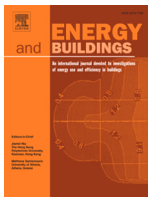
Energy & Buildings 354 (2026) 116913

Contents lists available at ScienceDirect

Energy & Buildings

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



Fault detection and diagnosis for heat recovery ventilation using 4S3F method: Impact of diverse sensor configurations

Ziao Wang [a], Chujie Lu [a,*], Arjen Meijer [a], Shalika Walker [b], Laure Itard [a]

[a] TU Delft, Delft, 2628 BL, The Netherlands
[b] Kropman B.V., Nijmegen, 6545 CG, The Netherlands

ARTICLE INFO

Keywords:
Fault detection and diagnosis
Diagnostic Bayesian network
Air handling units
Sensor configuration
Heat recovery system

ABSTRACT

Fault detection and diagnosis (FDD) are crucial to improving the efficiency of heating, ventilation, and air conditioning (HVAC) systems, reducing energy waste, and maintaining indoor comfort. Diagnostic Bayesian Networks (DBNs) present a compelling approach, offering robustness to uncertainty, adaptability to different sensor configurations, and interpretable inference. Existing FDD studies for air handling units (AHUs), however, are often limited to simulation or laboratory settings, seldom consider AHUs with heat recovery wheel (HRW) in operation, and rarely analyze how diagnostic performance changes under diverse sensor configurations. This study defined three practical sensor configurations (Sensor-Rich, Standard, and Limited) based on international guidelines and a practical survey, developed a corresponding DBN framework, and evaluated its performance on seventeen common faults using real-world data from an AHU in a Dutch office building. Existing FDD studies are often limited to simulation or specific Air Handling Unit (AHU) types with fixed sensor configurations, rarely investigating AHUs with heat recovery wheels, which are common in Europe. This study addresses these gaps by first defining three sensor configurations (Sensor-Rich, Standard, and Limited) based on international guidelines and a practitioner survey. A DBN-based FDD model was then developed for these configurations using historical data, expert knowledge and subsequently evaluated for its ability to diagnose seventeen common faults in an operational AHU with heat recovery wheel.The DBN correctly diagnosed fifteen, nine, and four faults for these configurations, respectively. The results show that increasing sensor availability improves overall diagnostic performance. However, certain cases demonstrate that additional measurements can also introduce conflicting evidence and reduce diagnostic accuracy. The study suggests that sensor selection must be combined with effective DBN modeling strategies to achieve robust diagnosis. Taken together, the analysis of key sensors and DBN modeling practices provides practical guidance for designing and implementing DBN-based FDD in common European AHU systems under diverse sensor configurations.The results indicate that increasing sensor quantity alone does not improve FDD performance; strategic sensor selection, placement, and effective data processing are also crucial.

1. Introduction

According to official reports from the European Parliament and Council, commercial and residential buildings collectively account for approximately 40% of total energy consumption in EU countries [1]. Among this, Heating, Ventilation, and Air Conditioning (HVAC) systems, which are essential to maintaining indoor air quality and climate comfort, are primary energy consumers in buildings [2]. However These systems are also prone to faults, such as component malfunctions, improper installations, sensor biases, and suboptimal setpoints, resulting in substantial energy waste in buildings [3,4]. Consequently, there is an urgent need to implement effective Fault Detection and Diagnosis (FDD) tools, not only to ensure a healthy and comfortable indoor environment but also to minimize energy waste and reduce maintenance costs.

This work focuses on FDD for Air Handling Units (AHU), essential HVAC components that connect primary heating and cooling sources with building zones, regulate fresh air intake and significantly impact building energy use [5]. FDD tools for AHUs can generally be divided into two categories: knowledge-based and data-driven approaches [6,7]. In knowledge-based approaches, fault detection rules are developed based on predefined rules or physical models, and diagnosis simulates the diagnostic reasoning of HVAC experts [8,9]. AHU performance assessment rule (APAR) is a classical knowledge-based FDD tool that uses a set of expert rules derived from mass and energy balances to

* Corresponding author.
E-mail addresses: Ziao.Wang@tudelft.nl (Z. Wang), C.J.Lu@tudelft.nl (C. Lu), A.Meijer@tudelft.nl (A. Meijer), shalika.walker@kropman.nl (S. Walker), l.c.m.itard@tudelft.nl (L. Itard).

https://doi.org/10.1016/j.enbuild.2025.116913
Received 3 September 2025; Received in revised form 25 November 2025; Accepted 24 December 2025
Available online 1 January 2026

Nomenclature

No.	Name of the Variable	Sign	Unit
1	Indoor air temperature	T_{in}	$^{\circ}C$
2	Indoor air temperature setpoint	$T_{in,set}$	$^{\circ}C$
3	Supply air temperature	T_{sa}	$^{\circ}C$
4	Supply air temperature setpoint	$T_{sa,set}$	$^{\circ}C$
5	Supply air temperature in distribution	T_{sad}	$^{\circ}C$
6	Preheat air temperature	T_{pre}	$^{\circ}C$
7	Return air temperature	T_{ra}	$^{\circ}C$
8	Return air temperature in distribution	T_{rad}	$^{\circ}C$
9	Outdoor air temperature	T_{oa}	$^{\circ}C$
10	Heating coil valve openness	U_{hc}	%
11	Predicted heating coil valve openness	$U_{hc,pred}$	%
12	Supply air static pressure	P_{sa}	Pa
13	Supply air static pressure setpoint	$P_{sa,set}$	Pa
14	Predicted supply air static pressure	$P_{sa,pred}$	Pa
15	Filter pressure difference	ΔP_{fi}	Pa
16	Supply fan pressure difference	ΔP_{sf}	Pa
17	Predicted fan pressure difference	$P_{sf,pred}$	Pa
18	Supply fan speed	N_{sf}	%
19	Predicted supply fan speed	$N_{sf,pred}$	%
20	Supply air flow rate	F_{sa}	m^3/h
21	Nominal supply air flow rate	$F_{sa,nm}$	m^3/h
22	Nominal heat recovery efficiency	$\eta_{hrw,n}$	%
23	Maximum heat recovery efficiency	$\eta_{hrw,max}$	%
24	Heat recovery efficiency (Supply)	η_{sa}	%
25	Heat recovery efficiency (Exhaust)	η_{eha}	%
26	CO_2 concentration	CO_2	ppm
27	Supply fan speed	N_{sf}	%
28	Predicted supply air temperature	$T_{sa,pred}$	$^{\circ}C$
29	Exhausted air temperature	T_{ea}	$^{\circ}C$
30	Supply coil water temperature	T_{sw}	$^{\circ}C$
31	Supply air flow rate in Zone 1.05	$F_{sa,105}$	m^3/h
32	Supply air flow rate in Zone North	$F_{sa,north}$	m^3/h
33	Supply air flow rate in Zone South	$F_{sa,sorth}$	m^3/h
34	Indoor air temperature in Zone 1.05	$T_{in,105}$	$^{\circ}C$
35	Indoor air temperature in Zone North	$T_{in,north}$	$^{\circ}C$
36	Indoor air temperature in Zone South	$T_{in,south}$	$^{\circ}C$
37	Indoor air temperature setpoint in Zone 1.05	$T_{in,105set}$	$^{\circ}C$
38	Indoor air temperature setpoint in Zone North	$T_{in,northset}$	$^{\circ}C$
39	Indoor air temperature setpoint in Zone South	$T_{in,south}$	$^{\circ}C$

diagnose faults [10]. The development of knowledge-based FDD approaches relies heavily on domain knowledge and may be less adaptive to the diversitydiversety of AHU design and operation. On the other hand, data-driven approaches reduce strongly dependency on domain knowledge and offer a more flexible alternative by learning from historical data to automatically detect and classify faults. Several popular machine learning algorithms have been successfully applied to diagnose faults in AHUs, including Support Vector Machine (SVM) [11,12], Extreme Gradient Boosting (XGBoost) [13–15], Convolutional Neural Network (CNN)[16–18] Transformer-based neural network (Transformer) [19], Multi-Task Network (MTN) [20], Generative Adversarial Network (GAN) [20,21]. Despite their strong diagnostic capabilities, data-driven approaches are highly dependent on large amounts of high-quality training data, which is often unavailable in real-world buildings [22,23]. Furthermore, since they are usually trained on specific datasets, their applicability is typically limited to particular AHU systems or operational conditions, limiting their generalizability [24–26]. Diagnostic Bayesian networks (DBNs), as probabilistic graphical models, provide a promising solution. They align with HVAC design and implementation processes, offering key advantages such as flexible modeling (accommodating both knowledge-based and data-driven approaches), robustness to incompleteness and inaccuracy in sensor measurements, and enhanced modeling interpretability [27–29]. Zhao et al. [28,30] successfully applied DBNs for AHUs of a typical small commercial building from the American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE) Project RP-1312. Li et al. [31] proposed a hybrid approach that leverages an improved genetic algorithm to optimize the DBN structure, enhancing its diagnostic performance and adaptability. Furthermore, they proposed a hierarchical object oriented DBN method for large-scale and complex systems [32], which was evaluated on an industrial building with 51 AHUs. Chen et al. [33] investigated a discrete DBN for cross-level faults in the AHU and the primary cooling subsystem of one campus building in the U.S. More systematically, Taal et al. [34] proposed the Four Symptoms and Three Faults (4S3F) reference architecture to guide DBN modeling for HVAC systems; it has been applied to demand-controlled ventilation in a Dutch university building and to full HVAC systems [35]. Moreover, Wang et al. [36] applied a DBN to the AHU heating component in a real office building using the same framework.

Sensor configurations play a significant role in influencing FDD performance. Most studies adopt a "top-down" approach, using feature selection to identify the important sensors for FDD in a sensor-rich laboratory setting or simulation software [20,37–40]. However, the idealized laboratory sensor setting may not accurately reflect the variability and constraints found in actual building environments. Real-world sensor configurations are influenced by factors such as budget, space limitations, and the specific needs of each building system [41,42], which can differ greatly from controlled lab conditions. Therefore, this study aims to better align with the practical realities of the building industry by adopting a "bottom-up" approach, which begins with a detailed investigation of existing sensor configurations in AHUs and then investigates their impact on FDD performance.

Most studies focus on specific types of AHUs with mixing box heat recovery system using the dataset from ASHRAE Project RP-1312 [5], which includes comprehensive field surveys, laboratory tests and performance evaluations of AHU under faulty conditions [5,43–45]. However, during the Covid-19 pandemic, air recirculation was prohibited to avoid high infection risk according to the guidelines of many professional HVAC associations [46]. In contrast, the European Union mandates the installation of alternative heat recovery devices, such as rotary and plate heat exchangers, rather than recirculation [47]. AHUs with heat recovery systems play a key role in providing safer and healthier ventilation, especially in the post-COVID era, and very few related FDD studies have been conducted. Nehasil et al. developed a rule-based FDD tool for ventilation units with heat recovery and validated it on several field AHUs [48]. Geoffroy et al. published an experimental dataset for an AHU with a fixed plate air-to-air heat exchanger and simulated faults to benchmark FDD algorithms, but did not develop a complete FDD scheme on top of this dataset [49]. Melgaard et al. introduced contextual KPIs and virtual sensors to track the performance of a rotary heat exchanger in an AHU during operation, focusing on performance monitoring rather than full fault diagnosis [50]. More recently, Lu et al. implemented a diagnostic Bayesian network for heat recovery ventilation units with rotary heat

recovery wheels (HRW) in a Dutch campus building, assuming a fixed Building Management System (BMS) sensor set and without systematically exploring alternative sensor configurations or fault-severity levels [51]. Collectively, these studies either focus on single-component performance or limited failure modes, lack a complete FDD framework for AHUs, or do not focus on model diagnosis performance under varying sensor configurations.Moreover, most studies focus on specific types of AHUs using the dataset from ASHRAE Project RP-1312 [5], which includes comprehensive field surveys, laboratory tests and performance evaluations of AHU under faulty conditions [5,43–45].

In terms of heat recovery, the AHUs investigated in ASHRAE Project RP-1312 rely on recirculated air through a mixing box [5]. But for improved energy efficiency, the European Union mandates the installation of heat recovery devices, such as rotary and plate heat exchangers, rather than recirculation [47]. Additionally, during the Covid-19 pandemic, air recirculation was prohibited to avoid high infection risk according to the guidelines of many professional HVAC associations [46]. Heat recovery devices can still be utilized, as long as the leakage is within acceptable limits, as Federation of European Heating, Ventilation and Air Conditioning Associations (REHVA) indicates that heat recovery device installation is justified for a ventilation system for leakage is below 5% [52].

This research therefore focuses on AHUs with HRW and extends previous work on DBN-based fault diagnosis by analyzing its real-world performance under diverse sensor configurations. To better align with the practical realities of the building industry, this study adopts a "bottom-up" approach, beginning with a detailed investigation of existing sensor configurations from both guidelines' recommendations and a practice survey of real-world AHUs in the Netherlands. Then three sensor configuration levels are defined: Limited, Standard, and Rich, ranging from the essential sensors required for basic AHU operation to advanced measurements for monitoring and optimization.The contribution of this paper can be summarized as follows:

- Defines three sensor configuration levels based on the investigation of recognized guidelines and practice survey.
- Develops a DBN framework, based on a real-world building's AHU equipped with a heat recovery wheel, and evaluates fault diagnosis performance across varying sensor configurations.
- Provides recommendations for sensor deployment and identifies key sensors in AHU systems for FDD.
- Investigates DBN's modeling practices.

The remainder of this paper is organized as follows. Section 2 outlines the research framework, including the DBN and the 4S3F reference model, which is a structured way to defined and link AHU symptoms and faults for AHU FDD. Section 3 describes the FDD sensor configurations from guidelines and the practice survey. Section 4 presents the case study, detailing the DBN structure, symptom and fault definitions, parameter modeling, and data collection, followed by model evaluation through Bayesian inference and fault isolation. Section 5 reports the results of symptom detection, case analyses across sensor configurations, and diagnostic sampling scenarios. Finally, Sections 6-7 provide the discussion and conclusion, highlighting key findings, limitations, and future research directions.

2. Methodology

2.1. Research framework

A systematic methodology was developed to investigate the impact of different sensor configurations on fault diagnosis performance within the AHU system. This methodology, illustrated in Fig. 1, encompasses the following key steps: (1) Sensor Configuration Survey, (2) FDD Modeling, (3)Experiment & Data Collection, (4) FDD Result Evaluation, and (5) Discussion.

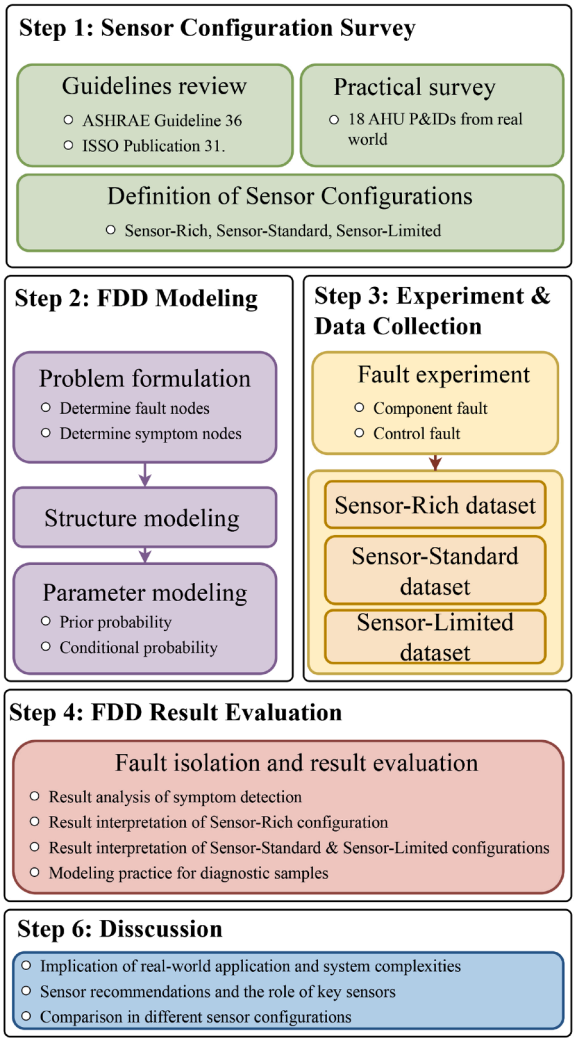


Fig. 1. Flowchart of the DBN-based fault diagnosis method in three sensor configurations based on recognized guidlines and practice survey.

- Step 1 (Section 3): Three sensor configurations (Rich, Standard, and Limited) were defined based on established guidelines, specifically ASHRAE Guideline 36 [53], which is widely used internationally, and ISSO Publication 31 from the Dutch Institute for the Study and Promotion of Research in Building Installations (ISSO) [54], relevant for the Dutch case study building. In addition, a practice survey of real-world buildings was conducted.
- Step 2 (Section 4), FDD modeling was developed. This process involved defining the model's key components: fault nodes, which represent potential system malfunctions (e.g., a stuck valve), and symptom nodes, which represent the observable data (e.g., sensor readings) used to detect those faults. These nodes were defined for three sensor configurations, fault-symptom relationships were established based on domain knowledge and expert input, and DBN parameters were modeled by defining prior and conditional probabilities.
- Step 3 (Section 4.5), An experiment was conducted in which faults were introduced into the AHU system and data was primarily collected from the Sensor-Rich configuration. Datasets for the other configurations were derived by reducing available sensors.
- Step 4 (Section 5), FDD results were evaluated, including fault isolation and interpretation of results in three sensor configurations.
- Step 5 (Section 6), the results were discussed, considering real-world application and system complexities, comparisons across sensor

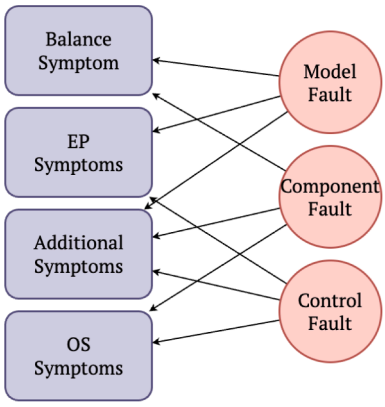


Fig. 2. 4S3F method illustrating relationships between symptoms and faults.

configurations, sensor recommendations and the role of key sensors, and modeling practices for diagnostic samples.

