a dynamic BN-based approach that incorporated the temporal dependencies of fault nodes between time steps using temporal conditional probabilities to improve accuracy for a whole building level fault diagnosis.

The aforementioned BN-based diagnosis methods highly rely on heuristics processes to learn causal relationships among fault status and symptoms, and their causal mechanism is primarily determined by the expert knowledge. While promising, heuristics processes by the domain knowledge may not be adequate and effective for the fault diagnosis in more complex buildings, especially those with multiple coupled subsystems. Other than being labor intensive, these approaches may not discover underlying coupling effects among the subsystems comprehensively. This motivates us to take a data-driven approach to interrogate the fault-symptom causalities to construct the BN structure. A notable emerging field is causal learning (CL) which uses the observational data to learn causality and we believe CL presents new opportunities to address domain specific challenges [25]. In general, CL research focuses on two categories [26]: (1) causal effects estimation; and (2) causal structure learning. Causal effects estimation is to investigate how much changing one variable will influence another given a causal structure assumption between these two variables. This can be done by the counterfactual inference [27,28] which assesses the strength of causality between two events by inferring the likelihood of one event not occurring when another is absent. Causal structure learning, on the other hand, is to induce the structure describing the causal relationships from variables to others, and BN is one of the prevailing causal structure learning tools as it has shown the ability to represent the probabilistically conditional independence in a graph model, providing an efficient and expressive way for knowledge representations and acquirements [29,30].

Despite the extensive research conducted on BN, there have been limited investigations into the causal structure construction from the causal effect estimation perspective using observation data, especially the BN construction for building fault diagnosis. As reviewed earlier, most building research using BN has heavily relied on the domain experts' knowledge. Additionally, the existing research on BN-based fault diagnosis primarily relied on the assumption of the symptom independence [13,31,32], which may not always be valid. In this research, we hypothesize that the building is an interconnected system comprising multiple subsystems, and when a fault occurs, co-evolving of multiple symptoms may present some unique patterns. Motivated by this idea, we introduce a new concept termed "synchronicity" to describe the co-evolving patterns, and propose a CL-based framework to induce the BN structures. Specifically, Eigen-Entropy (EE) [33], a metric derived from multi-variate time series is employed to characterize the "synchronicity". Next, causal inference is used for causation measurements between the fault status and symptoms to decide what symptoms should be included in the BN structure model. Finally, the performance of our proposed framework is validated by three fault test cases.

The paper is organized as follows. Background and methodology are detailed in sections 2 and 3. Experiments are summarized in section 4, followed by discussion in section 5. Finally, conclusion and future work is drawn in section 6.

2. Background

In this section, we review BN model for building fault diagnosis, and introduce the Pearl Causality, a commonly used causal effect estimation approach that can support the BN structure construction.

2.1. BN model for building system fault diagnosis

Bayesian Network (BN) is a probabilistic graphical model representing a set of variables and their conditional dependencies via a directed acyclic graph, which can be used to reveal causal relationships between faults and symptoms. Fig. 1. below illustrates an example of a BN model for fault diagnosis in the building systems.

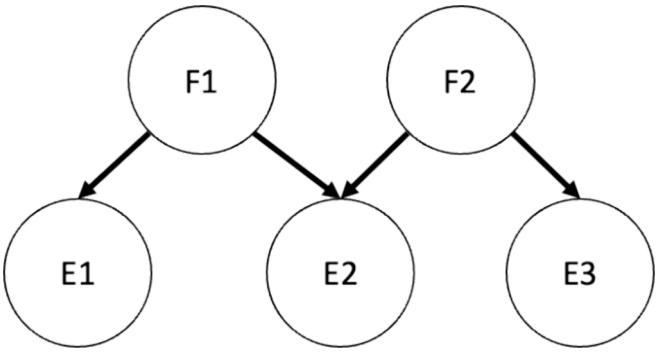


Fig. 1. A BN model for fault diagnosis.

In this BN model, F1 and F2 are fault nodes that represent two distinct faults, while E1, E2 and E3 are evidence nodes that are indicatives (symptoms) of the presence of a fault. Arcs from the fault nodes to the evidence nodes demonstrate the direct causation that a fault has on the occurrence of the evidence. By Bayes theorem [34], posterior probabilities of faults (F1 and F2) given these evidence nodes (E1, E2, and E3) can be calculated to infer which fault is more likely to affect the system. For instance, if the posterior probability of F1 is greater than that of F2, the fault is said to be F1.

Clearly, the link between faults and evidence nodes plays a crucial role in facilitating fault diagnosis in building systems based on BN. Pearl Causality, a method for estimating causal effects in fault-symptom relationships, is often employed to support the inference of BN structure. We provide a review of Pearl Causality basic in the next section.

2.2. Pearl causality

Pearl Causality [27], also known as counterfactual inference, assesses the likelihood that one event is the cause of another, which is usually evaluated by the probability of necessity. Given two binary-valued events, X and Y, let x and y stand for the propositions $X = 1$ and $Y = 1$, respectively, x' and y' stand for their complements ($X = 0$ and $Y = 0$). The probability of necessity (PN) is then defined as:

$$PN = P(Y_{X=0} = 0 | X = 1, Y = 1) = P(y'_{x'} | x, y) \tag{1}$$

Consequently, PN stands for the probability that event y would not have occurred in the absence of event x ($y'_{x'}$), given that x and y did actually occur.

Supposing the frequencies of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are as shown in Table 1, the PN can be calculated as [35]:

$$PN = P(y'_{x'} | x, y) = \frac{P(y) - P(y|x')}{P(x, y)} = \frac{\frac{n_{11}+n_{01}}{n} - \frac{n_{01}}{n_{01}+n_{00}}}{\frac{n_{11}}{n}} \tag{2}$$

where $n = n_{11} + n_{10} + n_{01} + n_{00}$. When $PN \geq 0.5$, the causal relationship from x to y is confirmed [36].

In this study, event $\boldsymbol{X}$ serves as an indication of the fault status in the building systems. Specifically, when $\boldsymbol{X} = 1$, it signifies the occurrence of a fault, indicating fault conditions. On the other hand, when $\boldsymbol{X} = 0$, it represents normal operations, indicating fault-free conditions. Additionally, event $\boldsymbol{Y} = 1$ is another indicator that describes specific properties related to multiple symptoms, referred to as synchronicity, and $\boldsymbol{Y} = 0$ is referred to as the complement to synchronicity (asynchronicity).

Table 1
Frequency data of events X and Y.

	$X = 1$	$X = 0$
$Y = 1$	n_{11}	n_{01}
$Y = 0$	n_{10}	n_{00}

To illustrate this idea, we consider AHU Cooling Coil Valve Stuck Fully Open (CoolCoiValStuck_0) as the example fault with two evidence nodes, AHU Cooling Coil Discharge Air Temperature (CC-DA-TEMP) and AHU Cooling Coil Valve Control Signal (CC-VLV). Therefore, $\boldsymbol{X} = 1$ indicates the occurrence of CoolCoiValStuck_0, $\boldsymbol{X} = 0$ indicates the nonoccurrence of CoolCoiValStuck_0 (fault-free condition); $\boldsymbol{Y} = 1$ indicates the synchronicity exist between CC-DA-TEMP and CC-VLV, and $\boldsymbol{Y} = 0$ is asynchronicity. Suppose$n_{11} = 16$ (the frequency of synchronicity under this fault) and$n_{01} = 14$ (the frequency of asynchronicity under this fault); $n_{10} = 484$ (the frequency of synchronicity under fault-free condition) and $n_{00} = 986$ (the frequency of asynchronicity under fault-free condition). Given these values, we obtain PN = 0.54 by Eq (2), which is greater than 0.5. We conclude that the synchronicity between CC-DA-TEMP and CC-VLV is attributed to CoolCoiValStuck_0.

Consequently, in the methodology section, we will provide a comprehensive explanation of this term and elucidate how we employ Pearl Causality to ascertain the BN structure. This BN structure will then be utilized to facilitate fault diagnosis in the building systems.

3. Methodology

To begin with, we introduce a new concept, “synchronicity”, to describe interactions among multiple time series in general. Next, we introduce details about our proposed entropy-based causal learning framework, termed Eigen-Entropy-based Causality Learning (EECL), for the BN construction to support fault diagnosis in building systems.

3.1. Information entropy and time series synchronicity

Information entropy [37] is to quantify the averaged information inherent to a random variable. Given a single variate with N possible values, its information entropy (H) is defined as

$$H = -\sum_{i=1}^{N} p_i \log p_i \tag{3}$$

where p_i is the probability of this single variate taking the value i, and $\sum_{i=1}^{N} p_i = 1$.

Eigen-Entropy (EE) is an entropy for multivariate data derived on eigenvalues extracted from the correlation magnitude matrix [33]. Given a dataset X with n samples and m features, EE is defined as

$$EE = -\sum_{i=1}^{m} \frac{\lambda_i}{m} \log \frac{\lambda_i}{m} \tag{4}$$

where λ_i is the eigenvalue corresponding to the correlation magnitude matrix on the feature space.

The EE method has been used to quantify the homogeneity (or heterogeneity) of a tabular dataset [33] to support sampling decisions. One use case of EE is to construct baseline for AFDD as demonstrated in [38,39]. The interest of this research is on multiple time-series, and here we introduce "synchronicity" to describe the multi-time series dataset property which can be measured by EE.

Definition 1. Multiple time series collected from the system may present some co-evolving patterns. If the trend of movements aligns over time (that is, the time series may increase, decrease or remain constant together with respect to time), we define this pattern as synchronicity.

Without loss of generality, let us consider two time series, X_1 and X_2, where $X_1 = [x_{11}, x_{12}, \cdots, x_{1n}]$, $X_2 = [x_{21}, x_{22}, \cdots, x_{2n}]$, and x_{ij} refers the time point j of time series i. At time t, cosine similarity between X_{1t} and X_{2t}, is defined as:

$$\cos < X_{1t}, X_{2t} >= \frac{\sqrt{\sum_{i=1}^{t} x_{1i} x_{2i}}}{\sqrt{\sum_{i=1}^{t} x_{1i}^2} \sqrt{\sum_{i=1}^{t} x_{2i}^2}} \tag{5}$$

The corresponding cosine similarity magnitude matrix on X_{1t} and X_{2t} is

$$\mathbf{C}_t^* = \begin{pmatrix} 1 & c_{12}^t \\ c_{21}^t & 1 \end{pmatrix} \tag{6}$$

where c_{12}^t is the magnitude of cosine similarity between X_{1t} and X_{2t}, ($c_{12}^t = |\cos < X_{1t}, X_{2t} > |$) and $c_{21}^t = c_{12}^t$. We derive eigenvalues λ_1^t and λ_2^t from $\mathbf{C}_t^*$ to obtain EE (see Eq (4)). Please note for multiple time-series, the dimension of the cosine similarity magnitude matrix increases accordingly.

