TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Mosteiro-Romero, M., Wang, Z., Taal, A., Lu, C., & Itard, L. (2026). Whole building HVAC diagnostic Bayesian networks for comfort-related fault detection and diagnosis: Application, evaluation and operational testing. *Building and Environment*, *305*, Article 115265. https://doi.org/10.1016/j.buildenv.2026.115265

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Building and Environment 305 (2026) 115265



Contents lists available at ScienceDirect

Building and Environment

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



Whole building HVAC diagnostic Bayesian networks for comfort-related fault detection and diagnosis: Application, evaluation and operational testing

Martín Mosteiro-Romero [a,*], Ziao Wang [a], Arie Taal [b], Chujie Lu [a], Laure Itard [a]

[a] Department of Architectural Engineering and Technology, Delft University of Technology, Julianalaan 134, Delft, 2628 BL, Netherlands
[b] Department Mechanical Engineering, The Hague University of Applied Sciences, Rotterdamseweg 137, Delft, 2628 AL, Netherlands

HIGHLIGHTS

- A Diagnostic Bayesian Network for whole-HVAC system fault detection and diagnosis was developed.
- The DBN uses both knowledge-based and data-driven methods for parameter modeling.
- DBN performance was evaluated using historical data and on-site AHU fault experiments.
- Despite sensor uncertainty, it achieved a classification rate of 87.8% for AHU fault experiments.
- DBN application during HVAC operation was demonstrated by diagnosing a suspected real-world fault.

ARTICLE INFO

Keywords:
Fault detection and diagnosis
Diagnostic Bayesian networks
Whole building HVAC systems
Cross-level faults

ABSTRACT

Automated fault detection and diagnosis (FDD) can support building energy performance and predictive maintenance by leveraging the vast amounts of data generated by building management systems. However, most work on FDD for HVAC systems focuses on individual components, neglecting complex system interactions and the cascading effects of component faults. Diagnostic Bayesian Networks (DBN) offer a particularly promising approach for whole-building FDD due to their robustness, flexibility and scalability. This paper presents the development of a DBN for whole-HVAC FDD and its application in an office building in the Netherlands. The DBN uses aggregated representations of all major building systems to detect which faulty components need further investigation and uses knowledge-based and data-driven methods for parameter estimation. The DBN's performance and the effects of input parameter assumptions are assessed using historical data as well as experimental data obtained by introducing simultaneous faults into the air handling unit (AHU) controls. Although fault injection was restricted to AHU controls due to safety concerns, the model architecture represents all major HVAC systems and simultaneously diagnoses their faults. The results using historical data identified persistent sensor-related anomalies in the building, as well as possibly associated system and control faults. Despite these sensor issues, the evaluation using experimental data showed that the DBN achieved a classification rate of 87.8% for AHU faults while maintaining a low false alarm rate (3.8%). However, the miss rate was high, exceeding 25% in all cases. Finally, the DBN is demonstrated through its application to diagnose a suspected fault during operation.

1. Introduction

Modern building systems are complex, with numerous components and interactions between systems, buildings and occupants. Through sensors deployed throughout the heating, ventilation and air conditioning (HVAC) system, building management systems (BMS) collect vast amounts of data, which potentially allow real-time monitoring of building operation and system performance optimization. However, the complexity arising from the coupling and interactions between components and controls makes it difficult to maintain fault-free operation [1]. Therefore, buildings continue to suffer from energy waste, uncomfortable indoor environments and poor air quality due to malfunctioning components and inadequate controls. Component degradation

* Corresponding author.
Email address: m.a.mosteiroromero@tudelft.nl (M. Mosteiro-Romero).

https://doi.org/10.1016/j.buildenv.2026.115265
Received 18 June 2026; Received in revised form 7 September 2026; Accepted 13 September 2026
Available online 14 September 2026
0360-1323/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Nomenclature