2.2. Diagnostic Bayesian network and 4S3F

This section introduces the FDD methodology for AHU, which includes DBN, the Noisy-OR model, and the 4S3F reference model.

A DBN is a probabilistic graphical model representing dependencies between fault and symptom nodes through a directed acyclic graph (DAG) [55]. Fault nodes indicate potential system faults (typically with two states: faulty and fault-free), while symptom nodes represent observable variables that respond to faults. The directed edges from the fault to the symptom nodes represent the causal relationships between them.

Posterior fault probabilities in the DBN are calculated using Bayes' theorem:

$$P(F_i|S) = \frac{P(S|F_i) \cdot P(F_i)}{P(S)} \quad (1)$$

This Bayesian inference framework enables the update of fault posterior probability as a new detection of symptoms.

Each node in the DBN is quantified by a conditional probability table (CPT), which specifies the probability of each node state for every possible combination of its parent-node states. For a binary symptom node with k binary fault parents, a full CPT requires 2^k probability parameters. In this study, the Noisy-OR model [56] is used to construct the CPTs for symptom nodes with multiple parent faults. By treating each fault as an independent cause, the model reduces the required parameters from 2^k to k. This efficiency makes it practical to parameterize the fault-symptom relationships within the proposed DBN.In this study, the Noisy-OR model [56] is introduced to reduce the exponential growth of parameters in conditional probability tables as the number of parent nodes increases.

For systematic and scalable AHU FDD development, the 4S3F reference model is adopted [27,29,34] as illustrated in Fig. 2. This structured two-stage model closely aligns with HVAC system diagrams and reflects the way HVAC systems and controls are designed [27].

In the first stage, 4S3F group symptoms into four categories: balance symptoms, which indicate deviations in energy, mass, or pressure balance within the system; energy performance (EP) symptoms, which reflect deviations in key performance metrics such as the coefficient of performance (COP) and are often derived from data-driven or model-based analyses; operational state (OS) symptoms, which capture abnormal system behavior, including a temperatures or flows and can be detected through historical data analysis or machine learning methods; and additional symptoms include external information sources such as maintenance logs and historical FDD output.

In the second stage, faults are classified into three types: faults in controls, faults in components, and model faults. Component faults originate from failures, degradation, or improper installation of system parts; control faults arise from setpoint errors, software issues, or problems in control algorithms; and model faults are related to inaccuracies or improper hypotheses in system modeling. In this study, only component and control faults are considered.

3. FDD sensor configurations in practice and guidelines

This section discusses sensor configurations for AHU FDD, first by examining the two recognized guidelines and then by performing a practice survey for Dutch office buildings. Based on the discrepancies found, three distinct configurations (Limited, Standard, and Rich) are defined to evaluate FDD performance. All sensors and measurement points referenced in the guidelines, practice survey, and defined sensor configurations are detailed in Table 1.

3.1. Guidelines

To evaluate the impact of sensor configurations on AHU FDD, this study uses ASHRAE Guideline 36, which has served as a foundational reference for AHU FDD and has been used in many studies [5,28,57], as well as the Dutch ISSO Publication 31, which is the standard guideline in the Netherlands. The list of recommended sensors by both guidelines can be found in the first columns of Table 1.Both documents specify sensor requirements to support energy efficiency, system stability, and FDD. A comparative analysis highlights the key findings below.

- ASHRAE Guideline 36 focuses on standardized control sequences for common HVAC systems to reduce energy use, enhance stability, and enable FDD, while ISSO Publication 31 emphasizes standardized measurement methods and optimal sensor placement for accurate data collection, system reliability and facilitation of FDD. Consequently, ASHRAE integrates FDD into advanced control logic, whereas ISSO structures its FDD framework around sensor classification for governance, monitoring, and evaluation.
- ASHRAE Guideline 36 focuses on the control and monitoring of the air distribution systems in air-side HVAC configurations, specifically multi-zone, single-zone, and dual-fan dual-duct Variable Air Volume Air Handling Unit (VAV). In contrast, ISSO Publication 31 covers both AHUs and hydronic systems, extending to complete climate systems and advanced heat recovery, and includes distribution, generation, and user modules. This broader scope requires additional monitoring and measurement for effective FDD.
- ISSO Publication 31 specifies a broader range of sensors and measurement points compared to ASHRAE Guideline 36, reflecting its inclusion of more diverse systems for FDD. While ASHRAE Guideline 36 lists only the minimum required control sensors and classifies measurements by operating states, both guidelines require core measurements such as supply, return, mixed, and outdoor air temperature, as well as supply and return water temperatures. These form the basis of FDD processes. Both guidelines also emphasize control signals for fan speed, dampers, and coil valves. In addition, ISSO 31 recommends further monitoring sensors, including pressure differential sensors across filters, humidity sensors for different air streams, water and airflow sensors, and air quality sensors at the return side, enabling more comprehensive diagnostics and robust system monitoring.
- Both guidelines emphasize the importance of control signals for FDD, including those for supply and return fans, dampers, heat recovery wheels, coil valves, and speed-controlled pumps. Combined with sensor measurements, these control signals form a comprehensive foundation for monitoring, diagnosing, and optimizing AHU performance.

Table 1
Comparison of sensor recommendations between ASHRAE Guideline 36 and ISSO Publication 31. Numbers 1 to 18 refer to different P&IDs surveyed in this study. The sensor configurations are defined as Sensor-Rich, Sensor-Standard, and Sensor-Limited to categorize sensor availability.

Sensor	ASHRAE	ISSO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Limited	Standard	Rich
Outdoor air temperature	✓	✓	✓				✓			✓			✓	✓		✓	✓	✓	✓	✓		✓	✓
Outdoor air relative humidity		✓															✓	✓	✓	✓			✓
Preheated air temperature	✓	✓			✓									✓			✓	✓	✓			✓	✓
Preheated air relative humidity		✓												✓									
Supply air temperature	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Supply air temperature setpoint	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Supply air relative humidity		✓			✓		✓		✓	✓		✓	✓			✓	✓	✓	✓	✓		✓	✓
Return air temperature	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
Return air relative humidity		✓			✓		✓		✓	✓		✓	✓	✓		✓	✓	✓	✓	✓		✓	✓
Exhaust air temperature	✓	✓				✓	✓							✓		✓	✓	✓	✓	✓		✓	✓
Exhaust air relative humidity		✓										✓			✓	✓							
Supply coil water temperature	✓	✓												✓	✓		✓					✓	✓
Return coil water temperature	✓	✓					✓			✓	✓	✓	✓	✓			✓	✓	✓	✓		✓	✓
Pressure difference supply filter		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pressure difference return filter		✓			✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓			✓
Pressure difference supply fan		✓			✓							✓					✓						✓
Pressure difference return fan		✓										✓											
Supply air flow rate		✓		✓													✓						✓
Return air flow rate		✓																					
Coil valve control signal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Supply fan control signal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Return fan control signal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Supply damper control signal		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Return damper control signal		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Air quality sensor		✓				✓		✓		✓					✓		✓						✓
Supply air static pressure	✓	✓					✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Duct static pressure setpoint	✓	✓					✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Coil water flow		✓										✓		✓					✓				
Interior air temperature		✓	✓	✓			✓		✓	✓		✓	✓			✓	✓	✓	✓	✓		✓	✓
Interior air temperature setpoint		✓	✓	✓			✓		✓	✓		✓	✓			✓	✓	✓	✓	✓		✓	✓

3.2. Practice survey on AHU sensor configuration

A survey of Pipe and Instrumentation Diagrams (P&IDs)P&ID diagrams from 18 AHUs in 18 Dutch public buildings was conducted by firstly selecting P&IDs and then systematically recording the presence of each component, sensor, and measurement point. The occurrence of these elements was then counted and summarized in Table 1.

The survey shows notable differences in sensor configurations among the 18 AHUs. Essential sensors like supply air temperature and return air temperature, along with most control signals for valves, fans, and dampers, were present in at least 16 of the 18 units. Sensors for outdoor air temperature, air humidity, coil water temperature, and select setpoints were moderately common, appearing in 9 to 15 AHUs. Conversely, advanced sensors for air quality, coil water flow, and fan pressure differences were found in 5 or fewer units, indicating less common installation. These sensor installation patterns align with findings from Gao and Berges [58].

14 out of 18 surveyed AHUs do not fully meet sensor recommendations from the two guidelines, as each unit is missing one or more required sensors. The absence of critical sensors in these AHUs, such as coil flow rates, supply and return water temperatures, air quality, and airflow, constrains the implementation of FDD method recommended by two guidelines, potentially leading to missed opportunities for energy savings and system performance improvements.

Discrepancies between guidelines and practice surveys arise from cost, system complexity, and lack of FDD standardization for sensor deployment in AHU. As a result, sensor installations in practice are largely focused on operational management rather than diagnostics or energy efficiency. Despite comprehensive guideline recommendations, practical implementation is often constrained by financial, technical, and logistical challenges. Therefore, it is crucial to investigate the impact of sensor configuration on practical FDD to better align guideline recommendations with real-world constraints and, oppositely, to showcase the positive effect of following to the guidelines.

3.3. Definition of sensor configurations

Based on practice surveys and the two referenced guidelines, this study defines three sensor configurations: Sensor-Limited, Sensor-Standard, and Sensor-Rich, as detailed in Table 1. Because the 18 AHUs from the practice survey and the two guidelines differ significantly, the three sensor configurations were extrapolated based on expert judgment.

The Sensor-Limited configuration represents a minimalist approach, containing only the essential sensors required for basic AHU operation and system safety, like cases 1 and 2. these sensors include supply air temperature, static pressure, and basic control commands. This configuration ensures fundamental system functionality, but lacks the capability for advanced diagnostics or energy optimization.

The Sensor-Standard configuration introduces additional sensors to enhance control accuracy and system feedback, like case 5. This configuration typically includes return air temperature, pre-heating air temperature, filter differential pressure, and supply water temperature sensors, supporting improved energy management and more comprehensive FDD, as recommended in previous studies [10].

The Sensor-Rich configuration extends beyond standard practice by incorporating a advanced sensors for detailed system monitoring and optimization, like case 14 and 15. Along with all sensors in the Standard configuration, it also includes room humidity, supply fan flow, and air quality sensors, in line with feature selection methods recommended by Li et al. [41] and the ASHRAE Project RP-1312 dataset. This configuration is expected to support advanced FDD strategies.

It should be noted that AHU sensor configurations vary, and some units may include most but not all sensors for a given definition. Thus, AHUs are classified by the nearest matching configuration, like Case 8. For example, an AHU with more sensors than Limited but not fully meeting Standard is classified as Standard. This approach accommodates practical variability in sensor deployment and enables systematic

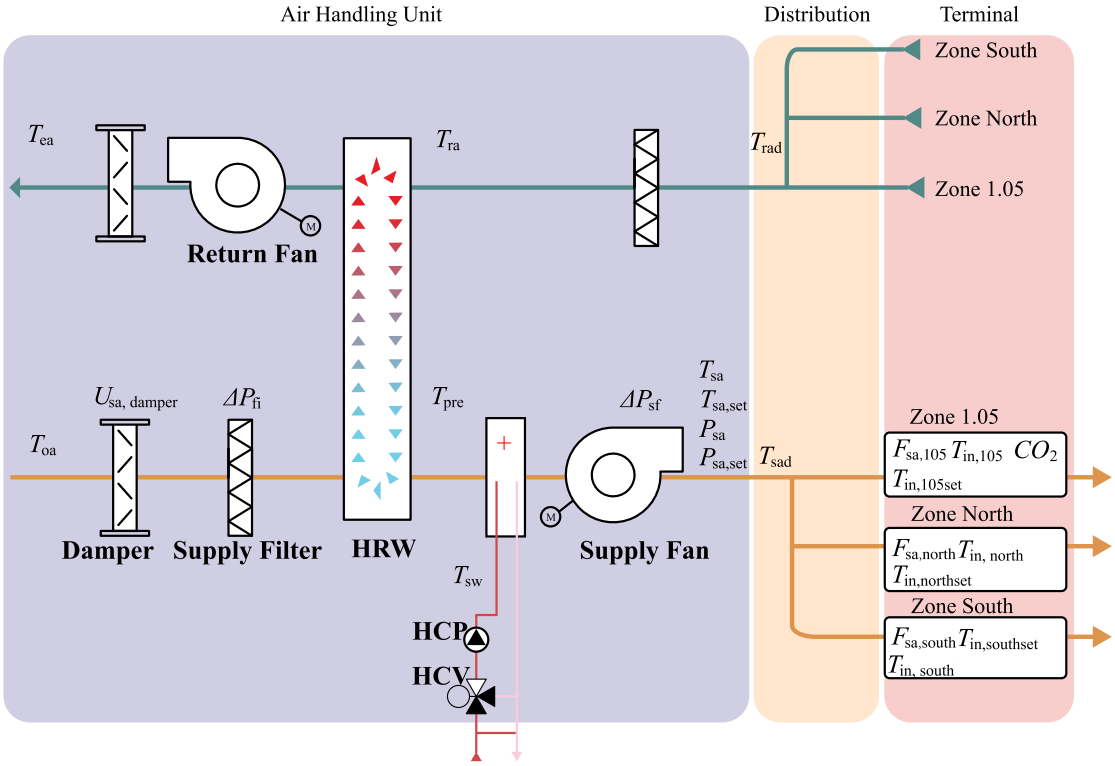


Fig. 3. A simplified version of process and instrumentation diagram (P& ID) diagram of the case study AHU system.

comparison of AHU capabilities from essential operation to advanced diagnostics.

4. Case study

This section describes the development of DBNs for FDD across different sensor configurations, including case description, DBN structure, assignment of symptom-fault relationships, data collection and model evaluation processes.

4.1. Case description

The case study building in the Netherlands is a typical Western European office building, equipped with comprehensive sensor and energy meter data similar to Sensor-Rich configuration. There was a possibility to conduct fault experiments in the case study building. The building is divided into three HVAC zones, each served by a central AHU with supply and return fans, filters, a heat recovery wheel, and a heating coil. Cooling is provided by three separate after-coolers for each zone, while heating uses a gas-fired boiler and zone hydronic system (radiators). Temperature control is managed by a master-slave Proportional-Integral-Derivative (PID) strategy. All three zones have an indoor temperature setpoint of 21.5 °C, and their controllers compute the needed supply air temperature setpoint based on the measured zone temperatures. The AHU controller sets the supply air temperature setpoint equal to the highest of these three demands and limits it to 26 °C, then controls the heating coil valve (HCV) so that the supply air temperature follows this setpoint. Temperature control is managed by a master-slave Proportional-Integral-Derivative (PID) strategy to adjust HCV based on the measurement of zones temperature. Supply air temperature and static pressure setpoints were dynamically adjusted throughout the day, with the system maintaining zones temperature at 21.5°C and supply static pressure at 250 Pa. Key measurement points and sensor locations are shown in Fig. 3.

4.2. DBN structure modeling

The DBN structure is developed by defining common faults in AHUs, followed by symptom definition and the connection between faults and symptoms.

4.2.1. Faults

Fault in this section representing common AHU issues [59], are named with initials F described in Table 2. The fault nodes correspond to key control faults of the AHU, including wrong supply air temperature setpoints (F_5), wrong supply air pressure setpoints (F_4), and component faults like a faulty heat recovery wheel (F_1), sensor biases (F_7 - F_{11}), and a supply fan stuck (F_2). Each fault node is defined with two possible states: faulty or normal, depending on the working state.

4.2.2. Symptoms

Symptom nodes are defined through logical analysis and the authors' expertise based on the 4S3F reference model, abbreviated as S, detailed in Table 3.

OS symptoms are defined as deviations in system state from the expected or intended condition, where the reference may be a control setpoint, historical or model-based norm. For example, supply air temperature and setpoint deviation (S_1) and supply air static pressure and setpoint deviation (S_2) both reflect how closely the system tracks its respective control targets. In addition to static setpoint comparisons, several OS symptoms are defined by the residual of prediction models, as summarized in Table 4. Several prediction models are classified as OS symptoms wherever the corresponding sensors are available, as summarized in Table 4. These prediction models, trained on historical fault-free data, establish a dynamic baseline for normal system operation. In these cases, the symptom is defined as a deviation from the model's predicted baseline that exceeds a predefined threshold, where the residual is calculated as the difference between the actual measured sensor value and this baseline The symptom is defined as the residual, which is the

Table 2
Fault nodes across sensor-rich, sensor-standard, and sensor-limited configurations.