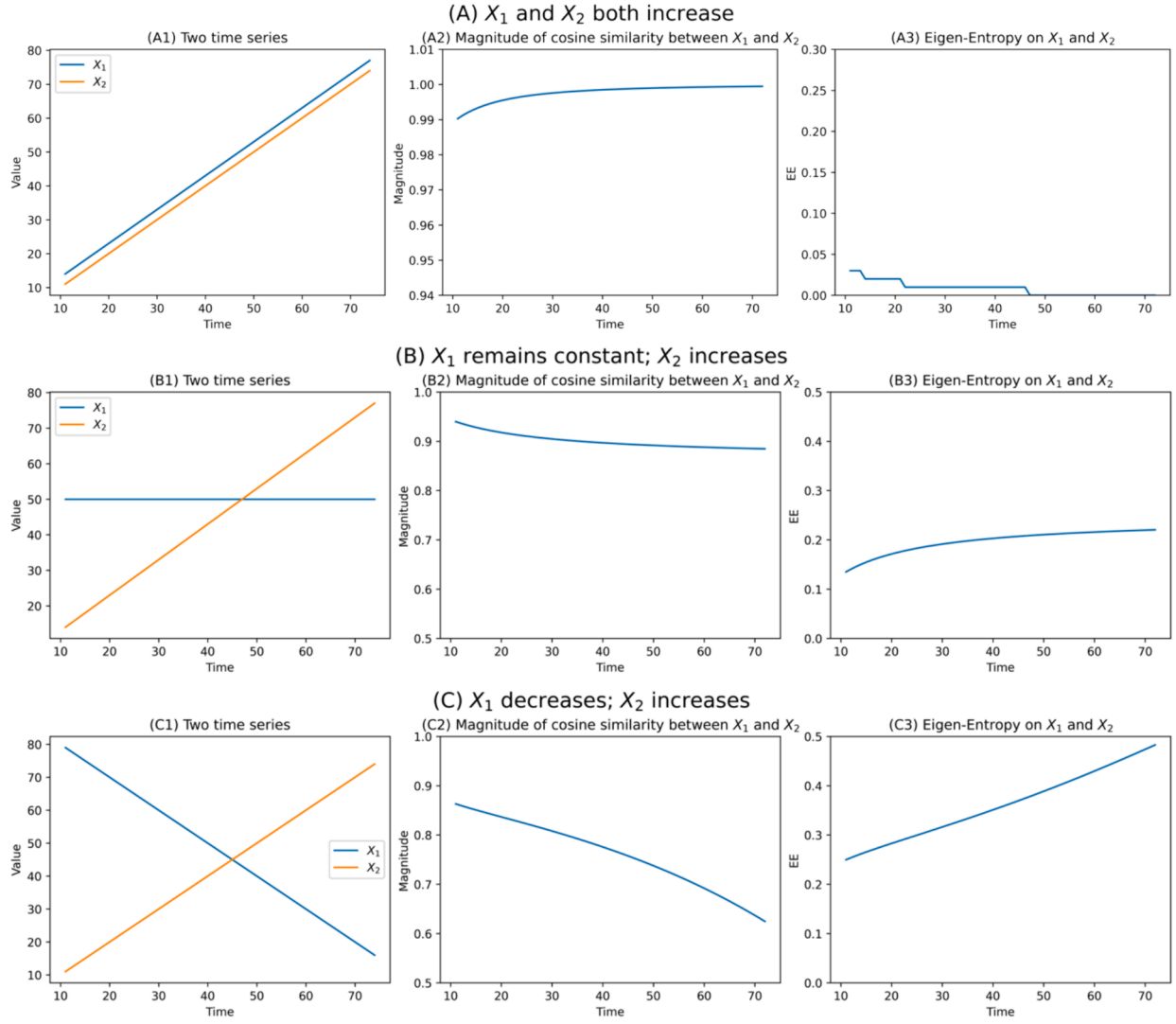


Fig. 2. Trend of movement, cosine similarity and EE between two time series X_1 and X_2 over time when (A) X_1 and X_2 both increase; (B) X_1 decreases and X_2 increases; (C) X_1 remains constant and X_2 increases. As shown in (A1), two time-series show the exactly same trend of movements (perfectly aligned movement), the corresponding magnitude of cosine similarity increases (A2) and EE decreases over the time (A3); As shown in (B1), two time-series show the different trend of movements, the corresponding magnitude of cosine similarity becomes decreasing (B2) and EE becomes increasing over the time (B3); As shown in (C1), two time-series show the exactly opposite trend of movements, the corresponding magnitude of cosine similarity becomes decreasing significantly (C2) and EE becomes increasing drastically over the time (C3). This indicates that EE can measure the degree of aligned movements between two time-series over the time.

Given the magnitude of cosine similarity between X_{1t} and X_{2t}, we can further obtain corresponding Eigen-Entropy by Eq (4). Let EE_{t1} be the Eigen-Entropy calculated for X_1 and X_2 at time point t_1 (X_{1t_1} and X_{2t_1}) and EE_{t2} is the Eigen-Entropy calculated for X_1 and X_2 at time point t_2 (X_{1t_2} and X_{2t_2}), where $t_1 < t_2$. If $EE_{t2} < EE_{t1}$, we assume that the trends of the time series have become more synchronous or aligning between time points t_1 and t_2, and vice versa.

Therefore, the value of EE indicates the degree of alignment between the movements of these time-series, or their synchronicity. If the time-series movements are well aligned or highly positively correlated, the value of EE would be zero or close to zero. As the movements become less aligned, the value of EE increases, reaching its maximum when the movements are completely misaligned or highly negatively correlated.

To illustrate this concept, we present a simple example in Fig. 2, which shows the relationships among the movement trends, cosine similarities, and EEs for these two time-series under different scenarios: (1) both X_1 and X_2 increase over time; (2) X_1 remains constant while X_2 increases over time; and (3) X_1 decreases while X_2 increases over time. In Fig. 2 (A), where X_1 and X_2 exhibit perfectly aligned movements over time (scenario (1)), indicating a strong positive correlation, the cosine similarity between X_1 and X_2 increases, and the corresponding EE decreases. In Fig. 2 (B), as X_1 and X_2 show divergent movements (scenario (2)), the cosine similarity between X_1 and X_2 decreases, and the corresponding EE increases. In Fig. 2 (C), X_1 and X_2 exhibit movements in completely opposite directions (scenario (3)), depicting a strong negative correlation, the cosine similarity decreases significantly, and the EE increases markedly over time.

Note that the patterns for cases where both X_1 and X_2 decrease, X_1 increases while X_2 decreases, and X_1 remains constant while X_2 decreases, are similar to Fig. 2 (A), (B), and (C) respectively. Therefore, we conclude that EE can be used as a metric to measure the synchronicity, describing the phenomenon of multiple time series showing trends of aligned movements over time.

3.2. Eigen-Entropy-based causality learning (EECL) for HVAC AFDD

In this research, we focus on the building HVAC fault diagnosis/root cause isolation. This involves the use of evidence nodes, which are comprised of sensor readings obtained from the building system over a specific time frame. These evidence nodes can be regarded as multiple time series. Upon analyzing the data, we have observed that when a system contains fault(s), the evidence nodes exhibit synchronicity. This has led to our assumption that the synchronicity among evidence nodes is attributed to the system fault(s). As a result, we have employed this causal assumption to identify evidence nodes for constructing BNs. That is, we need to decide which evidence nodes are associated with which fault.

Given k fault nodes (a.k.a., fault test cases), each fault node including one fault dataset (i.e., data from the system that contain fault(s)) and corresponding baseline dataset, each dataset consisting of d days' data, each day's data with t time points (t samples), and m evidence nodes (m symptoms), **Algorithm 1** presents the EECL method to determine the evidence nodes for each fault node so as to construct a BN for fault diagnosis. These include two parts: initialization and evidence node selection stages.

At the initialization stage, we need to identify a set of critical evidence nodes for each fault node as candidates to support the BN construction. This involves four main steps.

- We obtain feature importance score of each evidence node by training a machine learning model (e.g., random forest classifier) using all k fault datasets, and then select those evidence nodes whose importance scores are greater than a set value through sensitivity analysis (e.g., in this study, 0.05). This forms a set of critical evidence nodes that can differentiate all k fault datasets, say $\mathbf{E}_{all}$ (line 1).
- We follow a similar procedure to obtain multiple sets of evidence nodes, $\mathbf{E}_{i,j}$'s ($i \neq j$), each set containing critical evidence nodes that can distinguish its fault dataset from any other fault datasets for any individual test case i (e.g., F_i vs. F_j) (lines 2–3). We need to identify a set of critical evidence nodes that can differentiate between its fault dataset and baseline, saying $\mathbf{E}_i$ (line 4). Hence, we form a set of critical evidence nodes for the test case by taking the union of $\mathbf{E}_{all}$, multiple $\mathbf{E}_{i,j}$'s and $\mathbf{E}_i$, saying $\mathbf{E}_i'$, which contains critical evidence nodes from previous steps.
- Next is to assign a score to each evidence node in $\mathbf{E}_i'$ by taking the maximum value of its importance among $\mathbf{E}_{all}$, $\mathbf{E}_{i,j}$'s and $\mathbf{E}_i$ (lines 5–6).
- Finally, we rank all evidence nodes in $\mathbf{E}_i'$ in a descending order by their importance score to obtain a set of ranked, critical evidence nodes, $\widehat{\mathbf{E}}_i$ (line 7).

Fig. 3 illustrate an example given the scenario of three fault node, F_1, F_2 and F_3, one baseline node, B, and four evidence nodes, E_1, E_2, E_3, and E_4. Using a machine learning model (e.g., random forest), the important scores for E_1, E_2, E_3, and E_4 that can distinguish three fault nodes (F_1 vs. F_2 vs. F_3) are all 0.2; Next let us focus on the F_1, and assess these evidence nodes by distinguishing pairwise fault nodes (F_1 vs. F_2; F_1 vs. F_3): the scores for E_1, E_2, E_3, and E_4 are 0.15, 0.25, 0.20, and 0.05 for F_1 vs. F_2, and 0.23, 0.24, 0.36, 0.17 for F_1 vs. F_3 respectively; we obtain the scores for F_1 vs. B are 0.05, 0.62, 0.00 and 0.33. Therefore, from these scoring results, we rank evidence nodes according to their max score in a

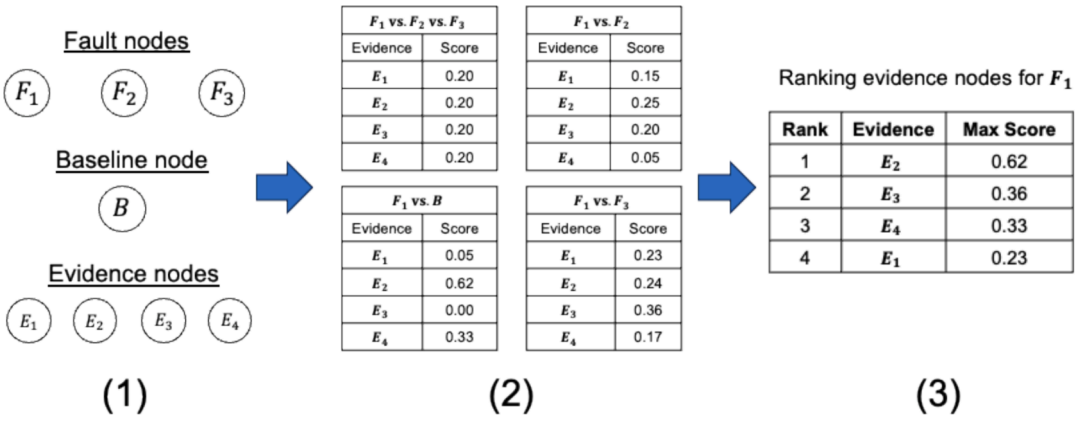


Fig. 3. An illustrative example of procedures of obtaining the set of ranked critical evidence nodes for one fault node. (1) 3 fault nodes and one baseline with 4 evidence nodes; (2) For one fault node, F_1, obtain (importance) scores for each evidence node under different comparison scenarios: (3) Rank evidence nodes by max score for F_1.

descending order (scores: $E_2 > E_3 > E_4 > E_1$).

Algorithm 1: BN construction by EECL for fault diagnosis

Input: k fault nodes, each fault node with one fault and one baseline datasets, each dataset consisting of d days' data, each data with m evidence nodes and t time points
Output: Bayesian networks for fault diagnosis,*BN*
Initialization:
1: Obtain a set of critical evidence nodes that can differentiate all k fault datasets, $\mathbf{E}_{all}$, according feature importance scores from the machine learning model
2: For fault node i, $i = 1, \cdots, k$
3: Repeat Step 1 to obtain multiple sets, each set containing critical evidence nodes that can differentiate its fault dataset from another fault dataset of fault node j, $\mathbf{E}_{i,j}, j = 1, \cdots, k, i \neq j$
4: Repeat Step 1 to obtain a set of critical evidence nodes that can differentiate its fault dataset baseline,$\mathbf{E}_i$
5: Obtain a set of critical evidence nodes by taking the union of $\mathbf{E}_{all}$, $\mathbf{E}_{i,j}$'s and $\mathbf{E}_i$, $\mathbf{E}_i'$
6: Assign importance score to each evidence node in $\mathbf{E}_i'$ by taking its maximum value of importance scores among $\mathbf{E}_{all}$, $\mathbf{E}_{i,j}$'s and $\mathbf{E}_i$
7: Rank each evidence node in $\mathbf{E}_i'$ by its importance score in a descending order to obtain a ranked evidence node set,$\widehat{\mathbf{E}}_i = \{e_{(1)}, e_{(2)}, \cdots\}$
Evidence nodes selection:
8: For fault node i, $i = 1, \cdots, k$
9: For day p, $p = 1, \cdots, d$
10: Normalize each evidence node from $\widehat{\mathbf{E}}_i$ of fault data with respect to its baseline data in day p
11: Calculate Eigen-Entropy (EE) by Eq (4) on $\widehat{\mathbf{E}}_i$ for each q time point on fault data in day p using the first qth samples,$EE_q, q = 2, \cdots, t$
12: Derive normalized EE for each q time point, NEE_q, where $NEE_q = EE_q/q$
13: Assign 1 for each q time point if $NEE_q < \varepsilon$; 0 otherwise
14: Repeat Steps 11–13 for baseline data in day p
15: Obtain frequency table given results from Steps 11–14, and calculate *PN* by the Eq (2)
16: If $PN < 0.5$, update $\widehat{\mathbf{E}}_i$ by removing the last ranked evidence node, and repeat Steps 11–14
17: Otherwise, stop and output $\widehat{\mathbf{E}}_i$
18: Return *BN* by linking k fault nodes with $\{\widehat{\mathbf{E}}_1, \cdots, \widehat{\mathbf{E}}_k\}$

Next, we start selecting evidence nodes for each fault node. The evidence node selection stage also involves four main steps.