Abbreviations

AHU	Air Handling Unit
BA	Balanced Accuracy
BMS	Building Management System
CR	Classification Rate
DBN	Diagnostic Bayesian Networks
FAR	False Alarm Rate
FDD	Fault Detection and Diagnosis
FN	False Negatives
FP	False Positives
HCV	Heating Coil Valve
HRW	Heat Recovery Wheel
HVAC	Heating, Ventilation and Air Conditioning
MR	Miss Rate
P&ID	Piping & Instrumentation Diagram
Pr	Precision
TN	True Negatives
TP	True Positives
VAV	Variable Air Volume

Symbols

σ	Standard deviation
ϵ	Threshold
DP	Pressure difference
DT	Temperature difference
F_1	F1-score
P	Pressure
RH	Relative Humidity
T	Temperature

Subscripts

$atrium$	Building zone Atrium
b	Boiler
c	Chiller
$calc$	Calculated setpoint
$comfort$	Comfort-related threshold
$cooling$	Cooling available
dev	Measurement deviation symptom
$dist$	Hydronic distribution
$east$	Building zone East
$heating$	Heating available
$LoRa$	LoRa sensor measurement
max	Maximum
$range$	Measurement range symptom
ret	Return
s	Sensor symptom-related threshold
set	Setpoint
$strong$	Strong link
sup	Supply
$west$	Building zone West

and malfunction have been found to lead to excess energy demand ranging from 15% [2] to 40% [3].

Automated fault detection and diagnosis (FDD) offers a promising solution to this problem by leveraging building automation data to detect faults in physical systems, diagnose their causes and potentially provide prognostic maintenance for degrading systems [2]. A number of methods for FDD in building systems can be found in the literature, including HVAC-level energy efficiency, chillers, air handling units (AHUs) and sensors. Most research has focused on individual components, and in particular AHUs, chillers, sensors, vapor compression systems, and variable refrigerant flow systems, which account for more than 80% of published literature in the field [1]. Chillers and AHUs have been identified as the energy-using components in an HVAC system that are most likely to cause serious and costly problems [4], which might explain their predominance in building system FDD.

Real-world HVAC systems, however, typically comprise multiple, highly coupled subsystems that need to be operated and controlled simultaneously [5]. Single-component FDD methods might neglect the interdependencies between different components, leading to inadequate fault diagnosis in practice [6]. Faults occurring in one subsystem may propagate and influence other subsystems, leading to false alarms in those other subsystems [7]. Therefore, ensuring fault-free operation in whole building HVAC systems requires an integrated approach that can overcome these limitations. However, a review on FDD for building systems revealed that only 12% of all studies in the literature focused on the whole building system [8].

Among the proposed FDD methods in the literature, Diagnostic Bayesian Networks (DBNs) are particularly promising for whole building HVAC system FDD due to their robustness, flexibility and scalability [9]. DBNs draw upon a Bayesian statistical approach, in which the probability of faults is calculated according to the occurrence of symptoms. Compared to other methods, DBNs can address a number of different and potentially simultaneous faults [1], which is especially relevant at the whole HVAC system scale.

Despite their potential advantages, the application of DBNs to whole building HVAC systems has so far been limited [9]. This is in part due to the difficulty of modeling complex, highly integrated building systems. Furthermore, the complexity of the whole building HVAC system makes it difficult to obtain labeled fault data in the real world, which further complicates the evaluation of DBN performance in practice. The use of DBNs for whole building HVAC systems, including all generation and distribution systems, their validation using fault data collected through experiments in building systems, and their evaluation using real-world operational data have not been demonstrated in the literature.

1.1. Scope of this paper

This paper presents the development of a DBN for a whole building HVAC system, as well as its testing, partial validation and application in a real case study building. Our methodology uses generic representations of individual systems within the building (boiler group, chiller group, hydronic groups, ventilation groups, and end user groups), such that the DBN can determine which subcomponents need to be further investigated by operators or by their own FDD methods where available. As part of the development of the model, both knowledge-based and data-driven methods for the definition of the DBN input parameters are tested and compared.