No.	Fault	Abbreviation	State	State definition	Rich	Standard	Limited
F_1	Heat recovery wheel stuck	HRW-Stuck	Faulty	Abnormal working state	✓	✓	
			Normal	Correct working state			
F_2	Supply fan stuck	SF-Stuck	Faulty	Abnormal fan speed control	✓	✓	✓
			Normal	Correct fan speed control			
F_3	Heating coil valve stuck	HCV-Stuck	Faulty	Abnormal valve openness	✓	✓	✓
			Normal	Correct valve openness			
F_4	Supply air static pressure setpoint wrong	Pset-Wrong	Faulty	Incorrect setpoint	✓	✓	✓
			Normal	Correct setpoint			
F_5	Supply air temperature setpoint wrong	Tset-Wrong	Faulty	Incorrect setpoint	✓	✓	✓
			Normal	Correct setpoint			
F_6	Supply air temperature sensor bias	Tsa-Bias	Biased	T_{sa} sensor biased	✓	✓	
			Normal	T_{sa} is normal			
F_7	Supply air static pressure sensor bias	Psa-Bias	Biased	P_{sa} sensor biased	✓		
			Normal	P_{sa} is normal			
F_8	Pre-heating air temperature sensor bias	Tpre-Bias	Biased	T_{pre} sensor biased	✓	✓	
			Normal	T_{pre} is normal			
F_9	Return air temperature sensor bias	Tra-Bias	Biased	T_{ra} sensor biased	✓	✓	
			Normal	T_{ra} is normal			
F_{10}	Supply heating coil water temperature sensor bias	Tsw-Bias	Biased	T_{sw} sensor biased	✓	✓	
			Normal	T_{sw} is normal			
F_{11}	Supply air flow rate sensor bias	Fsa-Bias	Biased	F_{sa} sensor biased	✓		
			Normal	F_{sa} is normal			

Table 3
Symptom nodes across sensor-rich, sensor-standard, and sensor-limited configurations.

No.	Symptom	Abbreviation	State	Rule of state definition	Rich	Standard	Limited
S_1	Supply air temperature and setpoint deviation	Tset-Tsa-Diff	Present	$\lvert T_{sa} - T_{sa,set}\rvert > \varepsilon_T$	✓	✓	✓
			Absent	$\lvert T_{sa} - T_{sa,set}\rvert \le \varepsilon_T$			
S_2	Supply air static pressure and setpoint deviation	Pset-Psa-Diff	Present	$\lvert P_{sa} - P_{sa,set}\rvert > \varepsilon_P$	✓	✓	✓
			Absent	$\lvert P_{sa} - P_{sa,set}\rvert \le \varepsilon_P$			
S_3	Room air temperature and setpoint deviation	Tin-Tinset-Diff	Present	$\lvert T_{in} - T_{in,set}\rvert > \varepsilon_{Tin}$	✓	✓	
			Absent	$\lvert T_{in} - T_{in,set}\rvert \le \varepsilon_{Tin}$			
S_4	Heating coil valve position and prediction deviation	Uhc-Upred-Diff	Present	$\lvert U_{hc} - U_{hc,pred}\rvert > \varepsilon_{hc}$	✓		
			Absent	$\lvert U_{hc} - U_{hc,pred}\rvert \le \varepsilon_{hc}$			
S_5	Pre-heating air temperature high	Tpre-Tra,oa-Max	Present	$T_{pre} > \max(T_{ra}, T_{oa}) - \varepsilon_{Tpre}$	✓	✓	
			Absent	$T_{pre} \le \max(T_{ra}, T_{oa}) - \varepsilon_{Tpre}$			
S_6	Supply air temperature and prediction deviation	Tsa-Tpred-Diff	Present	$\lvert T_{sa} - T_{sa,pred}\rvert > \varepsilon_{Tsa}$	✓	✓	
			Absent	$\lvert T_{sa} - T_{sa,pred}\rvert \le \varepsilon_{Tsa}$			
S_7	CO_2 concentration high	CO2-Max	Present	$CO_2 > \varepsilon_{CO_2}$	✓		
			Absent	$CO_2 \le \varepsilon_{CO_2}$			
S_8	Heat recovery efficiency high	HRW-Effi-Max	Present	$\eta_{hrw} > \eta_{hrw,max}$	✓	✓	
			Absent	$\eta_{hrw} \le \eta_{hrw,max}$			
S_9	Supply air static pressure and prediction deviation	Psa-Ppred-Diff	Present	$\lvert P_{sa} - P_{sa,pred}\rvert > \varepsilon_{Psa}$	✓		
			Absent	$\lvert P_{sa} - P_{sa,pred}\rvert \le \varepsilon_{Psa}$			
S_{10}	Heat recovery efficiency low	HRW-Effi-Min	Present	$\eta_{sa} \le \eta_{sa,nm}$	✓	✓	
			Absent	$\eta_{sa} > \eta_{sa,nm}$			
S_{11}	Filter pressure drop low	Pfi-diff-Min	Present	$\Delta P_{sf} < 350\,\text{Pa} \wedge \Delta P_{fi} < \varepsilon_{fi}$	✓	✓	✓
			Absent	$\neg(\Delta P_{sf} < 350\,\text{Pa} \wedge \Delta P_{fi} < \varepsilon_{fi})$			
S_{12}	Heat recovery efficiency ratio deviation	HRW-Effi1-Effi2	Present	$\lvert \eta_{sa} - \eta_{eha}\rvert > \varepsilon_{s,e}$	✓	✓	
			Absent	$\lvert \eta_{sa} - \eta_{eha}\rvert \le \varepsilon_{s,e}$			
S_{13}	Nearby supply air temperature sensor deviation	Tsa-Tsa2-Diff	Present	$\lvert T_{sa} - T_{sad}\rvert > \varepsilon_{T2}$	✓		
			Absent	$\lvert T_{sa} - T_{sad}\rvert \le \varepsilon_{T2}$			
S_{14}	Supply pressure setpoint control high	Pset-Toa-Ctrl	Present	$P_{sa,set} > 250\,\text{Pa} \wedge T_{oa} \le \varepsilon_{oa}$	✓	✓	✓
			Absent	$\neg(P_{sa,set} > 250\,\text{Pa} \wedge T_{oa} \le \varepsilon_{oa})$			
S_{15}	Supply temperature setpoint control high	Tset-Toa-Ctrl	Present	$T_{sa,set} > 26\,^\circ\text{C} \wedge T_{oa} \le \varepsilon_{oa}$	✓	✓	✓
			Absent	$\neg(T_{sa,set} \ge 26\,^\circ\text{C} \wedge T_{oa} \le \varepsilon_{oa})$			
S_{16}	Supply air flowrate low	SF-Fsa-Min	Present	$\lvert N_{sf} - N_{sf,nm}\rvert \le \varepsilon_N \wedge F_{sa} < F_{sa,nm}$	✓		
			Absent	$\lvert N_{sf} - N_{sf,nm}\rvert \le \varepsilon_N \wedge F_{sa} \ge F_{sa,nm}$			
S_{17}	Nearby return air temperature sensor deviation	Tra-Tra2-Diff	Present	$\lvert T_{ra} - T_{rad}\rvert > \varepsilon_T$	✓		
			Absent	$\lvert T_{ra} - T_{rad}\rvert \le \varepsilon_T$			
S_{18}	Fan pressure drop and prediction deviation	Psf-Psfpred-Diff	Present	$\lvert \Delta P_{sf} - \Delta P_{sf,pred}\rvert > \varepsilon_{\Delta P_{sf}}$	✓		
			Absent	$\lvert \Delta P_{sf} - \Delta P_{sf,pred}\rvert \le \varepsilon_{\Delta P_{sf}}$			
S_{19}	Fan control signal and prediction deviation	Nsf-Nsfpred-Diff	Present	$\lvert N_{sf} - N_{sf,pred}\rvert > \varepsilon_{N_{sf}}$	✓		
			Absent	$\lvert N_{sf} - N_{sf,pred}\rvert \le \varepsilon_{N_{sf}}$			

difference between the actual measured sensor or control signal value and the value predicted by the model. A small residual indicates normal operation, while a large, sustained residual indicates symptom detection. For example, fan pressure drop and prediction deviation (S_{18}) and fan control signal and prediction deviation (S_{19}) adopt methods from Zhao et al. [30] and Li and Wen [60]. The supply air temperature and prediction deviation (S_6) utilizes a polynomial regression model following Torabi et al. [61]. Supply air pressure and prediction deviation (S_9) is defined using an XGBoost model with features selected by expert knowledge. The heating coil valve position and prediction

Table 4
Prediction models for the 5 symptoms based on fault-free data. The input variables in the formulations (e.g., T_{sa}, F_{sa}) represent sensor readings and control signals as defiend in Nomenclature.

Model	Symptom No.	Model Formulation	RMSE
1	S_4	$U_{\text{hc,pred}} = f(T_{\text{sa}}, T_{\text{oa}}, T_{\text{in, 105}}, U_{\text{sa,damper}}, F_{\text{sa,105}})$	7.5 %
2	S_6	$T_{\text{sa}} = 3.6926 + 0.5025\,T_{\text{pre}} + 0.3618\,T_{\text{sw}}$	0.4 °C
3	S_9	$P_{\text{sa,pred}} = f(\Delta P_{\text{sf}}, T_{\text{pre}}, T_{\text{sa}})$	4.64 Pa
4	S_{18}	$\Delta P_{\text{sf}} = -541.6087 + 0.1412\,F_{\text{sa}} - 3.7897 \times 10^{-6}\,F_{\text{sa}}^2$	19.71 Pa
5	S_{19}	$N_{\text{sf}} = -112.3844 + 0.0313\,F_{\text{sa}} - 1.1846 \times 10^{-6}\,F_{\text{sa}}^2$	2.01 %

deviation (S_4), as proposed by Zhao et al.[28], is modeled in this study using XGBoost using the top five predictive features from the training dataset.

EP symptoms are defined as deviations from nominal performance metrics, reflecting how well the system or its components achieve expected operation. For example, the calculation of the performance of the heat recovery wheel is assessed using the supply side (η_{sa}) and exhaust side (η_{ea}) efficiencies, calculated following Pecceu and Cailou [62] under the assumption of balanced flows: heat recovery efficiency low (S_{10}) follows the method of Pecceu and Cailou [62], under the assumption of balanced flows:

$$\eta_{sa} = \frac{T_{oa} - T_{pre}}{T_{oa} - T_{ra}} \tag{2}$$

$$\eta_{ea} = \frac{T_{ra} - T_{eha}}{T_{ra} - T_{oa}} \tag{3}$$

Where all temperature variables are defined in the Nomenclature. The symptom heat recovery efficiency low (S_{10}) is then defined. This binary symptom transitions from 'Absent' to 'Present' when the calculated supply side efficiency (η_{sa}) drops below a nominal threshold ($\eta_{\text{sa,nm}}$). Furthermore, the efficiency ratio deviation (S_{12}) compares supply side (η_{sa}) and exhausted side (η_{ea}) efficiencies to detect abnormalities.

Balance symptoms include mass, energy, or pressure balance, and consistency checks between nearby sensors. For instance, the symptom of nearby air temperature sensor deviation (S_{13}) detects when there's a notable difference between nearby sensors measuring the same variable.

In this study, no external data sources such as maintenance logs or prior FDD outputs are available, and thus no additional symptoms are defined.

4.2.3. Symptom threshold definition

A symptom is detected when a measured value deviates from its expected value by more than a specified threshold ϵ_i, as summarized in Table B.1. Thresholds are determined based on system control documentation (e.g., S_{14}, S_{15}), manufacturer specifications (e.g., S_{11}, S_8), and expert knowledge accounting for sensor location (e.g., S_{17}, S_{13}, S_{18}). For symptoms based on prediction-based models (e.g., S_4, S_6, S_9, S_{19}, S_{19}), thresholds are established using fault-free historical data, with σ calculated via the t-student approach on prediction residuals. A symptom is detected if the deviation is out of the upper and lower limits of $\pm 3^*\sigma$. All coefficients for prediction models in this study are trained on fault-free data from the case study building. 70% of this dataset (24,586 data points) are randomly selected for model training, and the remaining 30% is used as test data to validate the models.

4.2.4. Fault-symptom relationships related to DBNs structure

Faults and symptoms are connected using logical analysis, first principles, and the authors' expertise. For instance, in the case of the HRW, when the HRW is stuck at a low rotation speed or not operating, heat recovery efficiency (S_{10}) typically decreases, as the temperature difference across the HRW (T_{pre} and T_{oa}) becomes small. In addition to efficiency, HRW faults can increase heating demand, which is reflected by higher heating coil valve openness compared to predictions under normal conditions (S_4). Deviations also occur in fan pressure drop and control signals, as the decreased rotation speed affects duct static pressure and airflow through the wheel matrix, leading to measurable differences between actual and predicted values based on normal conditions (S_{18}, S_4) as described by O'Connor and Calautit [63,64]. Therefore these four symptoms are connected to HRW fault in DBN.

The ability to diagnose a fault depends on the availability of measured symptoms; as the number of available symptoms decreases, the number of diagnosable faults also declines, resulting in a progressive reduction in DBN complexity from Sensor-Rich to Sensor-Limited configurations, as shown in Fig. 4(a) to (c). This reduction in DBN complexity occurs for two primary reasons. First, when a key sensor is absent, any faults and symptom nodes that rely on it are removed from the network. For example, in the Sensor-Standard and Sensor-Limited configurations (see Fig. 4(b) and (c)), the absence of F_{sa} sensor results in the elimination of the Fsa-Bias fault (F_{11}) in the DBN, and related Symptom SF-Fsa-Min (S_{16}), because it does not relate to any other faults. Second, the removal of certain sensors reduces the number of symptoms connected to specific fault nodes. For instance, the Fault HRW-Stuck node in the Sensor-Standard configuration loses several associated symptoms, such as Uhc-Uhcpred-Diff (S_5), Psf-Psfpred-Diff (S_{18}), and Nsf-Nsfpred-Diff (S_{19}) leaving only HRW-Effi-Min (S_{10}) linked to this fault. In the Sensor-Limited configuration, the absence of all relevant sensors means that all symptoms related to the Fault HRW-Stuck node are removed, resulting in the complete exclusion of this fault node and its associated symptoms from the DBN structure. In this study, eleven, nine, and four faults could be theoretically diagnosed in the Sensor-Rich, Sensor-Standard, and Sensor-Limited configurations, respectively, based on the DBN structures.

4.3. Parameter modeling

Once the DBN structure is established, parameters are assigned by specifying fault node prior probabilities, which represent fault frequencies, and conditional probabilities, which define the likelihood of symptom detection given each fault.

4.3.1. Prior probability

The initial prior probability distribution in this study is adopted from expert assessments and reported values for similar systems. Specifically, the assignment for each fault node divides the state into a faulty state and a fault-free state. Following the approach used by Taal et al. [34], each fault node is assigned a prior probability of 0.05 for the faulty state and 0.95 for the fault-free state. These values reflect an assumption that there is a 5% likelihood for any specific fault to occur, while the component is expected to remain fault-free with a 95% probability. These prior probabilities are intentionally kept uniform for all faults, as obtaining detailed fault occurrence data for real application buildings is often limited.

4.3.2. Conditional probability

The conditional probabilities between fault and symptom nodes are defined using a Noisy-OR model, with influence levels (high, medium, low) set to probabilities of 0.9, 0.5, and 0.1, respectively, as shown in Table B.2. These values are based on expert knowledge. For example, a "high" relationship degree means that the presence of a fault is highly likely to cause the associated symptom.