- For each day in a fault node, we first normalize each evidence node in $\widehat{\mathbf{E}}_i$ of the fault data with respect to its corresponding baseline (line 10). Then for both fault data and baseline, we calculate EE on $\widehat{\mathbf{E}}_i$ for each time point q and obtain corresponding normalized EE, NEE_q. We assign 1 if NEE_q is smaller than a certain threshold (ε), indicating the existence of synchronicity among evidence nodes at time point q (that is, evidence nodes exhibit the trend of aligned movements over time q); 0 otherwise (lines 11–14).
- After going through all d days' data, we obtain a frequency table for this fault node, where $X = 1$ indicating fault conditions, $X = 0$ indicating fault-free conditions, $Y = 1$ indicating synchronicity, otherwise, $Y = 0$. The frequency information is used for probability of necessity (*PN*) calculations to assess the causal relationship from fault status to synchronicity among $\widehat{\mathbf{E}}_i$ (line 15).
- If the $PN < 0.5$, it indicates the causal relationship does not hold between that fault condition and the synchronicity of the evidence nodes; thus, we remove the last ranked evidence node from $\widehat{\mathbf{E}}_i$ (note that in this approach we do not have any evidence node added) and continue the process; otherwise stop and output the final evidence nodes for this fault node (lines 16–17).
- We go through the procedures for all fault nodes, and finally construct BN by linking all fault nodes to corresponding evidence nodes selected (line 18).

4. Experiments on simulation datasets

4.1. Experimental datasets from simulation

A virtual HVAC system testbed developed using Modelica in Dymola environment [40] is used to generate experimental data in the experiment, and Fig. 4 shows the schematics of the HVAC system. The developed HVAC system model is for a one-floor, five-zone medium-sized office building, which has one Air Handler Unit (AHU) connected with five Variable Air Volume (VAV) terminal boxes serving five zones (four

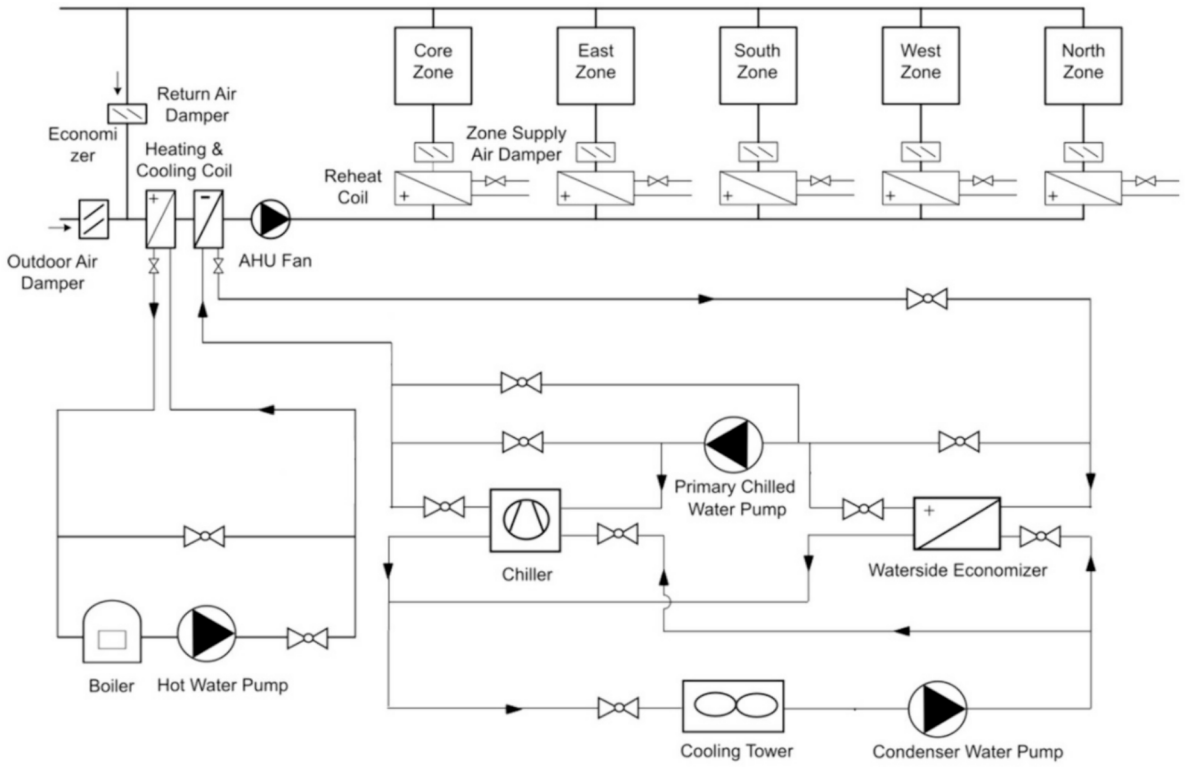


Fig. 4. Schematic diagram of the simulated HVAC system.

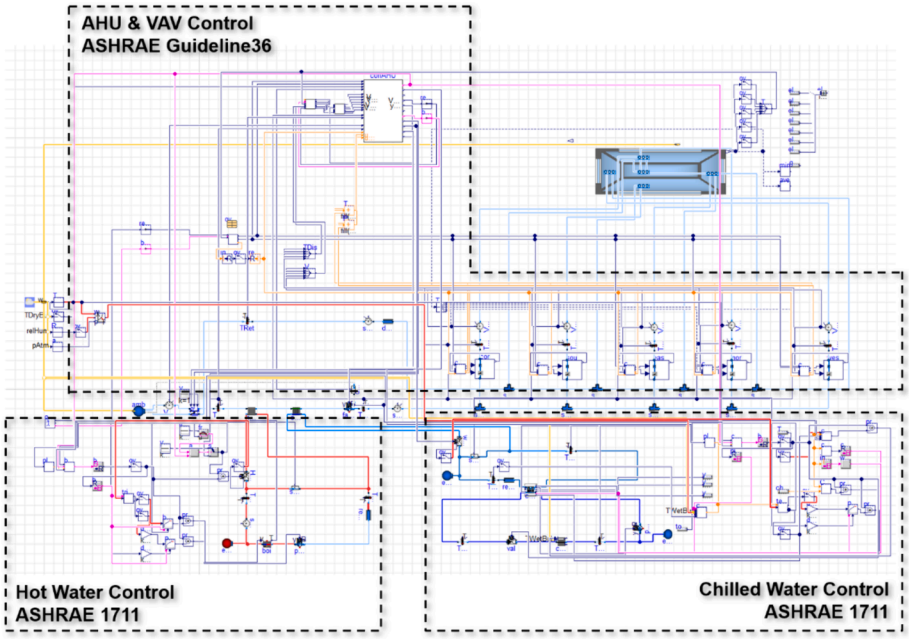


Fig. 5. Modelica implementation of the studied HVAC system for a commercial building.

Table 2
Comparison between the EnergyPlus prototype with the Modelica model.

Item	EnergyPlus Midfloor_Plenum	Modelica
Area [m^2]	1660.7	1662.7
AHU Fan Design Flow Rate [m^3/s]	4.2	4.8
AHU Fan Head [Pa]	1389	1381
Overall Efficiency	0.6	0.49
AHU Fan Rated Electric Power [W]	9685	13,530
Cooling Coil Capacity [W]	95,438	100,711
Heating Coil Capacity [W]	34,995	40,526
Chilled Water Pump Head [W]	Not applicable	255,000
Chilled Water Pump Flow [m^3/s]		0.004
Cooling Water Pump Head [W]		215,700
Cooling Water Pump Flow [m^3/s]		0.0043
Hot Water Pump Head [W]		157,000
Hot Water Pump Flow [m^3/s]		0.00132
Cooling Tower Fan Power [W]		4300

exterior zones, and one interior zone, respectively). Heating and cooling are delivered by a single-duct VAV system and the reheat in the VAV terminals is supplied by electric resistance coils. The chilled water is supplied by a central chiller plant which consists of a chiller, a waterside economizer, a cooling tower, and one chilled water pump and one condenser water pump. A boiler, fed by natural gas, supplies the hot water to the AHU heating coil.

Fig. 5 presents the Modelica model for the studied HVAC system, which was developed based on the open-source Modelica Buildings Library (MBL) [41] and verified against a medium-sized office DOE prototype model [42] developed by Pacific Northwest National Laboratory in EnergyPlus [43]. Table 2 shows a peak load and sizing comparison between the EnergyPlus prototype medium office model with the Modelica model. The peak cooling load (cooling coil capacity) is similar between these two models. The system model consists of three components, namely an HVAC system, a building envelope model, and a model for air flow through building leakage and through open doors based on wind pressure and flow imbalance of the HVAC system. The HVAC system is sized for Chicago, IL, USA in climate zone 5A. The HVAC system control complies with American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) standards and guidelines or literature-reported practices. For example, the air-side control sequences follow ASHRAE Guideline 36 [44] and the water-side control sequences follow ASHRAE project RP-1711 [45]. More details of this HVAC system model can be found in [46–49].

Table 3
Description of three fault nodes.

Fault No.	Fault Node Name	Fault Node Description
1	CoolCoiValStuck_0	AHU Cooling Coil Valve Stuck Fully Open
2	OADamStuck_100	AHU Outdoor Air Damper Stuck Fully Closed
3	SupDucLea_20	Supply Duct Leakage at a degradation rate of 20 %

In this study, three fault datasets and one fault-free dataset collected from this virtual testbed are used. Specifically, these three fault datasets are those collected when the virtual testbed is operated under one of the three different commonly-occurring physical fault conditions, namely, AHU Cooling Coil Valve Stuck Fully Open (CooCoiValStuck_0), AHU Outdoor Air Damper Stuck Fully Closed (OADamStuck_100) and Supply Duct Leakage at a degradation rate of 20 % (SupDucLea_20), respectively (see Table 3), while the fault-free dataset is collected when the virtual testbed is operated under normal conditions. Specifically, the fault-free dataset serves as the baseline for each fault node; in other words, the three fault nodes share the same baseline dataset. Since the HVAC system is sized for Chicago, IL, USA in climate zone 5A and the fault injection period starts at the beginning of the day on July 9 and continues for four weeks until August 5, the ranges of temperature and relative humidity are 24–29 °C and 50 % to 70 %. Each dataset (both fault and fault-free) consists of 28-days' data, each day's data containing 120 time points (samples) for the entire occupied hours (unoccupied hours excluded); consequently, there are 3,360 samples in total for each fault node. In our experiment, 15 days' data (1,800 samples) from each fault node are used as the training dataset while 5 days' data (6,00 samples) are used as the test datasets for validation, and the remaining 8 days are excluded because these days correspond to the periods when the building was unoccupied. Detailed information about the training and testing days can be found in Appendix A.

It is worth noticing that both fault and fault-free datasets have 132 evidence nodes. Out of these 132 evidence nodes, 50 are related to the AHU or chiller. Since building faults usually occur in AHUs or chillers, these evidence nodes are important and thus considered as the candidates for the BN construction. The descriptions of these evidence nodes

Table 4
AHU/Chiller-related evidence nodes.