The DBN was developed for a case study building in the Netherlands and subsequently tested, partially validated and applied to investigate a suspected fault during operation. First, the DBN is used to detect faults during a two-year period (2022–2023) using historical data to demonstrate that the DBN structure represents and actively diagnoses all major HVAC systems simultaneously. The performance of the DBN is then partially tested using data from fault experiments carried out in the case study building's AHUs during six days in March–April 2024. A sensitivity analysis is also carried out to verify the effects of input parameters on the model's performance. While the experiments were restricted to AHU controls due to safety concerns, the model architecture represents

and actively diagnoses all major HVAC systems simultaneously. The final network is subsequently used to detect faults in the building systems during two weeks when there was a suspected issue with the building's ventilation systems.

1.2. Contribution of this paper

The work presented in this paper differentiates itself from previous research in a few ways. First of all, the focus on the whole HVAC system, including heating, cooling and ventilation systems, has rarely been seen in DBN literature. For example, other papers focusing on interdependencies in building systems have focused on specific system couples, such as heating and ventilation [6] or cooling and ventilation [7,10]. Furthermore, the development of our DBN combines knowledge-based and data-driven methods, as opposed to purely knowledge-based approaches (e.g., [11]). Unlike other works in the literature, we further carry out both qualitative evaluation and partial quantitative evaluation of our model's performance. The use of normal building operation as well as on-site experimental data provides insight into the challenges and limitations of DBNs in real-world applications, as opposed to the use of model-based data for DBN development [5]. Thus, this paper demonstrates the application of DBNs to a real whole-building HVAC system, accounting for all system interactions, and its evaluation through both multi-year historical data and experimental fault injection under real operating conditions.

2. Background

2.1. FDD for whole building HVAC systems

Most works on whole-building FDD aim to detect anomalies in building energy demand data as potential symptoms for faulty building systems. Among these works, clustering and decision trees are the most popular techniques for anomaly detection at whole-building level [17]. Pattern recognition techniques, clustering and artificial neural ensembling networks have also been used to detect outliers and anomalies in building energy consumption [18,19]. However, such methods only focus on the whole building level, without providing any information on which systems might be leading to the detected anomalies.

Other authors have proposed methods to detect faults in specific building systems based on whole-building energy consumption data. For example, principal component analysis (PCA) and correlation analysis [20,21] and recursive deterministic perceptron (RDP) neural networks [22] have been used for these purposes. Feature-based PCA along with weather and schedule-based pattern matching has also been proposed to differentiate normal and abnormal operation in a whole building system [23,24]. However, this method is unable to isolate the root cause of cross-level faults and requires detailed analysis when various fault symptoms are observed [7]. Statistical Process Control (SPC) with Kalman filtering based on gray-box models has also been proposed to detect sudden and gradual faults in a whole-building HVAC system [25]. While the method is promising in terms of scalability and replicability, it has only been developed for two specific building systems and no validation with real data has been provided.

A key limitation of these fully data-driven methods is the need for labeled "normal" vs. "faulty" data, which can be difficult to collect in real-world applications. For example, while the most common faults for chillers have been well documented, it is hard to obtain labeled data for all common faults in the field [26]. Some recent advances in data-driven FDD seek to address this practical challenge by combining limited labeled samples with unlabeled data. In particular, data augmentation techniques such as generative adversarial networks (GANs) and variational autoencoders (VAEs), and semi-supervised learning approaches have been applied to address the issues of limited labeled data and data scarcity. These methods have been successfully applied at the component scale, such as for AHUs [27,28], refrigeration systems [29,30] and rooftop units [31]. However, these methods still exhibit important limitations in practical HVAC FDD applications. Synthetic samples are constrained by the distributions learned from existing observations, while semi-supervised learning approaches primarily exploit abundant unlabeled operational data and remain dependent on a small amount of labeled "faulty" data. Consequently, both approaches offer limited capability to generalize to unseen faults and changing operating conditions, especially for highly interconnected whole-building HVAC systems. Therefore, applications in whole-building HVAC FDD are less common and achieve much more modest performance due to the complex interactions that arise at this scale [31].