4.4. DBN inference and fault isolation procedure

This section describes the BN inference and fault isolation of the FDD model. In BN inference for fault diagnosis, the process generally consists of four key steps: (1) HVAC data collection from sensors, (2) symptom detection based on measured data, (3) DBN inference using detected symptoms as input, and (4) posterior probability calculation for potential faults [29]. The approach to diagnostic sampling between

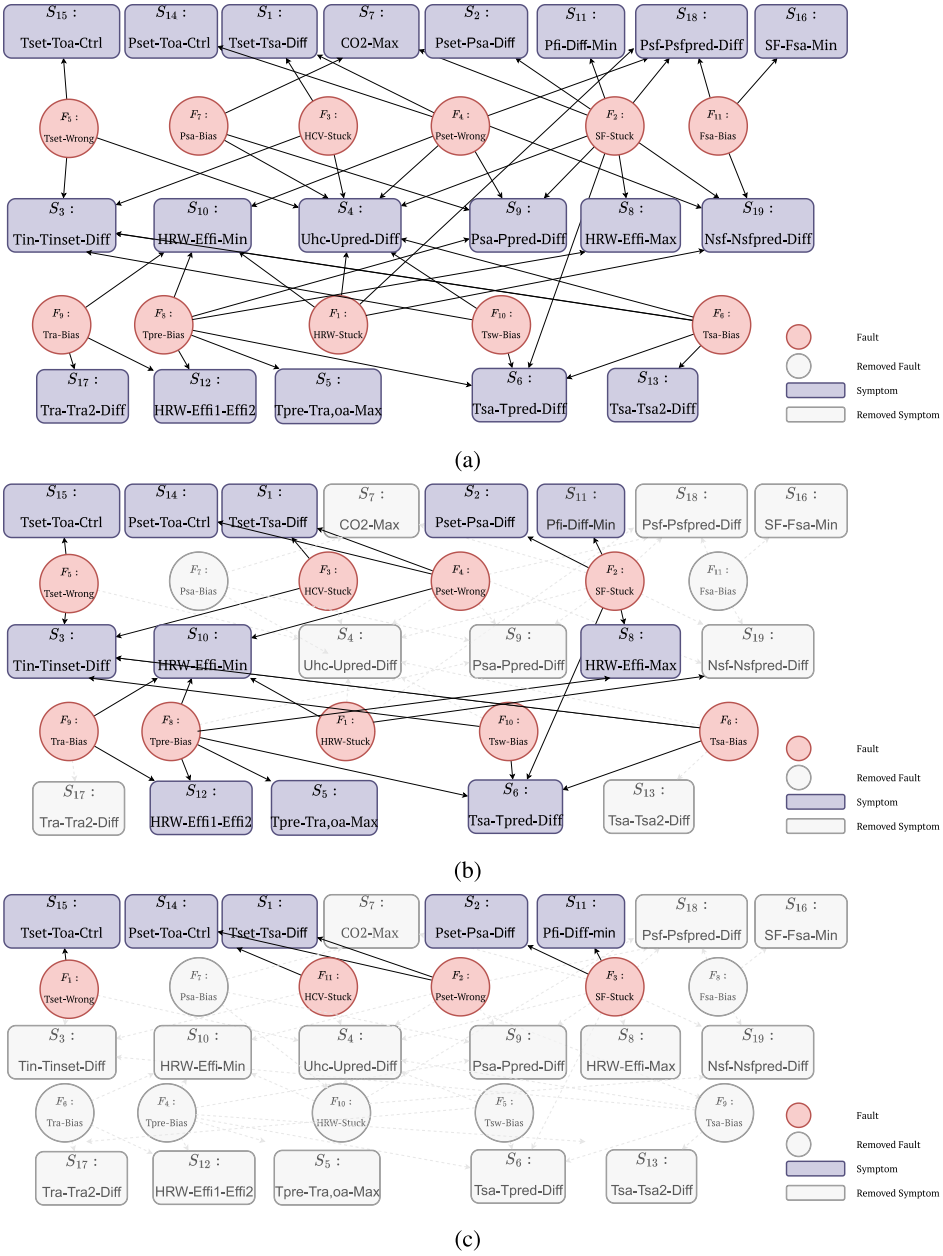


Fig. 4. DBN structures under three sensor configurations: (a) Sensor-Rich, where all symptoms are available; (b) Sensor-Standard, with a reduced symptoms and faults; and (c) Sensor-Limited, with limited symptoms and faults.

these steps varies in the literature, affecting how input data are aggregated and analyzed [27]. Three diagnostic sampling strategies are commonly used in BN-based FDD (Fig. 5). Scenario 1 aggregates each sensor time series over the evaluation window before symptom detection and runs one DBN inference per case. Scenario 2 detects symptoms at each time step, then aggregates their occurrence rates over the evaluation window and performs one DBN inference per case [65]. Scenario 3 runs DBN inference at every time step and then aggregates the resulting fault posterior probabilities [31,33,35]. This study adopts Scenario 2 because faults were maintained over the working day, and the diagnosis is intended to reflect aggregated daily symptom patterns rather than per-step variations [66].Three general strategies are used in BN-based FDD, as illustrated in Fig. 5. These involve aggregating sensor data before detection (Scenario 1), applying symptom detection before rate aggregation (Scenario 2) [65], and aggregating posterior probabilities after time-step inference (Scenario 3) [31,33,35]. In this study, Scenario 2 was adopted because it aligns with our experimental goal of reflecting aggregated daily symptom patterns, and its performance is further explored against Scenario 1 in Section 5.5.

The evaluation process proceeds as follows. Initially, "symptom present rates" were calculated. Methods for converting symptom detections into DBN inputs vary across FDD studies, depending on system dynamics and specific symptom definitions [27]. In this study, for each symptom, the rate is defined as the proportion of detected samples over the total number of samples in the evaluation window for a single experimental day.For each monitored symptom, this rate represents the fraction of time steps in the dataset where the symptom was detected as present. These rates were converted to a single binary states (present or absent) using a 0.5 threshold. This value serves as a hard threshold to detect persistent abnormalities within a fixed evaluation window. By applying this "majority rule," the method ensures that a symptom is considered detected for the DBN only if the symptom was present in more

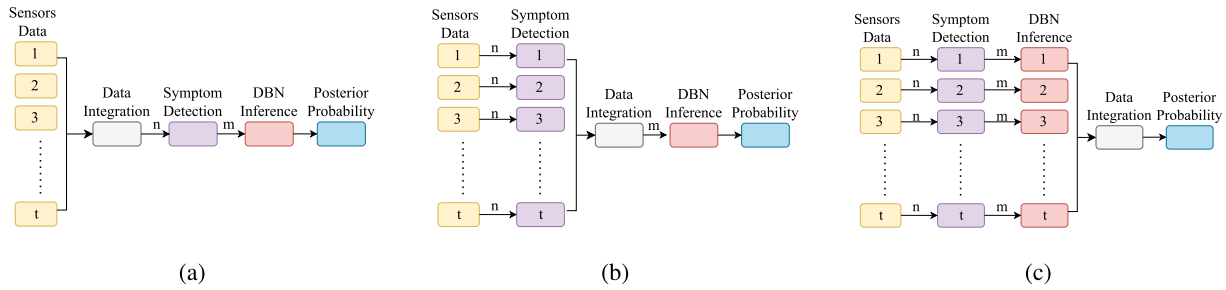


Fig. 5. Illustration of the three diagnostic sampling strategies: (a) Scenario 1-sensor aggregation, (b) Scenario 2-symptom aggregation, and (c) Scenario 3-posterior probability aggregation. Key parameters: n = cost of symptom extraction per time step, m = cost of a single BN inference, t = number of time steps per case.

Table 5
Summary of fault experiments. Cases 1a-1d correspond to F_1, Cases 2a-2b to F_2, Cases 3a-3b to F_3, and Cases 4-11 to F_4-F_{11} as defined in Table 2. Case 12 represents the fault-free day.

Case	Fault Description	Abbreviation	Experiment Scheme	Date
Case 1a	Heat recovery wheel stuck	HRW-Stuck-0	Fixing the heat recovery wheel rotation speed at 0%	03/22/23
Case 1b	Heat recovery wheel stuck	HRW-Stuck-30	Fixing the heat recovery wheel rotation speed at 30%	11/09/23
Case 1c	Heat recovery wheel stuck	HRW-Stuck-50	Fixing the heat recovery wheel rotation speed at 50%	03/27/23
Case 1d	Heat recovery wheel stuck	HRW-Stuck-80	Fixing the heat recovery wheel rotation speed at 80%	03/08/23
Case 2a	Supply fan stuck	SF-Stuck-30	Supply fan speed fixed at 30% control signal unchanged.	02/07/23
Case 2b	Supply fan stuck	SF-Stuck-65	Supply fan speed fixed at 65% control signal unchanged.	03/28/23
Case 3a	Heating coil valve stuck	HCV-Stuck-100	Fixing the heating coil valve openness at 100%	02/28/23
Case 3b	Heating coil valve stuck	HCV-Stuck-75	Fixing the heating coil valve openness at 75%	03/17/23
Case 4	Supply air pressure setpoint wrong	Pset-Wrong-50	Changing the supply pressure setpoint higher than desired (50Pa)	03/13/23
Case 5	Supply air temperature setpoint wrong	Tset-Wrong-3	Changing the supply temperature setpoint higher than desired (3°C)	03/15/23
Case 6	Supply air temperature sensor bias	Tsa-Bias-3	Changing the temperature sensor module reading with a positive bias (3°C)	02/10/23
Case 7	Supply air pressure sensor bias	Psa-Bias-30	Changing the pressure sensor module reading with a positive bias (30Pa)	02/13/23
Case 8	Pre-heating temperature sensor bias	Tpre-Bias-3	Changing the temperature sensor module reading with a positive bias (3°C)	01/03/23
Case 9	Return air temperature sensor bias	Tra-Bias-3	Changing the temperature sensor module reading with a positive bias (3°C)	01/05/23
Case 10	Supply water temperature sensor bias	Tsw-Bias-3	Changing the airflow sensor module reading with a negative bias (1000m^3/h)	03/10/23
Case 11	Airflow rate sensor bias	Fsa-Bias-1000		01/09/23
Case 12	Normal day	NM	Fault free date	02/03/23

than 50% of the evaluation window. were then converted to a single binary states (present or absent) using a 0.5 threshold, meaning that only if the symptom was present in more than 50% of the samples, it was considered to be present. This binary symptom data was subsequently used as input for DBN inference to calculate the posterior probability for each potential fault. Fault classification relied on these posterior probabilities. A fault was classified as present if its probability surpassed a predefined 0.15 threshold, adopting the method of Chen et al. [33]. If multiple faults met this criterion, the fault with the highest posterior probability was selected, as it is assumed that only a single fault is present at a time.

4.5. Data collection

Data were collected from the AHU in the case study building on working days from 9:00 to 16:00 during the 2023-2024 heating season. Experiments covered November, December, January, February and March, including both fault-free and faulty operations. All measurement data were recorded at a one-minute interval. The fault-free dataset consists of 420 one-minute samples per working day and was obtained by removing all periods affected by injected faults. This dataset is used to train the prediction models' baseline. Faulty datasets for each experiment contain between 260 and 420 one-minute samples, depending

on the duration of the fault injection, and are used only for validating the DBN FDD result. Experiments covered January, February, March, and November, including both fault-free and faulty operations with one-minute interval measurements. Table 5 details the implemented faults, covering components such as HCV, HRW, and fans, as well as several sensor and setpoint control faults.

5. Result

The DBN was evaluated under three sensor configurations. First, the results of symptom detection are presented. Next, the FDD performance of the Sensor-Rich configuration is discussed using selected cases, followed by the results for the Sensor-Standard and Sensor-Limited configurations. Finally, the modeling practice includes a comparison of different sampling approaches.

5.1. Result of symptom detection

This section analyzes the performance of the proposed symptom set by examining the relationship between known faults and their corresponding detected symptoms.

5.1.1. Overview of the symptom detection

The performance of the symptom set is summarized in the symptom detection matrix presented in Fig. 6. These "symptom present rates" are fundamental to the diagnostic process, as they are converted into binary inputs for the DBN. As detailed in Section 4.4, a symptom is considered 'present' (input as 1) only if its detection rate exceeds the 0.5 threshold. Thus, only symptoms with a sustained presence influence the final diagnostic outcome. For all three sensor configurations of the DBN, the symptom present rates remained identical for available symptoms.

5.1.2. Analysis of fault-robust symptoms

Several symptoms showed high robustness for specific fault categories. For example, Symptom HRW-Effi-Min (S_{10}), which indicates low heat recovery efficiency, had present rates above 0.5 for all HRW-Stuck (F_1) faults. Similarly, Symptom Pset-Psa-Diff (S_2) showed strong robustness for Fan-Stuck faults, with present rates of 1.0 across all Fan-Stuck cases (F_2). In contrast, Symptom Uhc-Upred-Diff (S_4) was detected in only one of the four variants of Case 1, while Symptom Nsf-Nsfpred-Diff (S_{18}) was detected in three of the four cases.

5.1.3. Analysis of general symptoms

Some symptoms functioned as general indicators for diverse fault categories. The Symptom Tin-Tinset-Diff (S_3), for instance, indicates a failure to meet the interior air temperature setpoint. It was detected with present rates well above the 0.5 threshold for several distinct faults, including HRW-Stuck-50 (Case 1c, present rate = 0.81), HCV-Stuck-75 (Case 3b, present rate = 0.95), and Tset-Wrong-3 (Case 5, present rate = 0.91). Additionally, several prediction-based symptoms (S_4, S_6, S_9, S_{18}, S_{19}) acted as general indicators across sensor faults, as model training relied on multiple sensor inputs (see Fig. B.1). A fault in any of these input sensors could trigger the corresponding symptom. For example, the Symptom Tsa-Tpred-Diff (S_6) was predictably triggered by input sensor faults such as Fault Tpre-Bias-3 (Case 8), Tsw-Bias-3 (Case 10), and Tsa-Bias-3 (Case 6). Another non-intuitive case was the detection of Symptom Nsf-Nsfpred-Diff (S_{19}) during a Fault Tset-Wrong-3 (Case 5). Although the temperature setpoint is not a direct model input, this detection likely occurred because the BMS unexpectedly changed the fan control logic to meet the higher temperature setpoint, indirectly affecting the supply temperature control loops. While these general symptoms provide valuable evidence, they may also introduce uncertainty in BN inference without more specific indicators.

5.1.4. Analysis of undetected symptoms

The matrix reveals critical undetected symptoms (i.e., a rate below 0.5) . An example is the small sensor bias fault, Psa-Bias-30 (Case 7). While this fault did trigger the Symptom CO2-Max (S_7), it failed to trigger the other two symptoms linked to it in the BN model. Another important case is the non-detection of Symptom Tset-Tsa-Diff (S_1) for HCV-Stuck faults (Case 3a, 3b) indicating one-third of expected evidence was absent. These undetected symptoms may impact BN inference, likely resulting in false negatives in fault diagnosis.

5.2. Fault diagnosis result of DBN in sensor-rich configuration

This section presents the results of the diagnosis for the Sensor-Rich configuration. In here, only Fault HRW-Stuck cases (Case 1a-1d), Tset-Wrong-3 (Case 5), and Tpre-Bias-3 ((Case 8)) are selected and described. The detail explanation of other cases are described in Appendix A. Fig. 7 presents the results of the fault diagnoses in diverse sensor configurations, in the Sensor-Rich configuration, 15 out of 17 cases are correctly diagnosed.

5.2.1. Fault HRW-stuck (Case 1a–1d)

As demonstrated in Fig. 7(a)–(d), Fault HRW-Stuck-0 (Case 1a) was accurately diagnosed with a posterior probability of 0.78. In contrast, Fault HRW-Stuck-30 (Case 1b) resulted in a false negative diagnosis with a posterior probability of 0.11. The remaining cases, Fault HRW-Stuck-50 and 80 (Cases 1c and 1d), were correctly diagnosed with a high posterior probability of 0.95. The primary diagnostic evidence was Symptom HRW-Effi-Min (S_{10}). To illustrate this, Fig. 8 presents the sensor readings for T_{pre}, T_{oa}, and T_{ra}, alongside both the nominal and calculated η of HRW.

In Case 1a, the complete failure of the HRW caused the heat recovery process to cease. Consequently, the T_{pre} and T_{oa} sensor readings became nearly identical, leading to η_{sa} approaching zero and triggering Symptom HRW-Effi-Min (S_{10}). Furthermore, the excessively low T_{pre} meant the AHU relied solely on the heating coil to heat the supply air. This resulted in difficulties in achieving the setpoint for T_{sa} ($T_{sa,set}$), thus detecting symptom Tset-Tsa-Diff (S_1), which is typically used to distinguish HCV faults.

In Cases 1b and 1c, with the HRW rotation speed fixed at 30% and 50% respectively, the η_{sa} remained below its nominal value, and Symptom HRW-Effi-Min (S_{10}) was detected. However, as shown in Fig. 6(d), Symptom Uhc-Upred-Diff (S_4) was not detected in these cases possibly because the general threshold (3 σ) used for this symptom is too wide, leading to high tolerance for the discrepancy between prediction and actual control signal. Additionally, the false negative diagnosis of Case 1b can be attributed to unexpected BMS control behavior. During this injected fault, an unrecorded drop in fan rotation speed, from the nominal 80% to between 50% and 70%, was observed. This unexpected operational shift likely altered the system's dynamic baseline, preventing the activation of other expected symptoms and leading to the misdiagnosis.

In Case 1d, the HRW rotation speed was fixed at 80%, which is close to the nominal operating condition. This resulted in Symptom HRW-Effi-Min (S_{10}) not being consistently detected. Nevertheless, as shown in Fig. 8(d), Symptom HRW-Effi-Min (S_{10}) was detected in more than 50% of the time steps throughout Case 1d, and therefore, it was still considered detected.