Evidence No.	Evidence Node Name	Evidence Node Description
E1	SA-TEMP	AHU Supply Air Temperature
E2	SA-TEMP-STP	AHU Supply Air Temperature Set Point
E3	OA-DB-TEMP	AHU Outdoor Air Dry Bulb Temperature
E4	OA-WB-TEMP	AHU Outdoor Air Wet Bulb Temperature
E5	MA-TEMP	AHU Mixed Air Temperature
E6	RA-TEMP	AHU Return Air Temperature
E7	CC-DA-TEMP	AHU Cooling Coil Discharge Air Temperature
E8	SF-SPD	AHU Supply Air Fan Speed
E9	OA-DMPR	AHU Outdoor Air Damper Control Signal
E10	RA-DMPR	AHU Return Air Damper Control Signal
E11	EA-DMPR	AHU Exhaust Air Damper Control Signal
E12	SA-CFM	AHU Supply Air Mass Flow Rate
E13	OA-CFM	AHU Outdoor Air Mass Flow Rate
E14	RA-CFM	AHU Return Air Mass Flow Rate
E15	EA-CFM	AHU Exhaust Air Mass Flow Rate
E16	CC-VLV	AHU Cooling Coil Valve Control Signal
E17	HC-VLV	AHU Heating Coil Valve Control Signal
E18	SAD-FLOW	AHU Supply Air Duct Static Pressure
E19	SAD-FLOW-STP	AHU Supply Air Duct Static Pressure Set Point
E20	CC-HTR	AHU Cooling Coil Heat Transfer Rate
E21	HC-HTR	AHU Heating Coil Heat Transfer Rate
E22	SF-PWR-CONS	AHU Supply Air Fan Power Consumption
E23	CHWS-TEMP	Chilled Water Loop: Chilled Water Supply Temperature
E24	CHWR-TEMP	Chilled Water Loop: Chilled Water Return Temperature
E25	CWS-TEMP	Condenser Water Loop: Cooling Water Supply Temperature
E26	CWR-TEMP	Condenser Water Loop: Cooling Water Return Temperature
E27	HWS-TEMP	Hot Water Loop: Hot Water Supply Temperature
E28	HWR-TEMP	Hot Water Loop: Hot Water Return Temperature
E29	CHWS-TEMP-STP	Chilled Water Loop: Supply Chilled Water Temperature Set Point
E30	CHW-DIFF-FLOW	Chilled Water Loop: Measured Differential Pressure
E31	CHW-FLOW-STP	Chilled Water Loop: Differential Pressure Setpoint
E32	HWS-TEMP-STP	Hot Water Loop: Supply Hot Water Temperature Set Point
E33	HW-DIFF-FLOW	Hot Water Loop: Measured Differential Pressure
E34	HW-FLOW-STP	Hot Water Loop: Differential Pressure Setpoint
E35	CHW-FLOW-CC	Chilled Water Loop: Chilled Water Flow Rate into the Cooling Coil
E36	CHL-CHW-FLOW	Chiller: Chilled Water Flow Rate of the Chiller
E37	WSE-CHW-FLOW	WSE: Chilled Water Flow Rate of the WSE
E38	CW-FLOW	Condenser Water Loop: Cooling Water Flow Rate
E39	CHL-CW-FLOW	Chiller: Cooling Water Flow Rate of the Chiller
E40	WSE-CW-FLOW	Cooling Water Flow Rate of the Water Side Economizer (WSE)
E41	HW-FLOW-HC	Hot Water Loop: Hot Water Loop Flow Rate into the Heating Coil
E42	BLR-HW-FLOW	Boiler: Boiler Hot Water Flow Rate
E43	HW-FLOW-BYPS	Hot Water Loop: Bypass Hot Water Flow Rate
E44	CHL-PWR	Chiller Power consumption
E45	DIFF-SAT-STP	AHU Supply Air Temperature and Supply Air Temperature Setpoint Difference
E46	DIFF-OAT-MAT	Difference between AHU Outdoor Air Temperature and Mixed Air Temperature
E47	CHW-COOLING	Chilled Water Cooling Capacity
E48	VAV-FLOW-SUM	Summation of VAV Flowrate
E49	MA-TEMP-1	AHU Mixed Air Temperature Curve Fit; MAT = f (OAT, RAT, SAflow, RAflow)
E50	MA-TEMP-2	AHU Mixed Air Temperature Curve Fit; MAT = f (OAT, RAT, SAflow, OAdmpr)

can be found in Table 4; besides, all these 50 evidence nodes are ranked according to its importance scores for each fault node, and the details about the ranking method can be found in Appendix B of this paper.

4.2. Evaluation metrics

In the BN model, we assess every fault sample for each fault node by utilizing posterior probabilities derived from prior and conditional probabilities. Let us define s_{ij} as the i^{th} fault sample from fault node j whose true label is $Y(s_{ij})$. Using BN, we can obtain k posterior probabilities, $P_1(s_{ij})$, $P_2(s_{ij})$, …,$P_k(s_{ij})$, indicating likelihoods of s_{ij} belonging to fault nodes 1, 2, …,k. Thus, the predicted label for s_{ij}, $\widehat{Y}(s_{ij})$, will be based on the maximum of these posterior probabilities, saying $\widehat{Y}(s_{ij}) = \text{argmax}_r\{P_r(s_{ij})\}$, where $r \in \{1, \cdots, k\}$. Therefore, for any s_{ij}, we have an indicator, $I(s_{ij})$, such that:

$$I(s_{ij}) = \begin{cases} 1, \widehat{Y}(s_{ij}) = Y(s_{ij}) \\ 0, otherwise \end{cases} \tag{7}$$

where, $I(s_{ij}) = 1$ indicating that the i^{th} fault sample from fault node j is correctly identified by the BN; $I(s_{ij}) = 0$ indicating this sample is incorrectly identified. If there are n fault samples in fault node j, we define that isolation accuracy (IA) for fault node j as:

$$IA_j = \frac{\sum_{i=1}^{n} I(s_{ij})}{n} \tag{8}$$

Once we have the IA for each fault node, we further define the average isolation accuracy (AIA) over all (say k) fault nodes, as:

$$AIA = \frac{\sum_{j=1}^{k} IA_j}{k} \tag{9}$$

If there are m evidence nodes in the BN, we define the sufficient isolation accuracy (SIA) for the BN as:

$$SIA = \frac{AIA}{m} \tag{10}$$

We expect to construct a robust BN containing as fewer evidence nodes as possible while maintaining satisfactory isolation accuracy, which can be measured by SIA. In other words, a robust BN has a higher SIA.

4.3. Experimental results

4.3.1. Sensitivity test on EECL

As is shown in **Algorithm 1**, a threshold ε is needed to determine if there is synchronicity among evidence nodes. As there is no set rule to determine ε to identify significant synchronicity, we conduct experiments by varying ε from 0.001 to 0.005 with increments of 0.001. This is because when ε is greater than 0.006, the PN is less than 0.5 indicating that there is no causal relationship according to [36]. Through observations, it is found that $\varepsilon = 0.005$ yields satisfactory results, as the SIA of the constructed BN under this threshold is 11.38 %, surpassing the results obtained from other values (refer to Table 5). Moreover, observing from Fig. 6, normalized EE values based on selected evidence nodes under fault conditions are below $\varepsilon = 0.005$, which agrees to our

Table 5
Results under different ε's ($\varepsilon = 0.005$ highlighted in grey).

ε	# of evidence nodes in BN	AIA	SIA
0.001	19	78 %	4.11 %
0.002	19	78 %	4.11 %
0.003	19	78 %	4.11 %
0.004	19	78 %	4.11 %
0.005	8	91%	11.38%

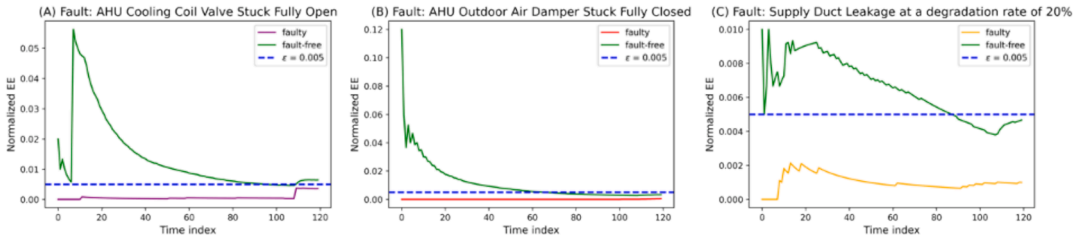


Fig. 6. Normalized EEs on selected evidence nodes over time for (A) AHU Cooling Coil Valve Stuck Fully Open; (B) AHU Outdoor Air Damper Stuck Fully Closed; and (C) Supply Duct Leakage at a degradation rate of 20 %. Each case shows normalized EEs under fault condition below the threshold ($\varepsilon = 0.005$), which agrees to our assumptions that fault conditions will lead to evidence synchronicity.

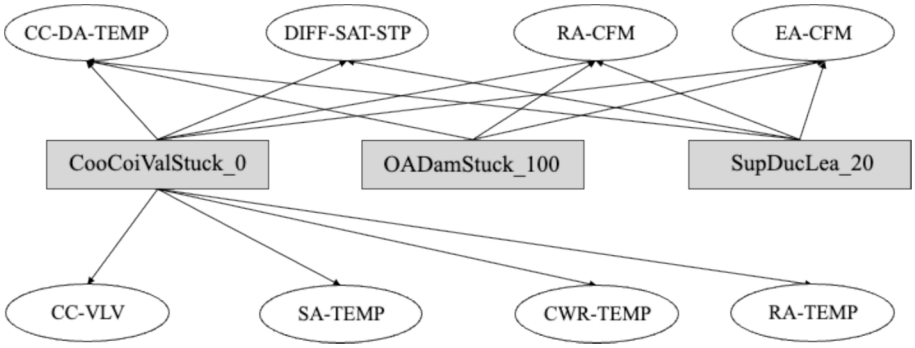


Fig. 7. BN constructed by EECL under $\varepsilon = 0.005$.

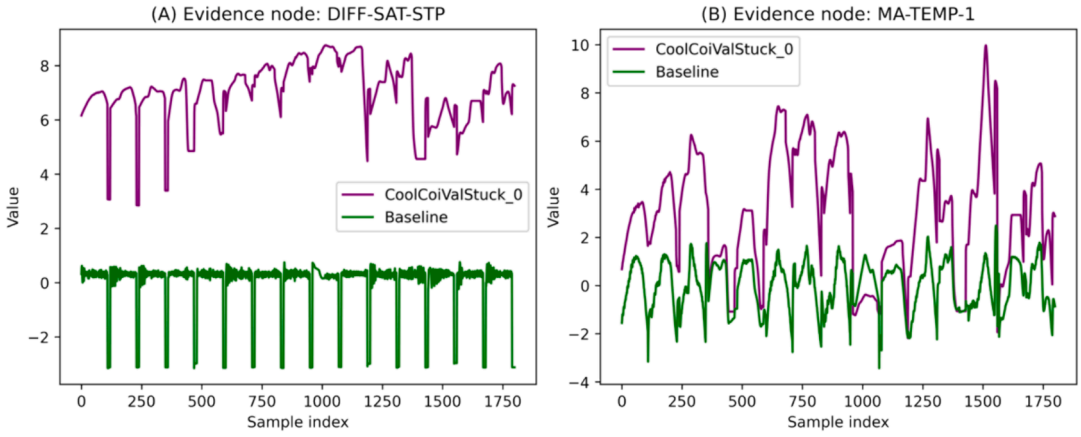


Fig. 8. Effects of the fault Cooling Coil Valve Stuck Fully Open on two evidence nodes: (A) DIFF-SAT-STP and (B) MA-TEMP-1.

assumption that fault conditions will lead to the synchronicity among evidence nodes. Therefore, we report the results with $\varepsilon = 0.005$ and the corresponding BN is shown in Fig. 7.

4.3.2. *BN derived from expert knowledge*

The structure of BN developed based on expert knowledge and physical analysis is shown as below. The values of an evidence node from a fault dataset are compared with those from a baseline dataset to

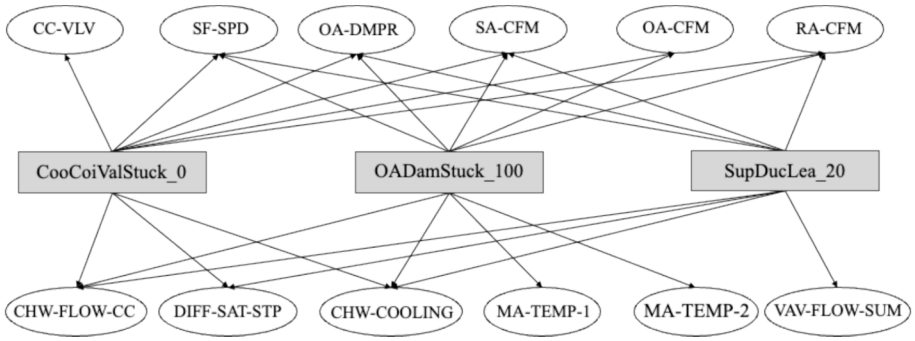


Fig. 9. BN constructed by expert knowledge [13].

observe if this evidence node demonstrates abnormality under a fault condition. Fig. 8 illustrates an example showing the effects of the fault: Cooling Coil Valve Stuck Fully Open on two evidence nodes, DIFF-SAT-STP (i.e., AHU Supply Air Temperature and Supply Air Temperature Setpoint Difference) and MA-TEMP-1 (i.e., Mixed Air Temperature). There are larger differences between the values of DIFF-SAT-STP under the fault scenario (see Fig. 8 (A) in purple) and those under baseline (see Fig. 8 (B) in green), while there are many overlaps between the values of MA-TEMP-1. Consequently, DIFF-SAT-STP rather than MA-TEMP-1 will be selected for the BN since this evidence node has shown significant abnormality under the fault condition according to the criteria described in [46]. Following the same procedures, nine evidence nodes are selected for the fault 'CooCoiValStuck_0', nine for the fault 'OADamStuck_100, and eight for the fault 'SupDucLea_20'. The final BN structure by expert knowledge is shown in Fig. 9, which includes twelve evidence nodes. The values of each evidence node under each fault node can be found in Appendix C.

4.3.3. BN derived from MIKK2 algorithm

The structure of BN developed based on Mutual Information-Krruskal-K2 (MIKK2) algorithm [50] is included as a benchmark. This algorithm initiates by computing the mutual information (MI) between variables, followed by utilizing the Kruskal algorithm from graph theory to develop the maximum spanning tree, thereby determining the maximum node in-degree. Subsequently, the maximum spanning tree undergoes a Depth First Search to establish the node order. Ultimately, the K2 algorithm utilizes both the node in-degree and the node order to learn and derive the optimal Bayesian Network structure. The final BN structure by MIKK2 is shown in Fig. 10, which includes nineteen evidence nodes.

4.3.4. Comparisons among three BN construction approaches

In this section, posterior probabilities for each fault node obtained using BNs by EECL, MIKK2 and expert knowledge are compared in the analysis. Two important parameters, prior and conditional probabilities predefined as in [51] are used for both faults and evidence nodes for posterior probability calculations. The corresponding comparison plots of posterior probabilities are shown in Fig. 11. The isolation accuracies for each individual fault using different BNs are as follows. For CooCoiValStuck_0, the isolation accuracy using BN by EECL is 98 %, significantly higher than that by MIKK2 (52 %), but slightly lower than that by expert knowledge (100 %); for OADamStuck_100, the isolation accuracy by EECL is 84 %, slightly higher than that by MIKK2 (81 %) but lower than that by expert knowledge (100 %); for SupDucLea_20, the isolation accuracy by EECL is 90 %, lower than that by MIKK2 (100 %) and that by expert knowledge (100 %). Observing from Fig. 12 (A) and (B), BN constructed by EE is with 8 evidence nodes and can achieve AIA of 91 %, which includes fewer evidence nodes and maintain higher AIA than that by MIKK2 (19 evidence nodes and AIA of 78 %); Although BN by expert knowledge can reach AIA of 100 %, it includes 50 % more evidence nodes than that by EECL. Moreover, as is observed from Fig. 12 (C), BN constructed by EECL reaches SIA of 11.38 %, higher than those by Expert (8.33 %) and by MIKK2 (4.11 %) respectively. This indicates the efficiency of EECL for BN construction, as EECL requires 33.3 % fewer evidence nodes and yields a 36.6 % higher SIA compared to expert knowledge, and 57.9 % fewer and 1.77 times higher compared to MIKK2, respectively.

As shown in Fig. 13, our constructed BN using EECL includes eight evidence nodes for CooCoiValStuck_0, three for OADamStuck_100, and four for SupDucLea_20; the numbers of evidence nodes by expert knowledge are nine, nine and eight, and those by MIKK2 are eleven, fifteen and seventeen, respectively; moreover, three evidence nodes are shared for CooCoiValStuck_0, one for OADamStuck_100, and two for SupDucLea_20 among three BNs, respectively. Therefore, our EECL method for BN construction is able to reach a satisfactory isolation accuracy for the cross-level fault diagnosis in the building systems for this given case study.

5. Discussions

In this research, the BN structure is constructed using a data-driven causal learning approach. To facilitate causal learning, the concept of "synchronicity" is introduced to describe the interactions among evidence nodes. The direction of causality from the fault status to the synchronicity is characterized by Pearl Causality. This process ultimately results in the construction of the BN structure that can be used to diagnose cross-level faults in a building HVAC system. As discussed in [13], automatic process of the BN structure construction is demanding due to the time-consuming and labor-intensive natures of determining BN structures by expert knowledge, and the developed EECL method has the potential for overcoming this deficiency since it is purely data-driven, and does not require any prior knowledge. Additionally, while expert knowledge method can determine the presence of the causal relationships between faults and evidence nodes, it does not provide any measures on the strength of these causations. In contrast, the proposed EECL method is able to quantify the causal relationships in terms of an evaluation metric (i.e., PN), which helps to reduce uncertainties of causality determined explicitly by expert knowledge.

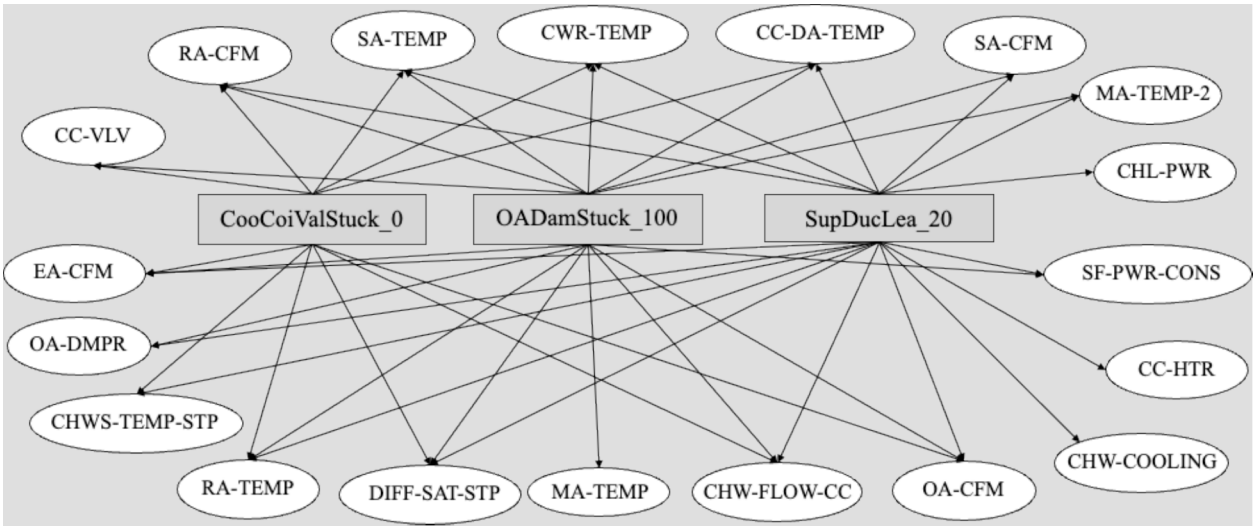


Fig. 10. BN constructed by MIKK2 algorithm [50].

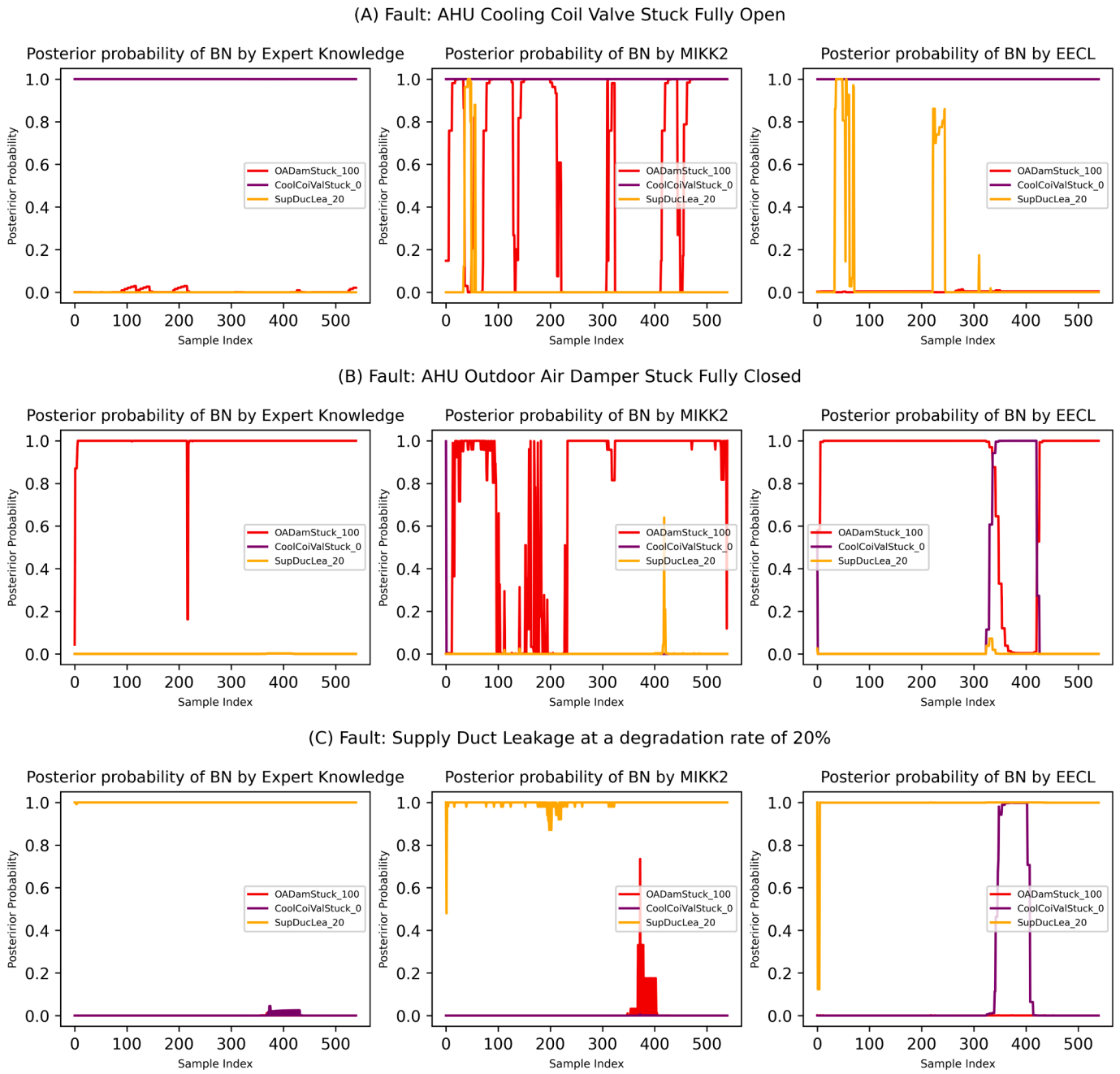


Fig. 11. Posterior probability results for test datasets from three fault nodes: (A) AHU Cooling Coil Valve Stuck Fully Open; (B) AHU Outdoor Air Damper Stuck Fully Closed; (C) Supply Duct Leakage at a rate of 20%. The left-hand sided plots are isolations from BN by expert knowledge, the middle sided plots are by MIKK2, and the right-hand sided plots are by EECL. All the posterior probabilities are generated by BayesFusion software [52,53].

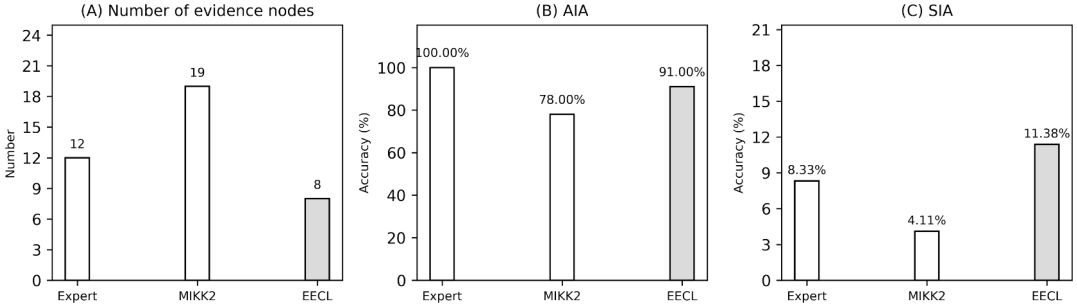


Fig. 12. Comparisons of BNs among three methods (Expert, MIKK2, EECL) in terms of (A) number of evidence nodes; (B) AIA and (C) SIA.