5.2.2. Fault Tset-Wrong-3 (Case 5)

In Case 5, as shown in Fig. 7(j), the fault was misdiagnosed as Fault Fsa-bias, with a posterior probability of 0.57, while the actual fault, Fault Tset-Wrong, was incorrectly assigned a low probability of 0.03. In this case, only one of three indicators, Symptom Tin-Tinset-Diff (S_3), was detected. Symptom Tset-Toa-Ctrl (S_{15}) was detected in less than 50% of the time steps as shown in Fig. 9(a). This is because, despite the positive bias, the dynamic $T_{sa,set}$ gradually dropped below the symptom detection threshold during the day.This may be because this symptom

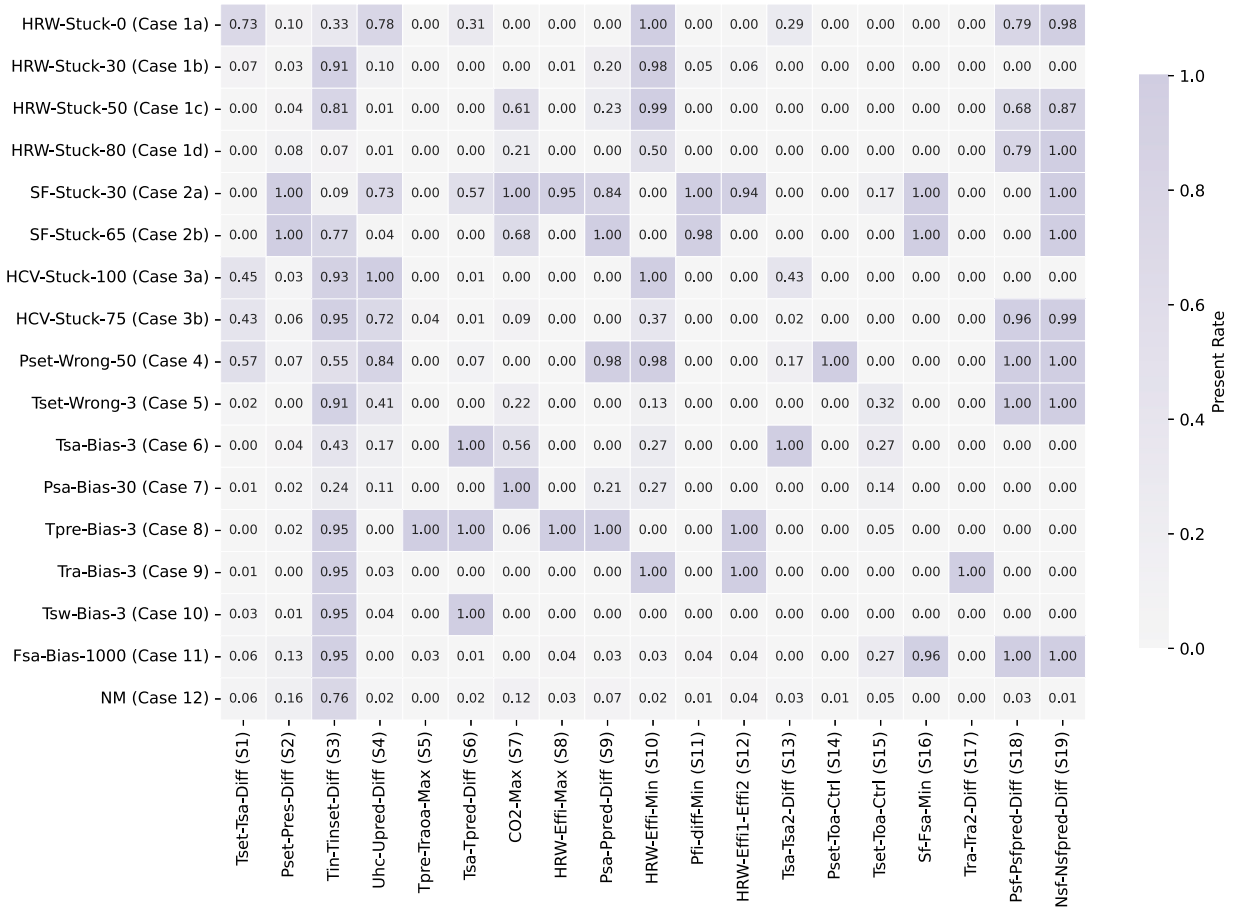


Fig. 6. Heatmap of the symptom present rates for the three sensor configurations. Each cell represents the rate at which a symptom (x-axis) is present for a given fault (y-axis).

is related to the BMS $T_{sa,set}$ control strategy. When outdoor temperature is increased during the day and the interior temperature is satisfied, $T_{sa,set}$ will decrease leading Symptom Tset-Toa-Ctrl (S_{15}) remains undetected. Additionally, the prediction-based symptom Uhc-UPred-Diff (S_4) was also undetected, as shown in Fig. 6, possibly due to the same wide threshold reason as described in Section 5.2.1 Case 1b.

5.2.3. Fault tpre-bias-3 (Case 8)

As shown in Fig. 7(m), Case 8 was correctly diagnosed as Fault Tpre-Bias with a posterior probability of 1.0. The Symptom HRW-Effi-Max (S_8) and HRW-Effi1-Effi2 (S_{12}) were both detected. Since T_{pre} is involved in the calculation of these two symptoms, a positive bias in T_{pre} increases η_{sa}, making it more likely to exceed the nominal value and trigger the detection of Symptom HRW-Effi-Max (S_8). It also creates a deviation between η_{sa} and η_{eha}, as shown in Fig. 9(b). Moreover, the Symptom Tpre-Traoa-Max (S_5) was also detected. As shown in Fig. 9(b), due to the bias, T_{pre} becomes higher than T_{ra}, violating the expected energy balance in the heat recovery wheel. Additionally, prediction-based symptom Tsa-Tpred-Diff (S_6) exceeded the present rate over 0.5, as shown in Fig. 6 due to T_{pre} function in model training as input .

5.3. Fault diagnostic result of DBN in sensor-standard configurations

This section presents the diagnostic performance of the DBN under Sensor-Standard configurations, which have reduced sensor availability. The results are compared to the Sensor-Rich configuration, referencing overall fault diagnosis counts from Fig. 7. In the Sensor-Standard configuration, with fewer available sensors, the DBN correctly diagnosed 9 out of the 17 fault cases. The detailed diagnosis results are as follows.

The reduction in sensors led to specific misdiagnoses. Fault HCV-Stuck-100 (Case 3a), depicted in Fig. 7(g), was misdiagnosed as Fault HRW-Stuck with a posterior probability of 0.46. This misdiagnosis happened because the missing Symptom Uhc-UPred-Diff (S_4) led the DBN to depend on the undetected Symptom Tset-Tsa-Diff (S_{15}). Meanwhile, Symptom HRW-Effi-Min (S_{10}) was detected due to the BMS's protective control mechanism, which reduced the HRW rotation speed in response to the HCV fault, misleading the DBN. A mutual false positive diagnosis occurred between Fault Tsa-Bias-3 (Case 6) and Tsw-Bias-3 (Case 10); each was incorrectly diagnosed as the other. This confusion arose from the lack of Symptoms Tsa-Tsa2-Diff (S_{13}) and Uhc-Upred-Diff (S_4), causing both faults to display identical evidential symptoms (Tsa-Tpred-Diff and Tin-Tinset-Diff) and similar conditional probability relationships in the DBN. Additionally, a normal operating day (Case 12) was falsely diagnosed as Fault Tset-Wrong with a posterior probability of 0.19, likely due to the absence of Symptom Uhc-Upred-Diff (S_4).

Some faults showed false negative. Fault Psa-Bias-30 (Case 6) was a false negative as its associated symptoms were absent in this DBN model configuration. Similarly, Fault Fsa-Bias-1000 (Case 11) went undiagnosed due to the exclusion of the relevant airflow sensor (F_{sa}) from this configuration.

Notably, Fault HRW-Stuck-30 (Case 1b) was correctly diagnosed with a posterior probability of 0.46. This relied on Symptom HRW-Effi-Min (S_{10}), which became crucial due to the lack of other symptoms in the Sensor-Rich configuration. Interestingly, Fault Tset-Wrong-3 (Case 5), previously a false negative in the Sensor-Rich configuration, was correctly diagnosed here with a posterior probability of 0.19. This improvement occurred because removing Symptoms Uhc-Upred-Diff (S_4), Nsf-Nsfpred-Diff (S_{19}), and Psf-Psfpred-Diff (S_{18}) due to sensor reductions

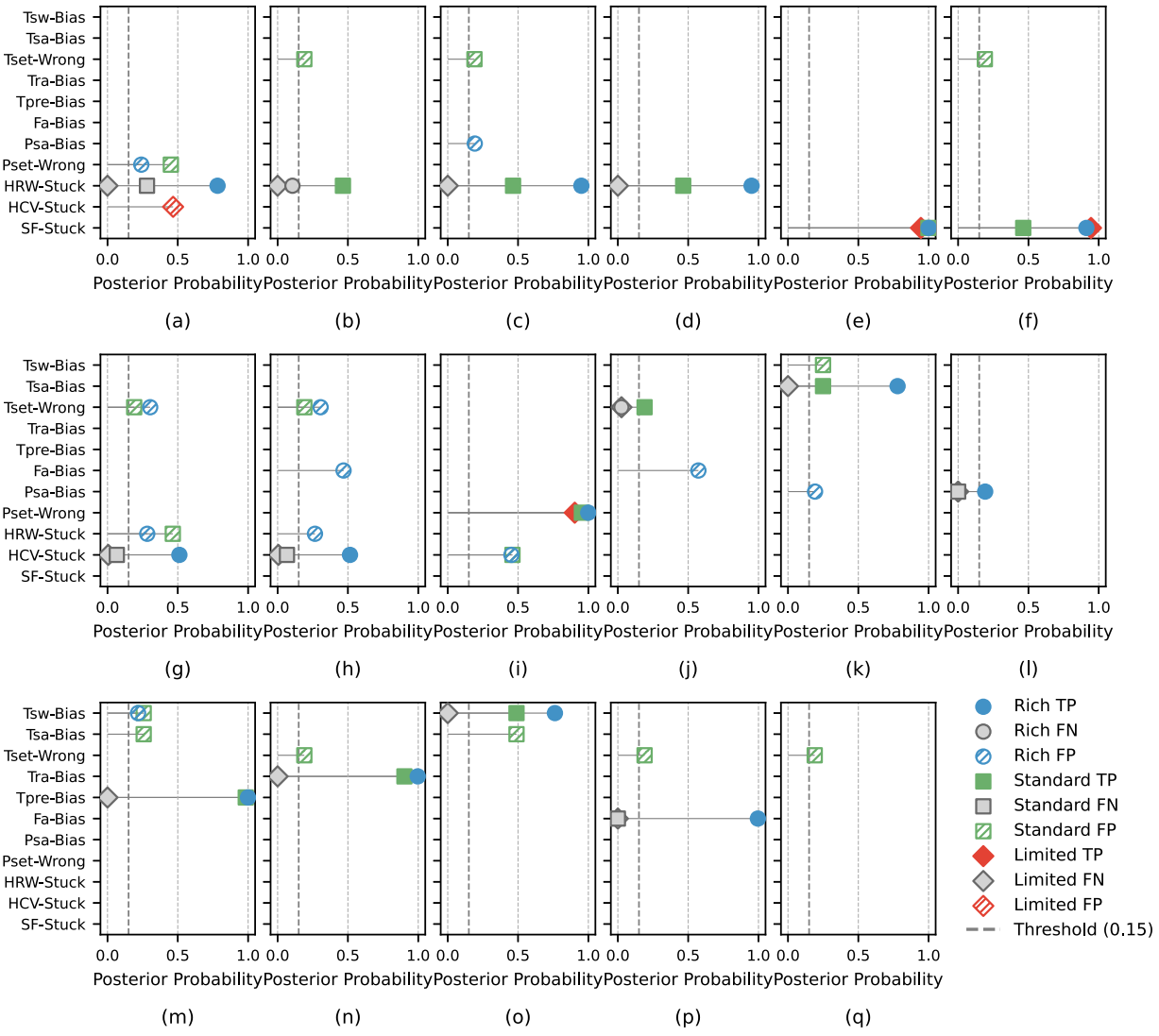


Fig. 7. Diagnostic performance of Rich, Standard, and Limited sensor configurations across 17 test cases. Each subplot shows posterior probabilities for a specific fault. A fault is classified as present if its probability exceeds the predefined 0.15 symptom detection threshold (dashed line). Markers indicate classification outcomes: True Positive (TP, solid color), False Negative (FN, solid gray), and False Positive (FP, hatched hollow). The specific fault case for each subplot is as follows: (a) HRW-Stuck-0, (b) HRW-Stuck-30, (c) HRW-Stuck-50, (d) HRW-Stuck-80, (e) Fan-Stuck-30, (f) Fan-Stuck-65, (g) HCV-Stuck-100, (h) HCV-Stuck-75, (i) Pset-Wrong, (j) Tset-Wrong, (k) Tsa-Bias, (l) Psa-Bias, (m) Tpre-Bias, (n) Tra-Bias, (o) Tsw-Bias, (p) Fa-Bias, and (q) NM (Normal).

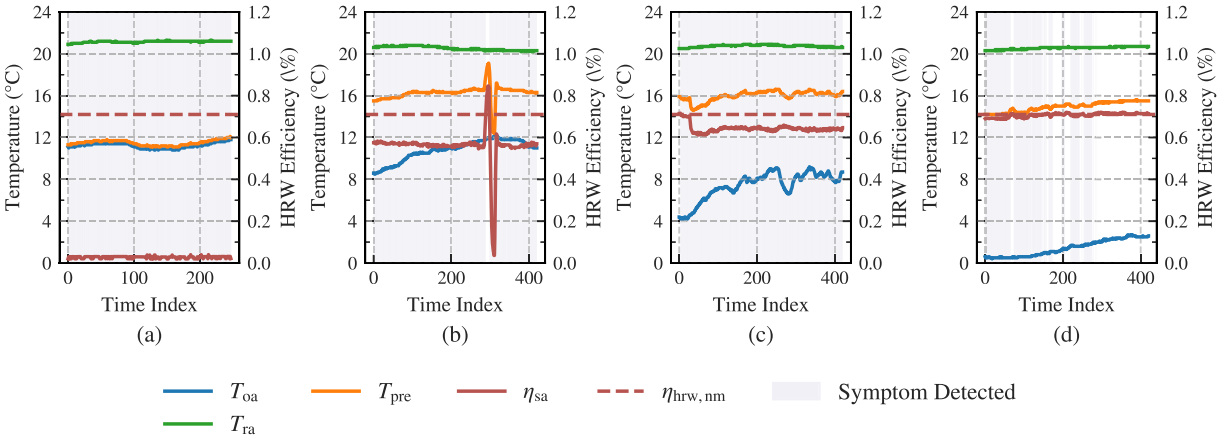


Fig. 8. Time series data illustrating key sensor readings (T_{oa}, T_{ra}, and T_{pre}) and the η_{sa} for Fault HRW-Stuck cases. (a) HRW-Stuck-0 (Case 1a). (b) HRW-Stuck-30 (Case 1b). (c) HRW-Stuck-50 (Case 1c). (d) HRW-Stuck-80 (Case 1d).

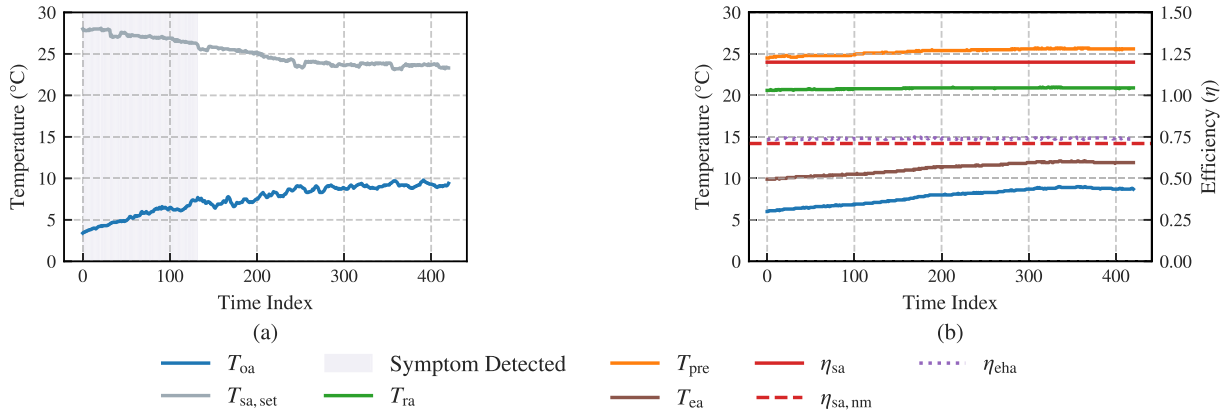


Fig. 9. Time series data illustrating key sensor readings (T_{oa}, T_{ra}, T_{ea}, T_{pre}, alongside the η_{sa}, and η_{eha}) values for (a) Fault Tset-Wrong (Case 5). (b) Tpre-Bias-3 (Case 8).

prevented conflicting diagnoses with other faults, allowing the DBN to assign a higher posterior probability to Fault Tset-Wrong.

5.4. Fault diagnostic result of DBN in Sensor-Limited configuration

Diagnostic capabilities were further reduced in the Sensor-Limited configuration, with only 4 out of 17 faults correctly diagnosed.

A case of misdiagnosis involved Fault HRW-Stuck-0 (Case 1a), falsely classified as Fault HCV-Stuck with a posterior probability of 0.47. This error arose because the sole symptom for Fault HCV-Stuck, Symptom Tset-Tsa-Diff (S_1), was detected in the Fault HRW-Stuck-0 case.

A significant number of faults went undiagnosed. All HRW-Stuck fault cases (Cases 1a-1d) were undiagnosable because all HRW-specific symptoms had been removed from this configuration. Additionally, other sensor-related faults were undiagnosable due to missing relevant sensors and symptoms. Specifically, Fault Tset-Wrong-3 (Case 5) was a false negative with a low posterior probability of 0.03. This was due to the removal of Symptom Tin-Tinset-Diff (S_3), leaving only Symptom Tset-Toa-Ctrl (S_{15}), which wasn't detected in this scenario.

Despite the significant reduction in sensors, Fault SF-Stuck (Cases 2a and 2b) was correctly diagnosed with a high posterior probability of 0.95 (Fig. 7(e), (f)). This accuracy was achievable because two key symptoms, Pset-Psa-Diff (S_2) and Pfi-Diff-Min (S_{11}), remained detectable even with limited sensors. Fault Pset-Wrong (Case 4) was also correctly diagnosed with a posterior probability of 0.90 (Fig. 7(i)), as its two remaining indicators, Symptoms Pset-Toa-Ctrl (S_{14}) and Tset-Tsa-Diff (S_1), were successfully detected. Notably, in contrast to the Sensor-Standard configuration, a normal operating day (Case 12) was correctly diagnosed as fault-free. This improvement resulted from removing Symptom Tin-Tinset-Diff (S_3), which previously caused a false positive in the Sensor-Standard configuration.