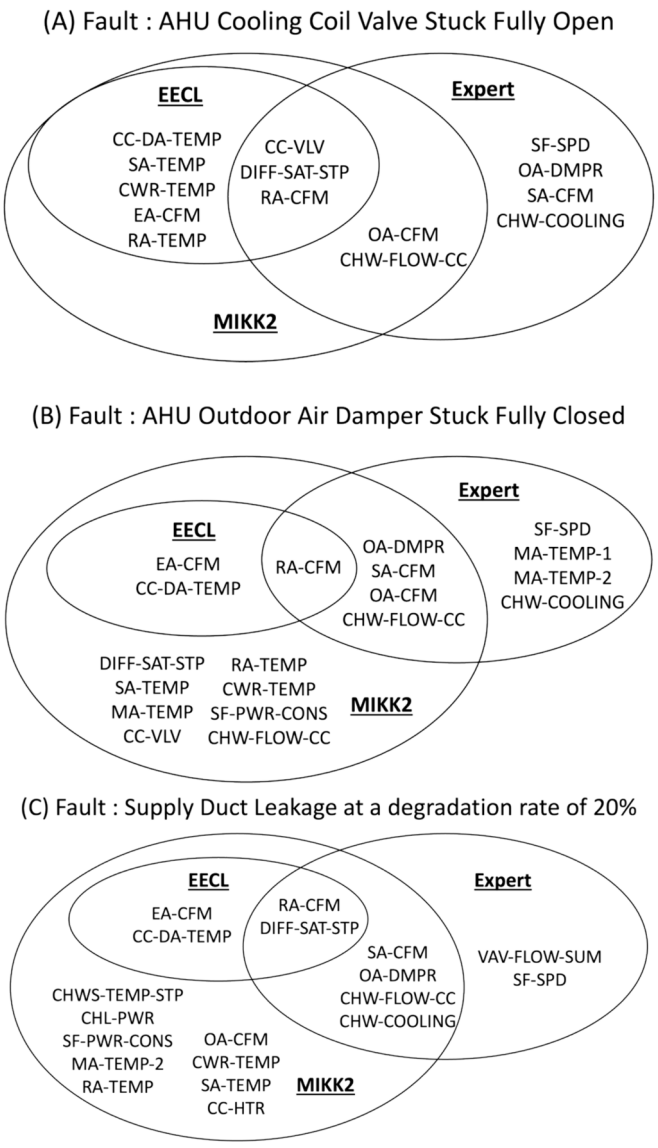


Fig. 13. Comparisons of BN evidence nodes by EECL, by MIKK2, and by expert knowledge for three fault nodes (common evidence nodes between three BNs are in the overlap of three circles).

In addition to causation characterizations, the proposed EECL method takes full considerations on interactions among the evidence nodes, which is measured by EE, an information entropy used for multivariate time-series. These interactions may reveal important and interesting patterns specific to a fault, but they may be overlooked by expert knowledge because experts often treat evidence nodes individually; This deficiency also applies in BN constructed by MIKK2 algorithm since evidence nodes are also treated as independent. The BN structure by EECL utilizes the parameter model that includes prior and conditional probabilities for the fault and evidence nodes determined by expert knowledge and is able to achieve satisfactory fault diagnosis/isolation accuracy (see Experiments).

As a matter of fact, several mainstream methods, such as randomized controlled trials, regression analysis, propensity score matching, have been widely used for causal effect estimations. These methods are to estimate the strength of the causal effect from one variable to another with a causal assumption given as a prior. In contrast, Pearl causality utilizes probability of necessity (PN) to assess whether the causal relationship is valid using frequency information. Specifically, decision criterion on causality (PN $\geq$ 0.5) is uniform; if PN is greater or equal to 0.5, the causal assumption holds true between the two variables. Since the objective of the study is to learn causal structure by determining which evidence nodes should be connected to the fault node, Pearl causality is adopted in this study.

While promising, it is worth noticing that the proposed EECL method like most data-driven approach is still influenced by several factors such as data volume, data quality, and data information. Despite the rapid development of data science and sensor technology, collections of a large amount of high-quality, information-rich data are still challenging with the current BAS. For example, the proposed EECL may discard a certain evidence node containing many missing values even though it is important and interpretable from a physical knowledge perspective. Another is that the BN by EECL still relies heavily on the expert knowledge to diagnose cross-level faults due to the same parameter model. There is a need for data-driven parameterizations (i.e., determination on prior and conditional probabilities) to support the diagnosability of BN by EECL. Moreover, Pearl causality in EECL acquires PN by using frequency information from binary outcomes (e.g., the outcome should be synchronicity/asynchronicity), which may not apply to a more complex scenario when outcomes are multi-class (e.g., weak, medium or strong synchronicity).

6. Conclusions and future work

This research develops an entropy-based causal learning method, termed eigen entropy-based causality learning (EECL), to support BN structure construction for fault diagnosis/isolation from the data-driven perspective. The proposed method includes two phases. In the first phase, Eigen-entropy is used for characterizing synchronicity, which describes the trends of movements among the evidence nodes over the time. In the second phase, counterfactual inference is applied to determine what and how evidence nodes should be connected to each fault node so as to build up the BNs to support fault diagnosis, including cross-level faults, in the building system. Compared to the traditional expert knowledge-based approach, the proposed method shares the following contributions: (1) it is a complete data-driven approach without the use of expert domain knowledge; (2) a term synchronicity is defined to capture the interactions, i.e., the trends of aligned movements over time, among multiple symptoms under fault status; (3) it utilizes cause effect estimations (counterfactual inferences) to induce the causal structure between faults and synchronicity among symptoms. The BN constructed by the developed EECL method is evaluated against that by expert knowledge based method using three cross-level fault test cases simulated using a virtual testbed. Experimental results show that the EECL based BN can achieve satisfactory isolation accuracy with fewer evidence nodes (average isolation accuracy of 91 % with 8 evidence nodes), indicating the efficacy of EECL approach for fault diagnosis.

Our current research focuses on a smaller set of fault test cases, each from one fault category. In the future, we are interested in investigating the capability of the proposed method for the fault diagnosis on multiple fault test cases, as well as the test cases from the same fault categories (i.e., various intensities under the same fault). Another interesting topic is to explore data-driven parameterizations for a robust diagnosability of BN by EECL. Finally, we use EE in conjunction with Pearl causality to construct BN, and there is a need to extend this EE causality framework to other types of causal structure learning methods such as functional causal models, score-based methods [25].

CRediT authorship contribution statement

Jiajing Huang: Writing – original draft, Methodology, Data curation, Conceptualization. **Naghmeh Ghalamsiah:** Writing – review & editing, Methodology, Data curation. **Abhidnya Patharkar:** Writing – review & editing, Conceptualization. **Ojas Pradhan:** Writing – review & editing, Methodology, Data curation. **Mengyuan Chu:** Data curation. **Teresa Wu:** Writing – review & editing, Supervision, Conceptualization.

Jin Wen: Writing – review & editing. **Zheng O'Neill:** Writing – review & editing. **Kasim Selcuk Candan:** Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research is supported by funds from the National Science Foundation award under the grant number 2309030 entitled "PIRE: Building Decarbonization via AI-empowered District Heat Pump Systems".

Appendix A:. Training and testing sets for the developed model

The HVAC system is sized for Chicago, IL, USA in climate zone 5A and the fault injection period starts at the beginning of the day on July 9 and continues for four weeks until August 5, the ranges of temperature and relative humidity are 24–29 °C and 50 % to 70 %.The data is collected at the sampling rate of 5 min, so 1,800 samples are used as the training set (representing 15 days with 10 h of operation each day). Also, 5 days are used for testing the developed BN network which has 6,00 samples (representing 5 days with 10 h of operation in each day). Table A-1 represents the same information (the excluded days in the table are unoccupied days that are removed from the dataset).

Table A1
Days considered for BN-based model training and validation.

Datasets	Occupied time
Training	Days 1–3; Days 6–10; Days 13–17; Days 20–21
Test	Days 22–24; Day 27–28

Appendix B:. Evidence node ranking method

In this study, we use the random forest classifier to characterize the importance score for each evidence node. Given three fault cases, CooCoi-ValStuck_0 (**C**), OADamStuck_100 (**O**), and SupDucLea_20 (**S**) and one baseline (**B**), we first identify critical evidence nodes that can differentiate all three fault cases (**C** vs. **O** vs. **S**). Next identify critical evidence nodes that can differentiate each pair of fault cases (**C** vs. **O**; **C** vs. **S**; **O** vs. **S**). Finally identify critical evidence nodes that can differentiate each fault case and baseline (**C** vs. **B**; **S** vs. **B**; **O** vs. **B**). The evidence nodes and corresponding importance scores can be referred to Table B-1.

Table B1
Critical evidence nodes and corresponding importance score information. Evidence nodes whose score > 0.05 are considered critical.

C vs. O vs. S		C vs. O		C vs. S		O vs. S	
Evidence	Score	Evidence	Score	Evidence	Score	Evidence	Score
CC-DA-TEMP	0.1024	CC-DA-TEMP	0.1741	CC-DA-TEMP	0.1773	RA-CFM	0.1889
RA-CFM	0.0912	DIFF-SAT-STP	0.1448	DIFF-SAT-STP	0.1492	EA-CFM	0.1754
EA-CFM	0.0880	SA-TEMP	0.1007	CWR-TEMP	0.0998	OA-CFM	0.1182
DIFF-SAT-STP	0.0826	CC-VLV	0.0786	SA-TEMP	0.0858	OA-DMPR	0.1100
SA-TEMP	0.0662	RA-TEMP	0.0766	CHWS-TEMP-STP	0.0758	SA-CFM	0.0635
OA-CFM	0.0546	CWR-TEMP	0.0739	CHW-FLOW-CC	0.0601	SF-PWR-CONS	0.0594
CWR-TEMP	0.0510	CHW-FLOW-CC	0.0512	RA-TEMP	0.0509		

C vs. B		O vs. B		S vs. B	
Evidence	Score	Evidence	Score	Evidence	Score
CC-DA-TEMP	0.1719	OA-DMPR	0.1628	EA-CFM	0.1416
DIFF-SAT-STP	0.1423	EA-CFM	0.1202	CHW-COOLING	0.1281
CC-VLV	0.1418	RA-CFM	0.1055	RA-CFM	0.0768
SA-TEMP	0.1024	OA-CFM	0.1041	CC-HTR	0.0759
RA-TEMP	0.0812	MA-TEMP-2	0.0992	CHL-PWR	0.0664
CWR-TEMP	0.0520	MA-TEMP	0.0971	SA-CFM	0.0644
				MA-TEMP-2	0.0568

Given critical evidence nodes identified and corresponding importance score under different criteria, we rank these evidence nodes for each test case according to the max importance score. Take the test case CooCoiValStuck_0 as an example (see Table B-2). Since the max value of importance score for CC-DA-TEMP is 0.1773 while that of DIFF-SAT-STP is 0.1492, then CC-DA-TEMP ranks before DIFF-SAT-STP. Similarly, we have the evidence nodes importance information for other two cases (see Tables B-3 & B-4). Finally, ranking information about the evidence nodes for each test cases are summarized in Table B-5.

Table B2
Critical evidence nodes and corresponding importance scores for CooCoiValStuck_0.

Evidence	C vs. O vs. S	C vs. O	C vs. S	C vs. B	Max Value
CC-DA-TEMP	0.1024	0.1741	0.1773	0.1719	0.1773
CC-VLV	0	0.0786	0	0.1418	0.1418
DIFF-SAT-STP	0.0826	0.1448	0.1492	0.1423	0.1492
CWR-TEMP	0.0510	0.0739	0.0998	0.0520	0.0998
RA-CFM	0.0912	0	0	0	0.0912
CHW-FLOW-CC	0	0.0512	0.0601	0	0.0601
SA-TEMP	0.0662	0.1007	0.0858	0.1024	0.1024
OA-CFM	0.0546	0	0	0	0.0546
EA-CFM	0.0880	0	0	0	0.0880
RA-TEMP	0	0.0766	0.0509	0.0812	0.0812
CHWS-TEMP-STP	0	0.0374	0.0758	0	0.0758

Table B3
Critical evidence nodes and corresponding importance scores for OADamStuck_100.

Evidence	C vs. O vs. S	C vs. O	O vs. S	O vs. B	Max Value
OA-DMPR	0.0428	0	0.1100	0.1628	0.1628
SA-TEMP	0.0662	0.1007	0	0	0.1007
RA-TEMP	0	0.0766	0	0	0.0766
EA-CFM	0.0880	0	0.1754	0.1202	0.1754
CHW-FLOW-CC	0	0.0512	0	0	0.0512
MA-TEMP	0	0	0	0.0971	0.0971
RA-CFM	0.0912	0	0.1889	0.1055	0.1889
CWR-TEMP	0.0510	0.0739	0	0	0.0739
DIFF-SAT-STP	0.0826	0.1448	0	0	0.1448
MA-TEMP-2	0	0	0	0.0992	0.0992
OA-CFM	0.0546	0	0.1182	0.1041	0.1182
SA-CFM	0.042	0	0.0635	0	0.0635
SF-PWR-CONS	0	0	0.0594	0	0.0594
CC-DA-TEMP	0.1024	0.1741	0	0	0.1741
CC-VLV	0	0.0786	0	0	0.0786

Table B4
Critical evidence nodes and corresponding importance scores for SupDucLea_20.

Evidence	C vs. O vs. S	C vs. S	O vs. S	S vs. B	Max Value
CHL-PWR	0	0.0456	0	0.0664	0.0664
OA-DMPR	0.0428	0	0.1100	0	0.1100
RA-CFM	0.0912	0	0.1889	0.0768	0.1889
DIFF-SAT-STP	0.0826	0.1492	0	0	0.1492
CHW-COOLING	0	0	0	0.1281	0.1281
OA-CFM	0.0546	0	0.1182	0	0.1182
CC-HTR	0	0	0	0.0759	0.0759
EA-CFM	0.0880	0	0.1754	0.1416	0.1754
MA-TEMP-2	0	0	0	0.0568	0.0568
CC-DA-TEMP	0.1024	0.1773	0	0	0.1773
SA-CFM	0.0424	0	0.0635	0.0644	0.0644
CWR-TEMP	0.0510	0.0998	0	0	0.0998
RA-TEMP	0	0.0509	0	0	0.0509
CHWS-TEMP-STP	0	0.0758	0	0	0.0758
SA-TEMP	0.0662	0.0858	0	0	0.0858
CHW-FLOW-CC	0	0.0601	0	0	0.0601
SF-PWR-CONS	0	0	0.0594	0	0.0594

Table B5
Ranking of critical evidence nodes for three test cases.

Rank	CooCoiValStuck_0		OADamStuck_100		SupDucLea_20	
	Evidence	Max Value	Evidence	Max Value	Evidence	Max Value
1	CC-DA-TEMP	0.1773	RA-CFM	0.1889	RA-CFM	0.1889
2	DIFF-SAT-STP	0.1492	EA-CFM	0.1754	CC-DA-TEMP	0.1773
3	CC-VLV	0.1418	CC-DA-TEMP	0.1741	EA-CFM	0.1754
4	SA-TEMP	0.1024	OA-DMPR	0.1628	DIFF-SAT-STP	0.1492
5	CWR-TEMP	0.0998	DIFF-SAT-STP	0.1448	CHW-COOLING	0.1281
6	RA-CFM	0.0912	OA-CFM	0.1182	OA-CFM	0.1182

(*continued on next page*)

Table B5 (*continued*)

Rank	CooCoiValStuck_0		OADamStuck_100		SupDucLea_20	
	Evidence	Max Value	Evidence	Max Value	Evidence	Max Value
7	EA-CFM	0.0880	SA-TEMP	0.1007	OA-DMPR	0.1100
8	RA-TEMP	0.0812	MA-TEMP-2	0.0992	CWR-TEMP	0.0998
9	CHWS-TEMP-STP	0.0758	MA-TEMP	0.0971	SA-TEMP	0.0858
10	CHW-FLOW-CC	0.0601	CC-VLV	0.0786	CC-HTR	0.0759
11	OA-CFM	0.0546	RA-TEMP	0.0766	CHWS-TEMP-STP	0.0758
12			CWR-TEMP	0.0739	CHL-PWR	0.0664
13			SA-CFM	0.0635	SA-CFM	0.0644
14			SF-PWR-CONS	0.0594	CHW-FLOW-CC	0.0601
15			CHW-FLOW-CC	0.0512	SF-PWR-CONS	0.0594
16					MA-TEMP-2	0.0568
17					RA-TEMP	0.0509

Appendix C:. Defining fault-evidence connections by expert knowledge

The following figures show the standardized value of evidence nodes in the BN constructed by expert knowledge for three fault nodes, 'CoolCoiValStuck_0′, 'OADamStuck_100′ fault and 'SupDucLea_20′ fault respectively. Evidence nodes with '**' are those associated with the specific fault.

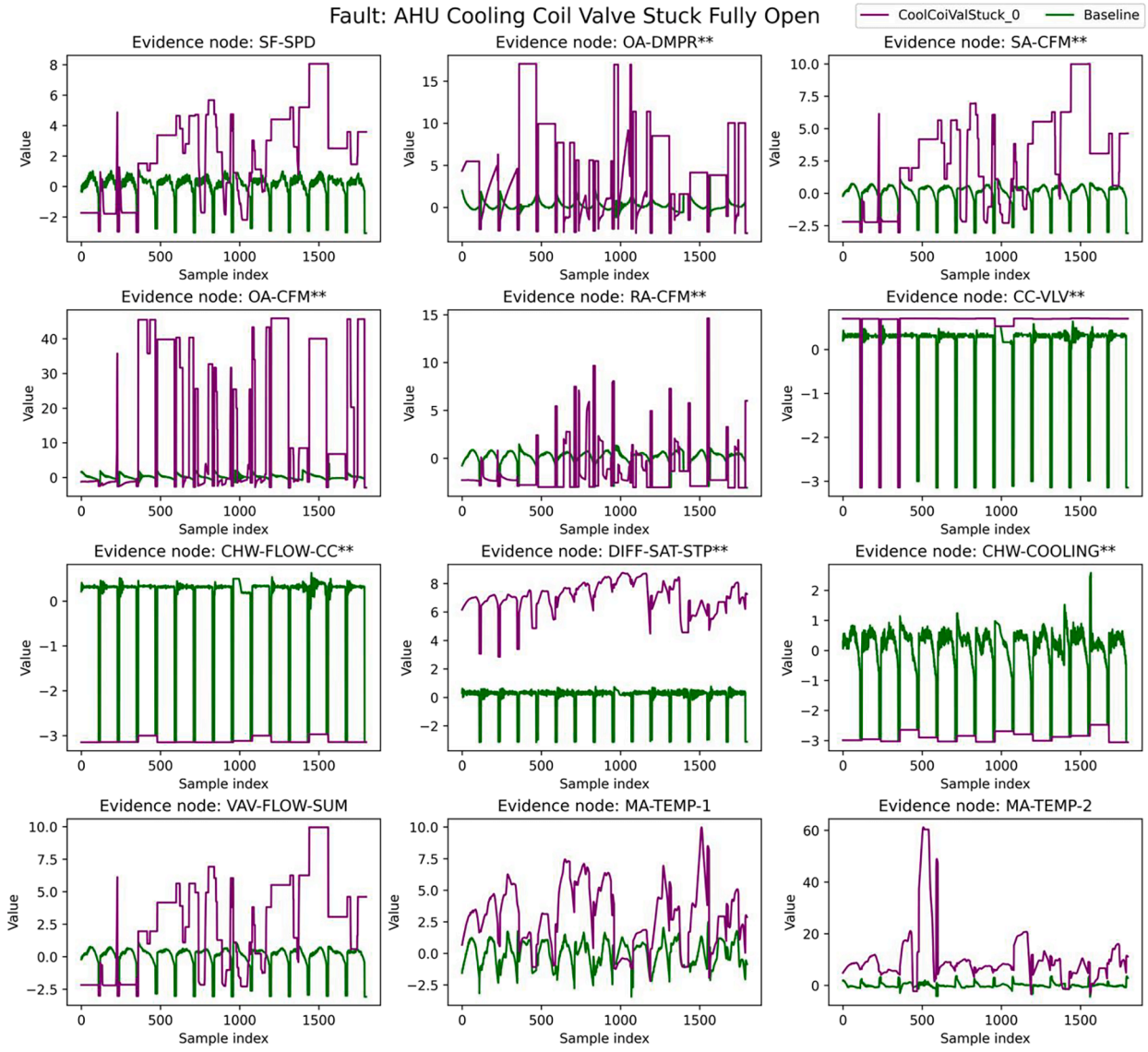


Fig. C14.

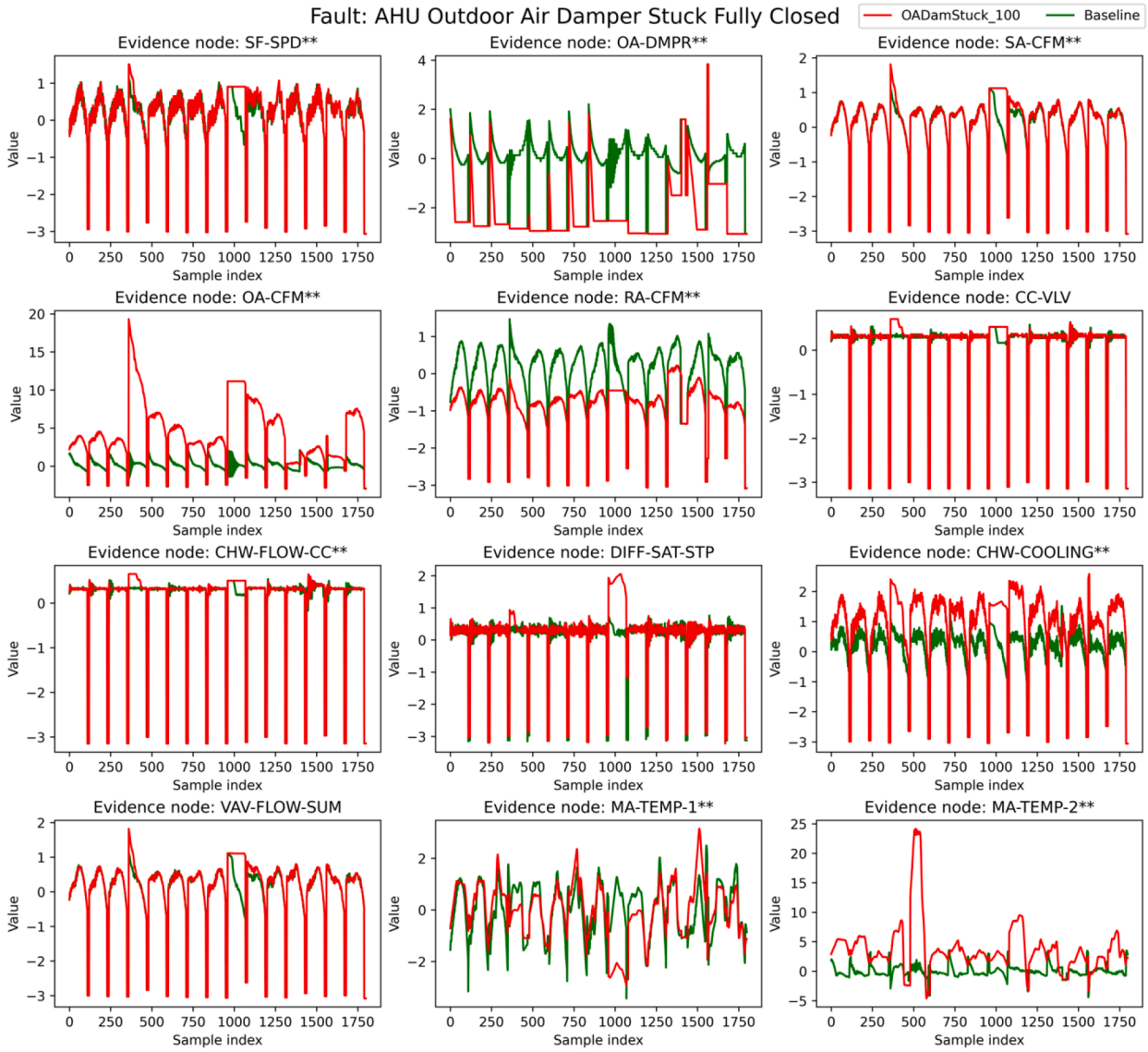

Fault: AHU Outdoor Air Damper Stuck Fully Closed
OADamStuck_100
Baseline
Evidence node: SF-SPD**
Evidence node: OA-DMPR**
Evidence node: SA-CFM**
Evidence node: OA-CFM**
Evidence node: RA-CFM**
Evidence node: CC-VLV
Evidence node: CHW-FLOW-CC**
Evidence node: DIFF-SAT-STP
Evidence node: CHW-COOLING**
Evidence node: VAV-FLOW-SUM
Evidence node: MA-TEMP-1**
Evidence node: MA-TEMP-2**
Value
Sample index

Fig. C15.