5.5. Modeling practice for diagnostic samples

Following the BN inference and fault isolation process in Section 4.4, this section evaluates how different approaches to diagnostic sampling affect the accuracy and computational complexity of fault detection.

In terms of accuracy, as shown in Table 6, Scenarios 2 and 3 achieved the highest diagnostic accuracy, each correctly identifying 15 out of 17 faults. Scenario 1 identified 14, performing less effectively on symptoms with transient information due to sensor data averaging; for example, Fault HCV-Stuck-75 (Case 3b) was misdiagnosed as Fsa-Bias because transient deviations in $U_{hc,pred}$ were lost. Scenario 3, while highly accurate, was sometimes overly sensitive to temporal fluctuations; for instance, in Fault HRW-Stuck-80 (Case 1d), it also indicated a high probability for Fsa-Bias during short periods when HRW efficiency appeared nominal, leading to temporary misclassifications.

Table 6
Comparison of diagnostic sampling strategies in terms of diagnostic accuracy, computational complexity, and execution time. Key parameters: n = cost of symptom extraction per time step, m = cost of a single BN inference, t = number of time steps per case.

Scenario	Correct Diagnosis	Time Complexity	Execution Time (s)
Scenario 1	14	$\mathcal{O}(n+m)$	0.2912
Scenario 2	15	$\mathcal{O}(t \cdot n+m)$	0.8331
Scenario 3	15	$\mathcal{O}(t \cdot (n+m))$	86.6769

In terms of computational complexity, Scenarios 1 and 2 had low runtimes of 0.29 and 0.83 s, while Scenario 3 required 86.68 s, as summarized in Table 6. Scenario 1 is the simplest, with a single inference; Scenario 2, though scaling with the number of time steps, remains moderate in complexity; Scenario 3 is the most complex due to performing inference at every time step. Notably, in this study, the cost of a single BN inference is represented as m for simplicity, although the actual complexity depends on the network structure, including the number of nodes, factor sizes, and the joint domain of the conditional probability tables. All experiments were run on a MacBook Pro (Apple M2 Pro, 12-core CPU, 16 GB memory, MacOS Sequoia 15.5). Posterior probability results are shown in Fig. B.2

In summary, Scenario 3 preserves detailed temporal information and achieves high accuracy, making it effective for detecting symptoms with transient information but computationally intensive. Scenario 1 is the least accurate due to aggregated input but is suitable for resource-constrained applications. Scenario 2 provides a balanced trade-off, combining high diagnostic accuracy and moderate computational cost, and is therefore recommended in this study.

6. Discussion

This section discusses the findings regarding the practical application of DBNs for FDD in AHUs across diverse sensor configurations, focusing on three key aspects: implication of real-world application and system complexities, comparison in different sensor configurations, sensor recommendations and the role of key sensors. The key findings and implications of this research are detailed below.

6.1. Implications of real-world application and system complexities

Unlike many studies assuming fixed sensor configuration in laboratory [28], simulation environments [67,68], or real buildings [69] this work addresses the challenge of sensor variability in real-world. By integrating guidelines with a practical survey, this study developed an adaptable DBN to evaluate FDD performance across varied sensor

configurations, identifying 15 of 17 faults in the Sensor-Rich configuration. However, this approach also revealed several practical challenges:

- Unexpected BMS control behavior (e.g., reducing HRW rotation speed as protection when a heating coil valve is stuck at high openness) and unpredictable operational factors (such as occupant behavior, weather, or undocumented maintenance) can lead to symptoms being missed or falsely detected, and therefore it remains an underexplored area of research [70]. For example, Symptom Nsf-Nsfpred-Diff (S_{19}) and CO2-Max (S_7) was false positively detected in Fault Tset-Wrong-3 (Case 5) and Tsa-Bias03 (Cases 6) respectively.
- For a number of symptoms (see Table 4), detection relied on prediction models applying fixed statistical thresholds, like $\pm 3\sigma$. Although this is common in laboratory studies [28,30], it can hinder symptom detection in real-world settings, as certain symptoms were undetected due to overly wide thresholds. For example, Symptom Uhc-UPred-Diff (S_4) was not detected in Fault Tsa-Bias-3 (Case 6). This suggests that fixed boundaries may reduce diagnostic accuracy in practical applications.

6.2. Comparison in different sensor configurations

Overall, the Sensor-Rich configuration achieves the highest number of correctly diagnosed faults, but the comparison across individual cases shows that more sensors do not always lead to better diagnoses. Although additional sensors generally improve diagnosis accuracy by providing more symptoms to the DBN, this study reveals that increasing the number of symptoms does not always guarantee improved results. For example, faults such as HRW-Stuck-30 (Case 1b) and Tset-Wrong-3 (Case 5) were correctly diagnosed with the Sensor-Standard configuration, but misdiagnosed in the Sensor-Rich configuration. These different results may be due to the removed symptoms Uhc-UPred-Diff (S_4), Nsf-Nsfpred-Diff (S_{19}), and Psf-Psfpred-Diff (S_{18}). These symptoms unexpectedlyunexpetctly detected or undetectedundetectin above cases mislead the BN inference, resulting in false positive or negative diagnoses. This finding highlights the potential for redundant or confounding information in DBNs to sometimes reduce diagnostic accuracy, aligning with the suggestion from Li et al. [31] that improper prior fault-symptom relationships should be removed to improve DBN's accuracy.

6.3. Sensor recommendations and the role of key sensors

Results in Fig. 7 indicate that the Sensor-Standard configuration supports basic FDD for AHUs with heat recovery, broadly consistent with the measurement points recommended by Schein et al. [10]. The importance of several sensors is evident from the number of faults they are associated with, such as outdoor air temperature (T_{oa}, 10 faults), supply coil water temperature (T_{sw}, 9 faults), and pre-heating air temperature (T_{pre}, 8 faults). From Sensor-Limited to Standard, adding T_{pre} and T_{sw} enables HRW efficiency and prediction-based symptoms (e.g. S_6: Tsa-Tpred-Diff), which allows correct diagnosis of the HRW fault (F_1), the supply temperature setpoint fault (F_5), and sensor faults such as Tsa-Bias (F_6), Tpre-Bias (F_8) and Tra-Bias (F_9). From Standard to Rich, additional supply water temperature (T_{sw}), airflow (F_{sa}) and fan pressure difference (ΔP_{sf}) sensors provide related symptoms that successfully diagnosed HCV-Stuck (F_3) and separated Tsw-Bias (F_{10}) from Fsa-Bias (F_{11}). Finally, adding CO_2 concentration provides the CO_2-Max symptom (S_{10}), which in combine with air flow sensors to identify Psa-Bias (F_7) as a pressure related fault rather than a fan fault. By linking these key measurements to more fault nodes in the DBN, the model gains stronger and more diverse symptom evidence, which improves DBN's ability to diagnose faults have similar symptoms. Therefore, upgrading the standard sensor configuration to include these key measurement points is recommended for improved FDD capability.Results in Fig. 7 indicate that the Sensor-Standard configuration supports basic FDD for AHUs with heat recovery, in line with the measurement points recommended by Schein et al. [10]. The importance of several sensors is evident from the number of faults they are associated with, such as outdoor air temperature (T_{oa}, 10 faults), supply coil water temperature (T_{sw}, 9 faults), and pre-heating air temperature (T_{pre}, 8 faults). Key sensors, including airflow rate (F_{sa}), supply fan pressure difference (ΔP_{sf}), supply coil water temperature and flow, and CO_2 concentration, further enhance diagnostic accuracy and broaden the scope of fault detection, approaching Sensor-Rich performance.

6.4. Limitations and data quality

This study has several limitations that should be considered related to data quality, modeling assumptions, and practical implementation. First, the accuracy of sensor data is critical for FDD performance. Although the DBN showed ability to diagnose sensor biases, the lack of sensor calibration, possible possible drift, and missing data may have influenced the detection of certain symptoms, particularly for prediction-based models that rely on consistent and reliable input data. While quality checks and data cleaning procedures were applied, residual uncertainties may remain, especially in periods with unrecorded maintenance or sensor malfunctions. Second, prediction-based symptoms and specific symptom thresholds depend on fault-free data. This method requires high-quality historical labeled data. Third, the conditional and prior probabilities in the DBN were assigned mainly based on expert knowledge due to the lack of comprehensive data. While this approach enables model construction when fault data are limited, it also introduces subjectivity and may not fully capture all real-world dependencies, particularly in complex or poorly understood fault scenarios. Fourth, the modeling practices adopted in this work, including the structure of the DBN and data focus on a specific dutch office building, present limitations to generalizability. Because the DBN was developed and validated using data from a single AHU, there is a risk of overfitting. As a result, the DBN framework and findings are likely transferable to other AHUs with similar architectures but may not be universally applicable to other AHUs with different configurations or operational profiles.

6.5. Future research

Based on the discussions and limitations, several directions for future research are recommendedsuggested. Firstly, further investigation is needed into the inconsistent symptom detection observed for certain faults and the diagnosis of multiple simultaneous faults, which were not extensively covered in the present work. Secondly, more extensive labelled fault datasets are needed to enable quantitative comparative analyses with other FDD methods and to support cross-building validation of symptom thresholds, prediction models, conditional probabilities and DBN structures across different AHU layouts, sensor configurations, building types and climatic conditionsmethodologies for cross-building adaptation should be explored to optimize symptom detection thresholds and improve the transferability and accuracy of prediction models across different buildings and HVAC systems. Thirdly, given the DBN's reliance on expert knowledge for defining fault-symptom relationships and conditional probabilities, future work should focus on developing structured methodologies or automated techniques for model construction [71,72] and knowledge elicitationautomated techniques to streamline the modeling process and integrate it into Bayesian frameworks more robustly, thereby reducing potential biases and the complexity of this process. Finally, further research should investigate the application and adaptability of DBNs to diverse climatic zones, building types, and operational conditions.

7. Conclusion

This study investigated the effectiveness of a DBN for fault diagnosis in a real-world AHU with diverse sensor configurations. The developed

DBN framework provides a robust diagnostic approach for AHUs with heat recovery, establishes a baseline for evaluating the cost-effectiveness of sensor selectiondeployment, and offers a monitoring solution for future engineering implementations of AHU FDD. The key findings of this research are summarized as follows:

- **Discrepancy between guidelines and practice:** This study defines three sensor configurations (Rich, Standard, Limited) for AHU FDD to bridge the gap between two recognized guidelines and the limited sensor installations in real-world, based on analysis of ASHRAE Guideline 36 and ISSO Publication 31 and the practice survey of Dutch AHUs.
- **DBN framework for AHU with heat recovery:** The proposed DBN was successfully applied to a real-world AHU with a heat recovery wheel, demonstrating its capability to diagnose a common operational and component faults across different sensor configurations.
- **Impact of sensor availability on diagnosis:** Diagnostic accuracy generally improves with increased sensor availability, with the Sensor-Rich configuration yielding the highest accuracy. However, simply adding more sensors do not always guarantee a better diagnosis for all faults. Diagnostic performance depends on strategic sensor selection and effective DBN modeling practices in order to achieve robust diagnosis.due to unexpected system behavior or inaccuracies in prediction-based symptoms.
- **Sensor configuration recommendations and key sensors:** The Sensor-Standard configuration provides a practical FDD baseline, while additional sensors for airflow, fan pressure, and key temperatures significantly enhance diagnostic performance toward the Sensor-Rich configuration.
- **Effectiveness of diagnostic sampling practices:** The method of processing diagnostic samples significantly affects FDD accuracy and computational efficiency. symptom aggregation followed by a single BN inference (Scenario 2) achieved the best balance between diagnostic accuracy and computational cost, outperforming both sensor aggregation and posteriorposteirorprobability approaches.

CRediT authorship contribution statement

Ziao Wang: Writing – original draft, Visualization, Investigation, Conceptualization; **Chujie Lu:** Writing – review & editing, Supervision, Conceptualization, Investigation, Methodology; **Arjen Meijer:** Writing – review & editing, Supervision; **Shalika Walker:** Data curation; **Laure Itard:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Data availability

The data that has been used is confidential.

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.

Acknowledgments

This study is part of the Brains4Buildings project, sponsored by the Netherlands Enterprise Agency MOOI funding scheme (Grant number: MOOI32004), which is established by the Dutch Ministry of Economic Affairs and Climate Change and the Ministry of the Interior and Kingdom Relations and executed by RVO Netherlands Enterprise Agency. The work was also supported by Kropman Installatietechniek. Special thanks to John Verlaan (Services and Building Automation teams, Breda branch); Shalika Walker at Kropman; and Srinivasan Gopalan and Karzan Mohammed for conducting and facilitating the experiment. Additionally, the authors sincerely appreciate Dr. Arie Taal for his 4S3F FDD and HVAC expertise, which greatly contributed to the valuable discussions.

Appendix A. Rest diagnostic result of sensor-rich configuration

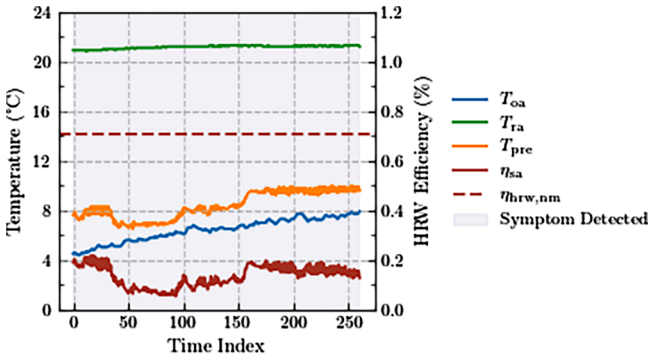


Fig. A.1. Time series data illustrating key sensor readings (T_{oa}, T_{ra}, and T_{pre}) and the η_{sa} for Fault HCV-Stuck-100 (Case 3a).

A.0.1. Fault SF-stuck (Case 2a–2b)

In cases 5 and 6, as illustrated in Fig. 7(e) and (f), Fault SF-Stuck was correctly diagnosed with posterior probabilities of 1.0 and 0.91, respectively. In Case 2a, the supply fan speed was fixed at 30%, resulting in a reduced supply airflow rate. Consequently, the supply air pressure (P_{sa}) failed to reach its setpoint, and both the filter pressure difference (P_{fi}) and the supply pressure (P_{sa}) decreased significantly. This triggered Symptom Pset-Psa-Diff and Pfi-Diff-Min. Moreover, the prediction-based symptoms Nsf-Nsfpred-Diff and Psa-Psapred-Diff were also detected, as shown in Fig. B.1(a) and (e). Furthermore, the lower fan speed reduced the volume of outdoor air supplied to the interior, limiting the dilution of CO_2 generated by occupants, which led to the detection of Symptom Co2-Max. Notably, Symptom HRW-Effi-Max and HRW-Effi1-Effi2 were also detected. This can likely be attributed to the decreased air velocity across the HRW at lower fan speeds, deviating from nominal operating conditions and allowing for excessive heat exchange between the exhaust and supply air streams.

In Case 2b, similar symptoms were detected. However, with the fan speed fixed at 65%, which was closer to normal operation, the HRW related symptoms (HRW-Effi-Max and HRW-Effi1-Effi2) were not detected.

A.0.2. Fault HCV-stuck (Case 3a–3b)

As shown in Fig. 7(g) and (h), the DBN correctly diagnosed Fault HCV-Stuck in Cases 7 and 8 with a posterior probability of 0.51. In Case 3a, two of the three primary symptoms, Symptom Uhc-UPred-Diff and Symptom Tin-Tinset-Diff, were detected, leading to the correct diagnosis. Notably, Symptom HRW-Effi-Min was unexpectedly detected, as illustrated in Fig. 6. When overheating occurs while the HCV remains stuck, the BMS might activate a protective control mechanism, reducing the HRW speed to lower the supply temperature to reach the temperature setpoint. As shown in Fig. A.1, the similarity between the T_{pre} and T_{oa} sensor readings supports this assumption. This mechanism may also explain why another primary symptom, Symptom Tset-Tsa-Diff, was detected in less than 50% of the time steps and therefore was not considered detected.

In Case 3b, similar to Case 3a, Symptoms Uhc-UPred-Diff and Tin-Tinset-Diff were detected, as shown in Fig. 6(h), leading to the correct diagnosis. Additionally, Symptom Tset-Tsa-Diff was detected in less than

50% of the time steps, meaning that it was not considered detected. This could be because when the HCV was stuck at 75% openness, it remained close to the nominal operational conditions.

A.0.3. Fault Pset-wrong (Case 4)

As shown in Fig. 7(i), Case 4 was correctly diagnosed as Fault Pset-Wrong with a posterior probability of 1.00. The expected operational state (OS) symptoms Pset-Toa-Ctrl, HRW-Effi-Min, and Tset-Tsa-Diff were detected. When the pressure setpoint is higher than expected, Symptom HRW-Effi-Min is consistently detected at every time step. This occurs because a higher pressure setpoint increases the fan speed. As the fan speed increases, air moves through the HRW more quickly, reducing the contact time between the HRW surface and the air, which in turn decreases the heat exchange efficiency. Additionally, the increased fan speed leads to an unstable supply temperature, causing the detection of Symptom Tset-Tsa-Diff as shown in Fig. 6. Moreover, two prediction symptoms, Uhc-UPred-Diff and Psa-Ppred-Diff, were also detected, as shown in Fig. B.1(d) and (e).

A.0.4. Fault Tsa-bias (Case 6)

As shown in Fig. 7(h), Case 6 was correctly diagnosed as Fault Tsa-Bias with a posterior probability of 0.78. Two of the three symptoms, Tsa-Tsa2-Diff and Tsa-Tpred-Diff, were detected, leading to the correct diagnosis. When T_{sa} is biased, the prediction Symptom Tsa-Tpred-Diff is expected to detect a deviation from the predicted T_{sa}, as shown in Fig. B.1(c). Similarly, Symptom Uhc-UPred-Diff is also expected to be detected, as T_{sa} is used as input during model training. However, it remained undetected, as shown in Fig. B.1(d).

A.0.5. Fault Psa-bias (Case 7)

As shown in Fig. 7(l), Case 7 was correctly diagnosed as Fault Psa-Bias with a posterior probability of 0.19. Only symptom CO2-Max was detected in this case. When P_{sa} is biased, the BMS incorrectly believes that the desired pressure has been reached or exceeded. As a result, it reduces the fan speed or maintains it at a lower level, leading to reduced airflow being supplied to the zone, which triggers the detection of Symptom CO2-Max. Furthermore, the prediction-based symptom Psa-Ppred-Diff is expected to detect a deviation from the predicted P_{sa}. However, this symptom was not detected, as shown inFig. B.1(e).

A.0.6. Fault Tra-bias (Case 9)

As shown in Fig. 7(n), Case 9 was correctly diagnosed as Fault Tra-Bias with a posterior probability of 0.10. Similar to Case 8, since T_{ra} is involved in the calculation of Symptoms HRW-Effi-Min and HRW-Effi1-Effi2, a positive bias in T_{ra} decreases η_{hrw}, causing it to fall below the nominal value and creating a deviation between η_{sa} and η_{eha}, thus triggering the detection of both Symptom HRW-Effi-Min and HRW-Effi1-Effi2. Another primary evidence, Symptom Tra-Tra2-Diff, was also detected at all time steps. This is because the symptom uses T_{rad}, measured in the distribution system, for comparison. When T_{ra} is biased, it becomes easier to detect this symptom.

A.0.7. Fault Tsw-bias (Case 10)

In Case 10, as shown in Fig. 7(o), Case 10 was correctly diagnosed as Fault Tsw-Bias with a posterior probability of 0.76. Two related diagnostic evidences Symptoms Tsa-Tpred-Diff and Tin-Tinset-Diff were detected as shown in Fig. 6(o). When T_{sw} is biased, Symptom Tsa-Tpred-Diff detects a deviation from the predicted T_{sa} as shown in Fig. B.1(c) and explained in Section 5.1. Another prediction symptom, Symptom Uhc-UPred-Diff, was not detected, as shown in Fig. B.1(d).

A.0.8. Fault Fsa-bias (Case 11)

As shown in Fig. 7(p), Case 11 was correctly diagnosed as Fault Fsa-Bias with a posterior probability of 1.0. All three related symptoms Nsf-Nsfpred-Diff, Psf-Psfpred-Diff, and SF-Fsa-Min were detected. When F_{sa} is biased, Symptom Nsf-Nsfpred-Diff detects a deviation from the predicted N_{sf}, as shown in Fig. B.1(a). The same explanation applies to Psf-Psfpred-Diff, as shown in Fig. B.1(b). Another key diagnostic symptom, Sf-Fsa-Min, was detected in nearly all time steps. This is because the biased F_{sa} reading fails to meet the nominal airflow value, leading to the detection of this symptom.

A.0.9. Normal day (Case 12)

As shown in Fig. 7(q), Case 12 was correctly diagnosed as a true negative across all faults. With the exception of Symptom Tin-Tinset-Diff, all other symptoms remained undetected in this case, leading to the accurate true negative diagnosis.

Appendix B. Additional material

This section contains supplementary data.

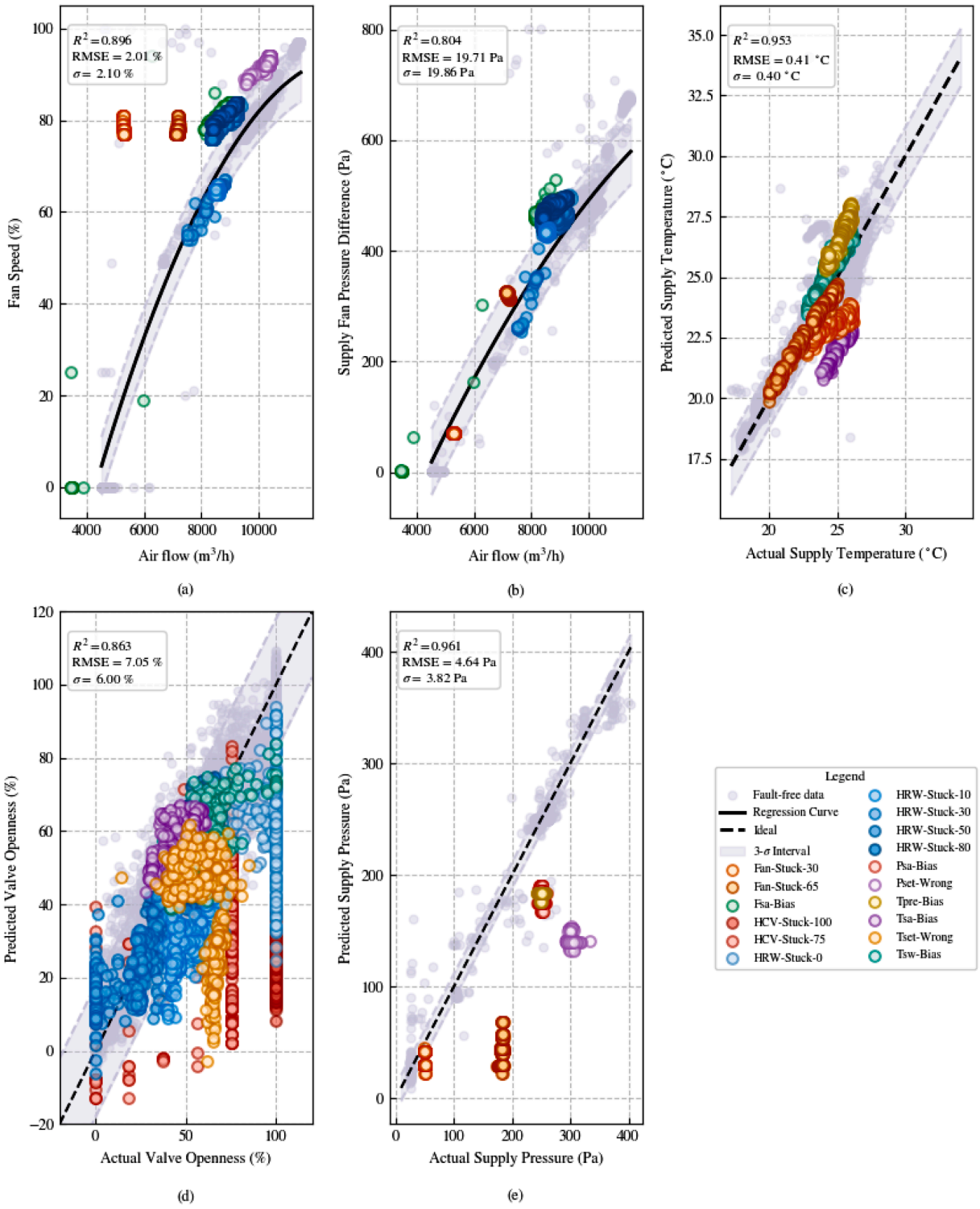


Fig. B.1. Comparison of measured and predicted values for symptom detection. (a) Fan control signal prediction symptom (S_{19}: $N_{sf,pred}$); (b) Fan pressure difference prediction symptom (S_{18}: $P_{sf,pred}$); (c) Supply air temperature prediction symptom (S_6: $T_{sa,pred}$); (d) Heating coil valve openness prediction symptom (S_4: $U_{hc,pred}$); (e) Supply air static pressure prediction symptom (S_9: $P_{sa,pred}$).

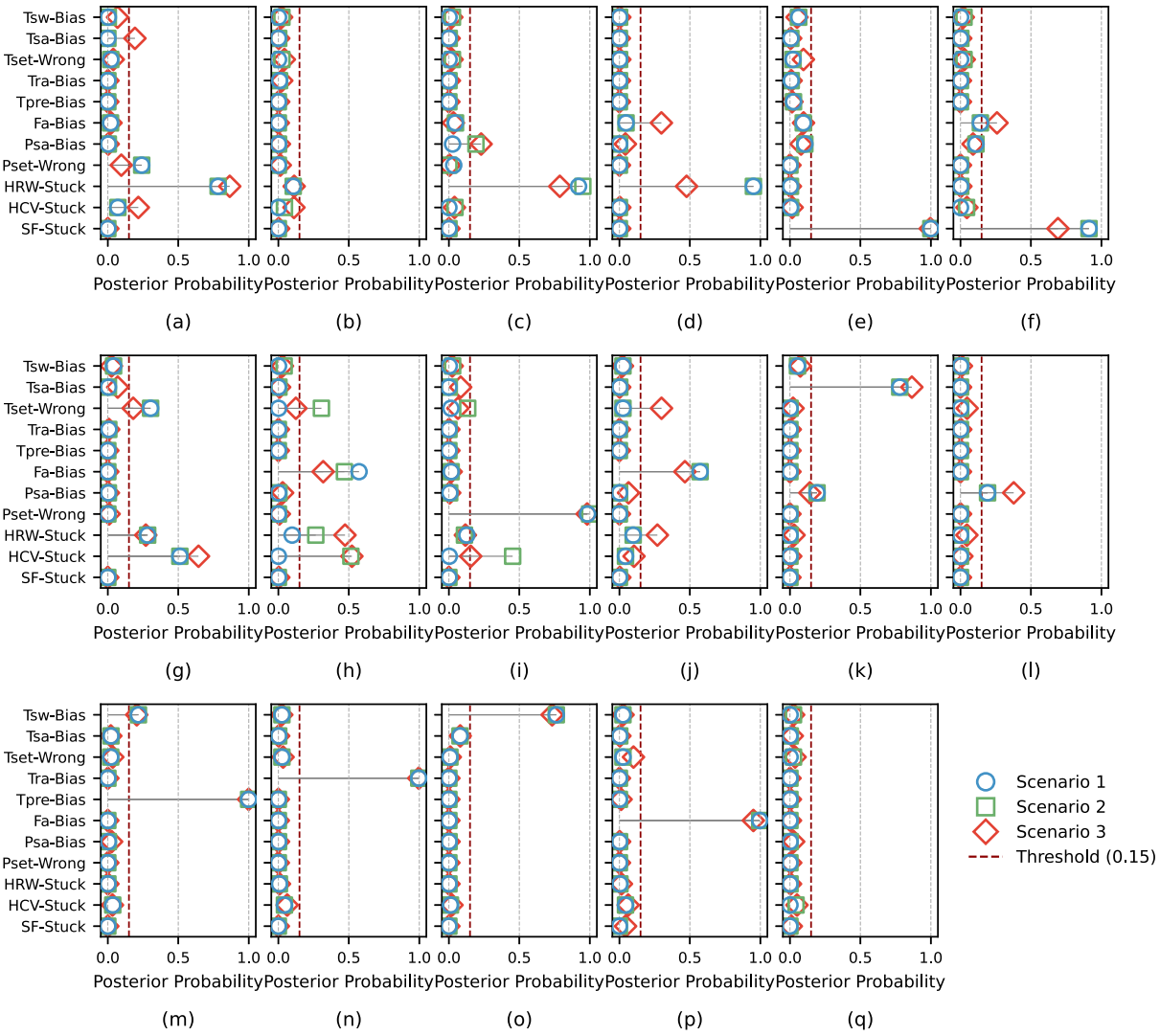


Fig. B.2. Comparative diagnostic performance of Scenario 1, Scenario 2, and Scenario 3, illustrating the posterior probability for seventeen fault cases. (a) HRW-Stuck-0. (b) HRW-Stuck-0. (c) HRW-Stuck-0. (d) HRW-Stuck-80. (e) Fan-Stuck-30. (f) Fan-Stuck-65. (g) HCV-Stuck-100. (h) HCV-Stuck-75. (i) Pset-Wrong. (j) Tset-Wrong. (k) Tsa-Bias. (l) Psa-Bias. (m) Tpre-Bias. (n) Tra-Bias. (o) Tsw-Bias. (p) Fa-Bias. (q) NM(normal).

Table B.1
ϵ values for symptom state definitions.

Symptom	ε_T	ε_P	ε_{Tin}	ε_{hc}	ε_{Tpre}	ε_{Tsa}	ε_{CO_2}	$\eta_{hrw,max}$	ε_{Psa}	$\eta_{sa,nm}$	ε_{fi}	$\varepsilon_{s,e}$	ε_{T2}	ε_{oa}	ε_N	$\varepsilon_{\Delta P_{sf}}$	$\varepsilon_{N_{sf}}$
Value	0.5	5	0.5	18	1	1.19	500	0.8	11.46	0.71	55	1	1	20	700	59.55	6.3
Unit	°C	Pa	°C	%	°C	°C	ppm	%	Pa	%	Pa	%	°C	°C	°C	Pa	%

Table B.2
Summary of defined fault-symptom relationships and their presence across sensor configurations along with the presence (✓) of each symptom under the Sensor-Rich, Sensor-Standard, and Sensor-Limited configurations, respectively.

No.	Fault	Symptom	Relationship	Rich	Standard	Limited
F_1	HRW-Stuck	HRW-Effi-Min	High	✓	✓	
		Uhc-Upred-Diff	Medium	✓		
		Nsf-Nsfpred-Diff	Medium	✓		
		Psf-Psfpred-Diff	Medium	✓		
F_2	SF-Stuck	CO2-Max	High	✓		
		Pset-Psa-Diff	High	✓	✓	✓
		HRW-Effi-Max	High	✓	✓	
		Pfi-Diff-Min	High	✓	✓	✓
		Psa-Ppred-Diff	Medium	✓		
		Tsa-Tpred-Diff	Medium	✓	✓	
		Uhc-Upred-Diff	Medium	✓		
		Psf-Psfpred-Diff	High	✓		
		Nsf-Nsfpred-Diff	High	✓		
F_3	HCV-Stuck	Tin-Tinset-Diff	High	✓	✓	
		Tset-Tsa-Diff	Medium	✓	✓	✓
		Uhc-Upred-Diff	High	✓		
F_4	Pset-Wrong	HRW-Effi-Min	High	✓	✓	
		Psa-Ppred-Diff	High	✓		
		Pset-Toa-Ctrl	High	✓	✓	✓
		Tset-Tsa-Diff	High	✓	✓	✓
		Uhc-Upred-Diff	High	✓		
		Nsf-NsfPred-Diff	Medium	✓		
		Psf-PsfPred-Diff	Medium	✓		
F_5	Tset-Wrong	Tin-Tinset-Diff	Medium	✓		
		Tset-Toa-Ctrl	Medium	✓	✓	✓
		Uhc-Upred-Diff	High	✓	✓	
F_6	Tsa-Bias	Tin-Tinset-Diff	Medium	✓		
		Tsa-Tsa2-Diff	High	✓		
		Tsa-Tpred-Diff	High	✓	✓	
		Uhc-Upred-Diff	Medium	✓		
F_7	Psa-Bias	CO2-Max	High	✓		
		Uhc-Upred-Diff	Medium	✓		
		Psa-Ppred-Diff	Medium	✓		
F_8	Tpre-Bias	HRW-Effi-Min	High	✓	✓	
		HRW-Effi1-Effi2	High	✓	✓	
		HRW-Effi-Max	High	✓	✓	
		Psa-Ppred-Diff	Medium	✓		
		Tpre-Traoa-Max	High	✓	✓	
		Tsa-Tpred-Diff	High	✓	✓	
F_9	Tra-Bias	HRW-Effi-Min	High	✓	✓	
		HRW-Effi1-Effi2	High	✓	✓	
		Tra-Tra2-Diff	High	✓		
F_{10}	Tsw-Bias	Tsa-Tpred-Diff	High	✓		
		Uhc-Upred-Diff	Medium	✓		
		Tin-Tinset-Diff	Medium	✓	✓	
F_{11}	Fsa-Bias	Sf-Fsa-Min	High	✓		
		Psf-Psfpred-Diff	High	✓		
		Nsf-Nsfpred-Diff	High	✓		

References

[1] E. Parliament, Directive (EU) 2024/1275 of the European parliament and of the council of 24 April 2024 on the energy performance of buildings (recast), 2024, (*Official Journal of the European Union*, L 127, 8 May 2024, pp. 1–50). http://data.europa.eu/eli/dir/2024/1275/oj.

[2] G. Martinopoulos, K.T. Papakostas, A.M. Papadopoulos, A comparative review of heating systems in EU countries, based on efficiency and fuel cost, Renew. Sustain. Energy Rev. 90 (2018) 687–699. https://doi.org/10.1016/j.rser.2018.03.060

[3] S. Katipamula, M.R.B. and, Review article: methods for fault detection, diagnostics, and prognostics for building systems-a review, Part I, HVAC&R Res. 11 (1) (2005) 3–25. https://doi.org/10.1080/10789669.2005.10391123

[4] Z. Li, J.-G. Wen, W. Cai, T.-W. Wong, Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future, Renew. Sustain. Energy Rev. 109 (2019) 85–101. https://doi.org/10.1016/j.rser.2019.04.021

[5] J. Wen, S. Li, RP-1312 – Tools for Evaluating Fault Detection and Diagnostic Methods for Air-Handling Units, Technical Report, ASHRAE, 2011.