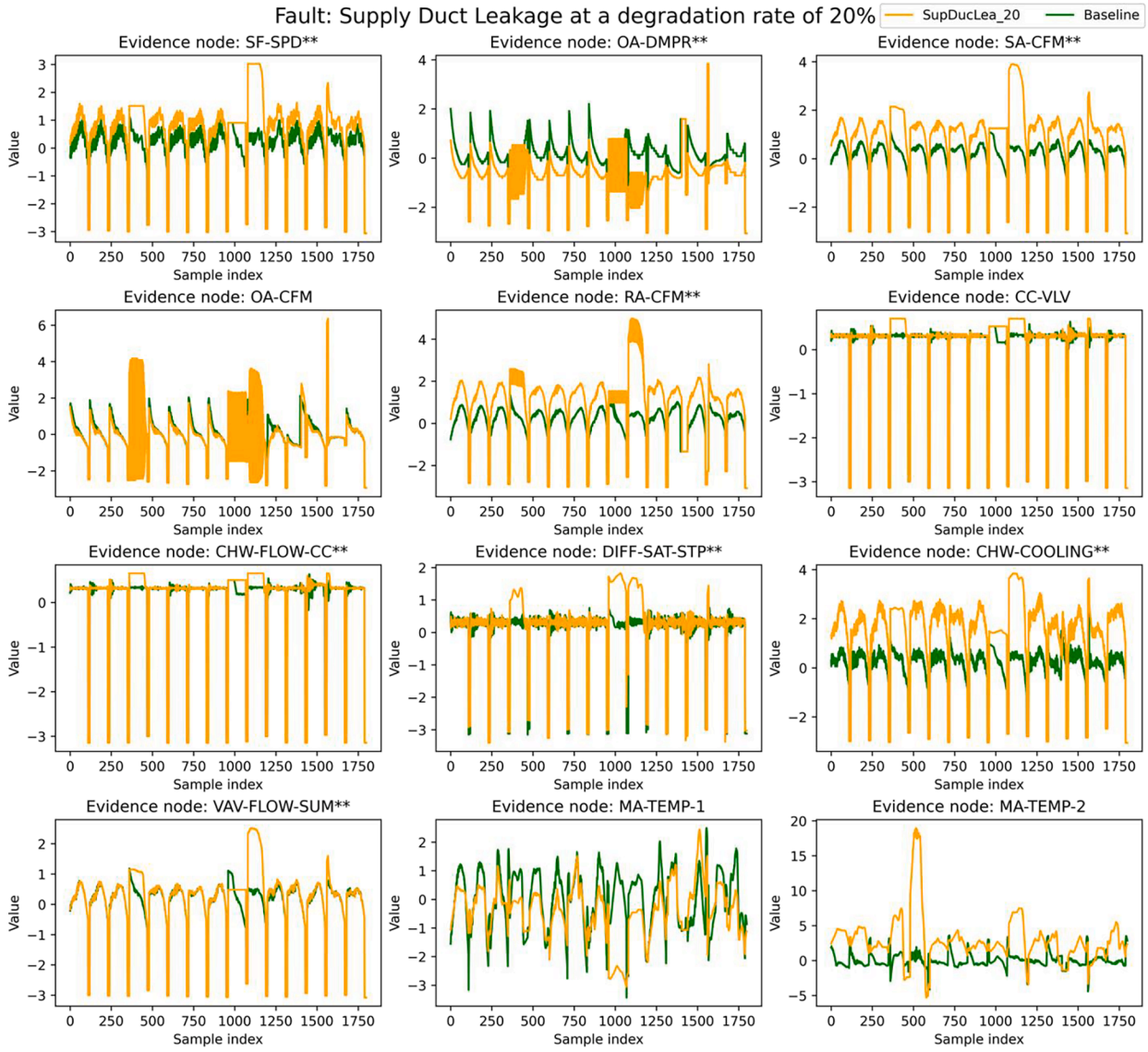


Fig. C16.

References

[1] United Nations Environment Programme. 2022 global status report for buildings and construction: towards a zero-emission, efficient and resilient buildings and construction sector. 2022.

[2] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy Build 40 (3) (2008) 394–1338.

[3] IEA - Annex 25 - Real Time Simulation for Building Optimization, Fault Detection and Diagnosis: https://www.iea-ebc.org/Data/publications/EBC_Annex_25_source_book.pdf.

[4] M.R. Brambley, S. Katipamula, Commercial building retuning: a low-cost approach to improved performance and energy efficiency, ASHRAE J 51 (10) (2009) 12–23.

[5] K.W. Roth, D. Westphalen, P. Llana, M. Feng, The energy impact of faults in us commercial buildings, in: Proceedings of the 2004 International Refrigeration and Air Conditioning Conference, 2004, p. 665.

[6] M.A. Piette, S.K. Kinney, P. Haves, Analysis of an information monitoring and diagnostic system to improve building operations, Energy Build 33 (8) (2001) 783–791.

[7] J. Granderson, G. Lin, A. Harding, P. Im, Y. Chen, Building fault detection data to aid diagnostic algorithm creation and performance testing, Sci Data 7 (2020) 65.

[8] B.L. Capehart, M.R. Brambley, Automated diagnostics and analytics for buildings, first ed., River Publishers, New York, 2014.

[9] CIBSE Guide H: Building Control Systems, Routledge, Oxford, British, 2007.

[10] S. Ginestet, D. Marchio, O. Morisot, Evaluation of faults impacts on energy consumption and indoor air quality on an air handling unit, Energy Build 40 (2008) 51–57.

[11] Y. Yan, P.B. Luh, K.R. Pattipati, Fault diagnosis of HVAC air-handling systems considering fault propagation impacts among components, IEEE Trans Autom Sci Eng 14 (2017) 705–717.

[12] N. Cauchi, K.A. Hoque, M. Stoelinga, A. Abate, Maintenance of smart buildings using fault trees, ACM Trans Sen Netw 14 (2018) 1–25.

[13] Y. Chen, J. Wen, O. Pradhan, L.J. Lo, T. Wu, Using discrete Bayesian networks for diagnosing and isolating cross-level faults in HVAC systems, Appl Energy 327 (2022) 120050.

[14] Z. Chen, Z. O'Neill, J. Wen, O. Pradhan, T. Yang, X. Lu, et al., A review of data-driven fault detection and diagnostics for building HVAC systems, Appl Energy 339 (2023) 121030.

[15] J. Wang, Z. Yang, J. Su, Y. Zhao, S. Gao, X. Pang, D. Zhou, Root-cause analysis of occurring alarms in thermal power plants based on Bayesian networks, Int J Electr Power Energy Syst 103 (2018) 67–74.

[16] A. Lokrantz, E. Gustavsson, M. Jirstrand, Root cause analysis of failures and quality deviations in manufacturing using machine learning, Procedia CIRP 72 (2018) 1057–1062.

[17] N. Liu, M. Hu, J. Wang, Y. Ren, W. Tian, Fault detection and diagnosis using Bayesian network model combining mechanism correlation analysis and process data: Application to unmonitored root cause variables type of faults, Process Saf Environ Prot 164 (2022) 15–29.

[18] M.T. Amin, F. Khan, S. Ahmed, S. Imtiaz, A data-driven Bayesian network learning method for process fault diagnosis, Process Saf Environ Prot 150 (2021) 110–122.

[19] M.T. Amin, F. Khan, S. Imtiaz, Fault detection and pathway analysis using a dynamic Bayesian network, Chem Eng Sci 195 (2019) 777–790.
[20] Y. Wang, Z. Wang, S. He, Z. Wang, A practical chiller fault diagnosis method based on discrete Bayesian network, Int J Refrig 102 (2019) 159–167.
[21] Y. Chen, J. Wen, T. Chen, O. Pradhan, Bayesian Networks for whole building level fault diagnosis and isolation, in: Proceedings of the 2018 International High Performance Buildings Conference, 2018, p. 266.
[22] Z. Wang, L. Wang, Y. Tan, J. Yuan, X. Li, Fault diagnosis using fused reference model and Bayesian network for building energy systems, J Build Eng 34 (2021) 101957.
[23] A. Taal, L. Itard, P&ID-based automated fault identification for energy performance diagnosis in HVAC systems: 4S3F method, development of DBN models and application to an ATES system, Energy Build. 224 (2020) 110289.
[24] Pradhan O, Wen J, Chen Y, Lu X, Chu M, Fu Y, et al. Dynamic bayesian network-based fault diagnosis for ASHRAE guideline 36: high performance sequence of operation for HVAC systems. In: Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. 2021, p. 365-68.
[25] L. Cheng, R. Guo, R. Moraffah, P. Sheth, K.S. Candan, H. Liu, Evaluation methods and measures for causal learning algorithms, IEEE Trans Artif Intell 3 (2022) 924–943.
[26] B. Schölkopf, Causality for machine learning, in: H. Geffner, R. Dechter, J. Y. Halpern (Eds.), Probabilistic and Causal Inference: the Works of Judea Pearl, first ed., Association for Computing Machinery, New York, 2022, pp. 765–804.
[27] J. Pearl, Causality: models, reasoning and inference, second ed., Cambridge University Press, New York, 2009.
[28] J. Pearl, Probabilistic reasoning in intelligent systems: networks of plausible inference, first ed., Morgan Kaufmann Publishers Inc, San Francisco, 1988.
[29] L. Jiang, L. Zhang, C. Li, J. Wu, A correlation-based feature weighting filter for naïve Bayes, IEEE Trans Knowl Data Eng 31 (2) (2019) 201–213.
[30] L. Jiang, L. Zhang, L. Yu, D. Wang, Class-specific attribute weighted naïve Bayes, Pattern Recognit 88 (2019) 321–330.
[31] Y. Chen, J. Wen, L.J. Lo, Using weather and schedule based pattern catching and feature based PCA for whole building fault detection — part I development of the method, ASME J Eng Sustain Build Cities 3 (1) (2022) 011001.
[32] Y. Chen, J. Wen, L.J. Lo, Using weather and schedule based pattern catching and feature based PCA for whole building fault detection — part II field evaluation, ASME J Eng Sustain Build Cities 3 (1) (2022) 011002.
[33] J. Huang, H. Yoon, T. Wu, K.S. Candan, O. Pradhan, J. Wen, Z. O'Neill, Eigen-Entropy: a metric for multivariate sampling decisions, Inf Sci 619 (2023) 84–97.
[34] D. Barber, Bayesian reasoning and machine learning, first ed., Cambridge University Press, Cambridge, 2012.
[35] J. Pearl, Probabilities of causation: three counterfactual interpretations and their identification, Synthese 121 (1999) 93–149.
[36] J. Tian, J. Pearl, Probabilities of causation: bounds and identification, Annals of Mathematics and Articial Intelligence 28 (2000) 287–313.
[37] C.E. Shannon, A mathematical theory of communication, Bell Syst. Technol 27 (1948) 379–423.
[38] J. Huang, H. Yoon, O. Pradhan, T. Wu, J. Wen, Z. O'Neill, K.S. Candan, A cosine-based correlation information entropy approach for building automatic fault detection baseline construction, Sci Technol Built Environ 28 (9) (2022) 1138–1149.
[39] J. Huang, T. Wu, H. Yoon, O. Pradhan, J. Wen, Z., O'Neill, Automatic fault detection baseline construction for building HVAC systems using joint entropy and enthalpy, in: Proceedings of 2021 Institute of Industrial and Systems Engineers, 2021, pp. 536–541.
[40] P. Fritzson, Principles of object oriented modeling and simulation with Modelica 3.3: a cyber-physical approach, 2nd ed.,, Wiley-IEEE Press, Piscataway, 2014.
[41] M. Wetter, W. Zuo, T.S. Nouidui, X. Pang, Modelica buildings library, J Build Perform Simul 7 (4) (2014) 253–270.
[42] S. Goel, R.A. Athalye, W. Wang, Enhancements to ASHRAE standard 90.1 prototype building model, Richland, WA (United States): Pacific Northwest National Lab, (PNNL) (2014).
[43] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W.F. Buhl, Y.J. Huang, C.O. Pedersen, et al., EnergyPlus: creating a new-generation building energy simulation program, Energy Build 33 (4) (2001) 319–331.
[44] ASHRAE, ASHRAE Guideline 36-2018, High-Performance Sequences of Operation for HVAC Systems. Tullie Cir NE, Atlanta, GA; 2018.
[45] Taylor Engineering LLP, S.T. Taylor, B. Gill, R. Kiriu, Advanced sequences of operation for HVAC systems – Phase II Central plants and hydronic systems. Technical report research project 1711. Task 5: Reporting of findings, Atlanta, GA: ASHRAE; 2019.
[46] Y. Fu, Z. O'Neill, Z. Yang, V. Adetola, J. Wen, L. Ren, et al., Modeling and evaluation of cyber-attacks on grid-interactive efficient buildings, Appl Energy 303 (2021) 117639.
[47] Y. Fu, Z. O'Neill, V. Adetola, A flexible and generic functional mock-up unit based threat injection framework for grid-interactive efficient buildings: A case study in Modelica, Energy Build 250 (2021) 111263.
[48] X. Lu, Y. Fu, Z. O'Neill, J. Wen, A holistic fault impact analysis of the high-performance sequences of operation for HVAC systems: Modelica-based case study in a medium-office building, Energy Build 252 (2021) 111448.
[49] G. Li, Z. Yang, Y. Fu, Z. O'Neill, L. Ren, O. Pradhan, J. Wen, A hardware-in-the-loop (HIL) testbed for cyber-physical energy systems in smart commercial buildings, Sci Technol Built Environ (2024) 1–18.
[50] X. Li, H. Yu, Bayesian Network Structure Learning Algorithm Based on Node Order Constraint, In: 2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI). 2022, p. 214-217.
[51] O. Pradhan, A,, dynamic Bayesian network framework for data-driven fault diagnosis and prognosis of smart building systems. Doctoral dissertation, Drexel University, 2023.
[52] BayesFusion. SMILE Wrappers Programmer's Manual. 2021.
[53] BayesFusion. GeNIe Modeler User Manual Version 2.2.4. 2018.