[6] Y. Zhao, T. Li, X. Zhang, C. Zhang, Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future, Renew. Sustain. Energy Rev. 109 (2019) 85–101. https://doi.org/10.1016/j.rser.2019.04.021

[7] J. Chen, L. Zhang, Y. Li, Y. Shi, X. Gao, Y. Hu, A review of computing-based automated fault detection and diagnosis of heating, ventilation and air conditioning systems, Renew. Sustain. Energy Rev. 161 (2022) 112395. https://doi.org/10.1016/j.rser.2022.112395

[8] H. Yang, S. Cho, C.-S. Tae, M. Zaheeruddin, Sequential rule based algorithms for temperature sensor fault detection in air handling units, Energy Convers. Manage. 49 (8) (2008) 2291–2306. https://doi.org/10.1016/j.enconman.2008.01.029

[9] H. Wang, Y. Chen, C.W.H. Chan, J. Qin, An online fault diagnosis tool of VAV terminals for building management and control systems, Autom. Constr. 22 (2012) 203–211. Planning Future Cities-Selected papers from the 2010 eCAADe Conference. https://doi.org/10.1016/j.autcon.2011.06.018

[10] J. Schein, S.T. Bushby, N.S. Castro, J.M. House, A rule-based fault detection method for air handling units, Energy Build. 38 (12) (2006) 1485–1492. https://doi.org/10.1016/j.enbuild.2006.04.014

[11] P.M. Van Every, M. Rodriguez, C.B. Jones, A.A. Mammoli, M. Martínez-Ramón, Advanced detection of HVAC faults using unsupervised SVM novelty detection and Gaussian process models, Energy Build. 149 (2017) 216–224. https://doi.org/10.1016/j.enbuild.2017.05.053

[12] V. Martinez-Viol, E.M. Urbano, K. Kampouropoulos, M. Delgado-Prieto, L. Romeral, Support vector machine based novelty detection and FDD framework applied to building AHU systems, in: 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 1, 2020, pp. 1749–1754. https://doi.org/10.1109/ETFA46521.2020.9212088

[13] M. Meas, R. Machlev, A. Kose, A. Tepljakov, L. Loo, Y. Levron, E. Petlenkov, J. Belikov, Explainability and transparency of classifiers for air-handling unit faults using explainable artificial intelligence (XAI), Sensors 22 (17) (2022). https://doi.org/10.3390/s22176338

[14] D. Chakraborty, H. Elzarka, Early detection of faults in HVAC systems using an XGBoost model with a dynamic threshold, Energy Build. 185 (2019) 326–344. https://doi.org/10.1016/j.enbuild.2018.12.032

[15] Y. Gao, H. Han, H. Lu, S. Jiang, Y. Zhang, M. Luo, Knowledge mining for chiller faults based on explanation of data-driven diagnosis, Appl. Thermal Eng. 205 (2022) 118032. https://doi.org/10.1016/j.applthermaleng.2021.118032

[16] J. Gao, H. Han, Z. Ren, Y. Fan, Fault diagnosis for building chillers based on data self-production and deep convolutional neural network, J. Build. Eng. 34 (2021) 102043. https://doi.org/10.1016/j.jobe.2020.102043

[17] Y. Yu, H. Zhang, W. Peng, R. Wang, C. Li, Images based fault diagnosis of air handling unit via combining kernel slow feature analysis and deep learning method, J. Build. Eng. 56 (2022) 104718. https://doi.org/10.1016/j.jobe.2022.104718

[18] Y. Gao, S. Miyata, Y. Akashi, How to improve the application potential of deep learning model in HVAC fault diagnosis: based on pruning and interpretable deep learning method, Appl. Energy 348 (2023) 121591. https://doi.org/10.1016/j.apenergy.2023.121591

[19] C. Fan, Y. Lei, Y. Sun, L. Mo, Novel transformer-based self-supervised learning methods for improved HVAC fault diagnosis performance with limited labeled data, Energy 278 (2023) 127972. https://doi.org/10.1016/j.energy.2023.127972

[20] Y. Yan, J. Cai, Y. Tang, L. Chen, Fault diagnosis of HVAC AHUs based on a BP-MTN classifier, Build. Environ. 227 (2023) 109779. https://doi.org/10.1016/j.buildenv.2022.109779

[21] J. Bi, K. Yan, Y. Du, End-to-end residual learning embedded ACWGAN for AHU FDD with limited fault data, Build. Environ. 270 (2025) 112529. https://doi.org/10.1016/j.buildenv.2025.112529

[22] K. Yan, A. Chong, Y. Mo, Generative adversarial network for fault detection diagnosis of chillers, Build. Environ. 172 (2020) 106698. https://doi.org/10.1016/j.buildenv.2020.106698

[23] Z. Du, K. Chen, S. Chen, J. He, X. Zhu, X. Jin, Deep learning GAN-based data generation and fault diagnosis in the data center HVAC system, Energy Build. 289 (2023) 113072. https://doi.org/10.1016/j.enbuild.2023.113072

[24] C. Fan, W. He, Y. Liu, P. Xue, Y. Zhao, A novel image-based transfer learning framework for cross-domain HVAC fault diagnosis: from multi-source data integration to knowledge sharing strategies, Energy Build. 262 (2022) 111995. https://doi.org/10.1016/j.enbuild.2022.111995

[25] B. Feng, Q. Zhou, J. Xing, Q. Yang, Y. Chen, Z. Deng, Attention-empowered transfer learning method for HVAC sensor fault diagnosis in dynamic building environments, Build. Environ. 250 (2024) 111148. https://doi.org/10.1016/j.buildenv.2023.111148

[26] J. Liu, Q. Zhang, X. Li, G. Li, Z. Liu, Y. Xie, K. Li, B. Liu, Transfer learning-based strategies for fault diagnosis in building energy systems, Energy Build. 250 (2021) 111256. https://doi.org/10.1016/j.enbuild.2021.111256

[27] C. Lu, Z. Wang, M. Mosteiro-Romero, L. Itard, Diagnostic Bayesian network in building energy systems: current insights, practical challenges, and future trends, Energy Build. 341 (2025) 115845. https://doi.org/10.1016/j.enbuild.2025.115845

[28] Y. Zhao, J. Wen, S. Wang, Diagnostic Bayesian networks for diagnosing air handling units faults – Part II: faults in coils and sensors, Appl. Thermal Eng. 90 (2015) 145–157. https://doi.org/10.1016/j.applthermaleng.2015.07.001

[29] A. Taal, L. Itard, P&ID-based symptom detection for automated energy performance diagnosis in HVAC systems, Autom. Constr. 119 (2020) 103344. https://doi.org/10.1016/j.autcon.2020.103344

[30] Y. Zhao, J. Wen, F. Xiao, X. Yang, S. Wang, Diagnostic Bayesian networks for diagnosing air handling units faults – Part I: faults in dampers, fans, filters and sensors, Appl. Thermal Eng. 111 (2017) 1272–1286. https://doi.org/10.1016/j.applthermaleng.2015.09.121

[31] T. Li, Y. Zhao, C. Zhang, J. Luo, X. Zhang, A knowledge-guided and data-driven method for building HVAC systems fault diagnosis, Build. Environ. 198 (2021) 107850. https://doi.org/10.1016/j.buildenv.2021.107850

[32] T. Li, Y. Zhou, Y. Zhao, C. Zhang, X. Zhang, A hierarchical object oriented Bayesian network-based fault diagnosis method for building energy systems, Appl. Energy 306 (2022) 118088. https://doi.org/10.1016/j.apenergy.2021.118088

[33] 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. https://doi.org/10.1016/j.apenergy.2022.120050

[34] A. Taal, L. Itard, W. Zeiler, A reference architecture for the integration of automated energy performance fault diagnosis into HVAC systems, Energy Build. 179 (2018) 144–155. https://doi.org/10.1016/j.enbuild.2018.08.031

[35] 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. https://doi.org/10.1016/j.enbuild.2020.110289

[36] Z. Wang, C. Lu, A. Taal, S. Gopalan, K. Mohammed, A. Meijer, L. Itard, Bayesian network-based fault detection and diagnosis of heating components in heat recovery ventilation, in: Proceedings of RoomVent 2024, Stockholm, Sweden, 2024, p. 417.

[37] F. Cheng, W. Cai, X. Zhang, H. Liao, C. Cui, Fault detection and diagnosis for air handling unit based on multiscale convolutional neural networks, Energy Build. 236 (2021) 110795. https://doi.org/10.1016/j.enbuild.2021.110795

[38] P. Movahed, S. Taheri, A. Razban, A bi-level data-driven framework for fault-detection and diagnosis of HVAC systems, Appl. Energy 339 (2023) 120948. https://doi.org/10.1016/j.apenergy.2023.120948

[39] Y. Gao, H. Han, Z.X. Ren, J.Q. Gao, S.X. Jiang, Y.T. Yang, Comprehensive study on sensitive parameters for chiller fault diagnosis, Energy Build. 251 (2021) 111318. https://doi.org/10.1016/j.enbuild.2021.111318

[40] Z. Wang, Z. Wang, X. Gu, S. He, Z. Yan, Feature selection based on Bayesian network for chiller fault diagnosis from the perspective of field applications, Appl. Thermal Eng. 129 (2018) 674–683. https://doi.org/10.1016/j.applthermaleng.2017.10.079

[41] D. Li, Y. Zhou, G. Hu, C.J. Spanos, Optimal sensor configuration and feature selection for AHU fault detection and diagnosis, IEEE Trans. Ind. Inform. 13 (3) (2017) 1369–1380. https://doi.org/10.1109/TII.2016.2644669

[42] L. Zhang, M. Leach, J. Chen, Y. Hu, Sensor cost-effectiveness analysis for data-driven fault detection and diagnostics in commercial buildings, Energy 263 (2023) 125577. https://doi.org/10.1016/j.energy.2022.125577

[43] M.S. Piscitelli, D.M. Mazzarelli, A. Capozzoli, Enhancing operational performance of AHUs through an advanced fault detection and diagnosis process based on temporal association and decision rules, Energy Build. 226 (2020) 110369. https://doi.org/10.1016/j.enbuild.2020.110369

[44] K. Yan, J. Huang, W. Shen, Z. Ji, Unsupervised learning for fault detection and diagnosis of air handling units, Energy Build. 210 (2020) 109689. https://doi.org/10.1016/j.enbuild.2019.109689

[45] W.-S. Yun, W.-H. Hong, H. Seo, A data-driven fault detection and diagnosis scheme for air handling units in building HVAC systems considering undefined states, J. Build. Eng. 35 (2021) 102111. https://doi.org/10.1016/j.jobe.2020.102111

[46] W. Zheng, J. Hu, Z. Wang, J. Li, Z. Fu, H. Li, J. Jurasz, S.K. Chou, J. Yan, COVID-19 impact on operation and energy consumption of heating, ventilation and air-conditioning (HVAC) systems, Adv. Appl. Energy 3 (2021) 100040. https://doi.org/10.1016/j.adapen.2021.100040

[47] J. Laverge, A. Janssens, Heat recovery ventilation operation traded off against natural and simple exhaust ventilation in Europe by primary energy factor, carbon dioxide emission, household consumer price and exergy, Energy Build. 50 (2012) 315–323. https://doi.org/10.1016/j.enbuild.2012.04.005

[48] O. Nehasil, L. Dobiasova, V. Mazanec, J. Siroky, Versatile AHU fault detection – design, field validation and practical application, Energy Build. 237 (2021) 110781. https://doi.org/10.1016/j.enbuild.2021.110781

[49] H. Geoffroy, J. Berger, E. Gonze, C. Buhe, Experimental dataset for an AHU air-to-air heat exchanger with normal and simulated fault operations, J. Build. Perform. Simul. 16 (3) (2023) 268–290. https://doi.org/10.1080/19401493.2022.2097311

[50] S.P. Melgaard, R.L. Jensen, K.H. Andersen, P.K. Heiselberg, AHU heat exchanger performance tracking during the building operational phase, Appl. Thermal Eng. 270 (2025) 126204. https://doi.org/10.1016/j.applthermaleng.2025.126204

[51] L. van Koetsveld van Ankeren, C. Lu, L. Itard, Implementing diagnostic Bayesian networks for heat recovery ventilation in real-world scenarios: a dutch case study, J. Build. Eng. 111 (2025) 113527. https://doi.org/10.1016/j.jobe.2025.113527

[52] V. REHVA (Federation of European Heating, A.C. Associations), REHVA COVID-19 guidance document, 2021, Version 4.1, Accessed May 7, 2025, https://www.rehva.eu/fileadmin/user_upload/REHVA_COVID-19_guidance_document_V4.1_15042021.pdf.

[53] ASHRAE, Guideline 36: High Performance Sequences of Operation for HVAC Systems, ASHRAE Guideline 36-2018, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, USA, 2018.

[54] ISSO, Meetpunten en Meetmethoden Voor Klimaatinstallaties: Regelen, Sturen, Beveiligen En Monitoren, ISSO-publicatie 31, Stichting ISSO, Rotterdam, Nederland, 2014.

[55] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, 1988. https://doi.org/10.1016/C2009-0-27609-4

[56] I.J. Good, A causal calculus (I), Br. J. Philos. Sci. 11 (44) (1961) 305–318. https://doi.org/10.1093/bjps/XI.44.305

[57] 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. https://doi.org/10.1016/j.enbuild.2021.111448

[58] J. Gao, M. Bergés, A large-scale evaluation of automated metadata inference approaches on sensors from air handling units, Adv. Eng. Inform. 37 (2018) 14–30. https://doi.org/10.1016/j.aei.2018.04.010

[59] S. Gopalan, A. Rijs, S. Chitkara, A. Thamban, R. Kramer, Fault prioritisation for air handling units using fault modelling and actual fault occurrence data, Energy Build. 319 (2024) 114476. https://doi.org/10.1016/j.enbuild.2024.114476

[60] S. Li, Development and validation of a dynamic air handling unit model, Part I, ASHRAE Trans. 116 (2010) 45.

[61] N. Torabi, H. Burak Gunay, W. O'Brien, R. Moromisato, Inverse model-based virtual sensors for detection of hard faults in air handling units, Energy Build. 253 (2021) 111493. https://doi.org/10.1016/j.enbuild.2021.111493

[62] S. Pecceu, S. Caillou, Efficiency of heat recovery ventilation in real conditions: feedback from several measurement campaigns, in: Proceedings of the 40th AIVC - 8th TightVent - 6th venticool Conference, Ghent, Belgium, 2019, pp. D1_S3B–04. https://www.aivc.org/sites/default/files/D1_S3B-04.pdf.

[63] D. O'Connor, J. Calautit, B.R. Hughes, Effect of rotation speed of a rotary thermal wheel on ventilation supply rates of wind tower system, Energy Procedia 75 (2015) 1705–1710. Clean, Efficient and Affordable Energy for a Sustainable Future: The 7th International Conference on Applied Energy (ICAE2015). https://doi.org/10.1016/j.egypro.2015.07.432

[64] J.K. Calautit, D. O'Connor, P.W. Tien, S. Wei, C.A.J. Pantua, B. Hughes, Development of a natural ventilation windcatcher with passive heat recovery wheel for mild-cold climates: CFD and experimental analysis, Renew. Energy 160 (2020) 465–482. https://doi.org/10.1016/j.renene.2020.05.177

[65] T. Parhizkar, F. Aramoun, S. Esbati, Y. Saboohi, Efficient performance monitoring of building central heating system using Bayesian network method, J. Build. Eng. 26 (2019) 100835. https://doi.org/10.1016/j.jobe.2019.100835

[66] M. Hu, H. Chen, L. Shen, G. Li, Y. Guo, H. Li, J. Li, W. Hu, A machine learning Bayesian network for refrigerant charge faults of variable refrigerant flow air conditioning system, Energy Build. 158 (2018) 668–676. https://doi.org/10.1016/j.enbuild.2017.10.012

[67] F. Xiao, Y. Zhao, J. Wen, S. Wang, Bayesian network based FDD strategy for variable air volume terminals, Autom. Constr. 41 (2014) 106–118. https://doi.org/10.1016/j.autcon.2013.10.019

[68] Y. Li, Z. O'Neill, An innovative fault impact analysis framework for enhancing building operations, Energy Build. 199 (2019) 311–331. https://doi.org/10.1016/j.enbuild.2019.07.011

[69] D. Dey, B. Dong, A probabilistic approach to diagnose faults of air handling units in buildings, Energy Build. 130 (2016) 177–187. https://doi.org/10.1016/j.enbuild.2016.08.017

[70] Z. Wang, C. Lu, M. Mosteiro-Romero, L. Itard, Simultaneous presents faults detection by using diagnostic Bayesian network in air handling units, in: Proceedings of ASim2024, The 5th Asia Conference of the IBPSA, IBPSA, Osaka, Japan, 2024, pp. 1613–1619. https://doi.org/10.69357/asim2024.1273

[71] C. Lu, S. Walker, C. Struck, L. Itard, D. Saelens, Poster abstract: P&ID-to-graph: exploring LLM-assisted approaches for HVAC diagrams digitalization, in: Proceedings of the 12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '25), ACM, Golden, CO, USA, 2025. https://doi.org/10.1145/3736425.3772104

[72] C. Lu, L. Itard, Leveraging LLM for P&ID-based automated code generation in HVAC fault detection and diagnosis, in: Proceedings of the 15th REHVA HVAC World Congress – CLIMA 2025, Springer Nature, Milan, Italy, 2025.