Other challenges in real-world applications include insufficient sensors, poor sensor accuracy, non-standardized sensor installation and control logic, missing data and an imbalance between faulty and fault-free training data [32]. As a result, most data-driven FDD techniques in the literature use data from laboratory experiments or simulations to validate their methods, while most works using real building data only focus on building-scale energy anomalies [17]. However, laboratory experiments on single systems and simulations might not be able to replicate real-world conditions in highly integrated, large-scale HVAC systems. Furthermore, fully data-driven methods can have low interpretability [4] and neglect valuable information in building operation, such as expert knowledge from facilities management and the perception and feedback of building occupants [33].

2.2. DBNs for whole building systems

Diagnostic Bayesian Networks (DBN) are probabilistic graphical models in which faults and symptoms are represented as nodes in a network connected by edges representing the causal relationships between faults and symptoms (Fig. 1). Fault nodes typically have two states (fault present or absent), whereas symptom states may also be binary or indicate the strength of the evidence of the symptom (e.g., strong, medium, weak). Prior probabilities for each fault and conditional probabilities of a symptom being detected given the presence of a given fault are assigned based on either expert knowledge or historical data. By applying Bayes' theorem, posterior probabilities for the presence of fault F_i given the presence of symptom S_j can be calculated from the prior and conditional probabilities:

$$P(F_i|S_j) = \frac{P(F_i) \cdot P(S_j|F_i)}{\sum_{x \in X} P(F_x) \cdot P(S_j|F_x)} \quad (1)$$

DBNs have several advantages that make them attractive for whole building HVAC system FDD. Unlike most data-driven methods, DBNs are interpretable, as the causal relationships between faults and symptoms are explicitly modeled [35]. This is an attractive feature for real-world applications, as building professionals tend to be suspicious of FDD methods when they are not able to understand the model inference mechanism [36]. Additionally, most FDD methods in the field of building energy systems have traditionally accounted only for either

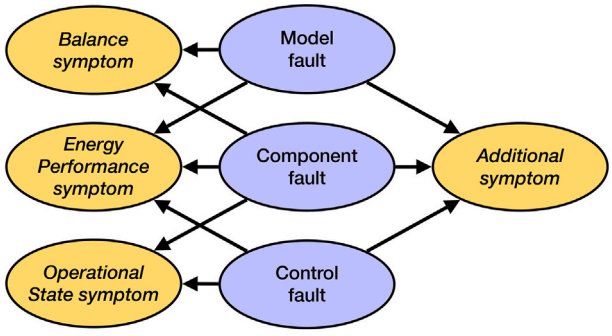

Fig. 1. Graphical representation of the DBN structure used in the four symptoms and three faults (4S3F) framework [34].

Table 1
Overview of papers in the literature focusing on DBNs for whole-HVAC or cross-level FDD.

Ref.	Scope	Systems	Structure	Parameter	Evaluation	Basis/Data source
[7]	Cooling and Ventilation	Chiller, AHU, VAV	Knowledge-based	Hybrid	Quantitative	Experimental data
[10]	Cooling and Ventilation	Chiller, AHU	Knowledge-based	Knowledge-based	Quantitative	Experimental data
[6]	Heating and Ventilation	Boiler, AHU, Radiator, Zone	Knowledge-based	Data-driven	Quantitative	Simulated data
[12]	Whole-HVAC	Only AHU faults	Knowledge-based	Knowledge-based	Quantitative	Simulated data
[5]	Whole-HVAC	Only AHU and duct faults	Data-driven	Data-driven	Quantitative	Simulated data
[11,13]	Whole-HVAC	Boiler, Heat pump, Hydronics, Storage	Knowledge-based	Knowledge-based	Qualitative	Historical data analysis
[14]	Whole-HVAC	Boiler, Hydronics, Radiator, Zone	Knowledge-based	Knowledge-based	Qualitative	Simulated data
[15]	Whole-building	HVAC and Envelope	Knowledge-based	Knowledge-based	Qualitative	Comfort prediction
[16]	Whole-building	HVAC and Envelope	Knowledge-based	Knowledge-based	Qualitative	Historical data analysis

sensor or component faults [32]. DBNs can consider both kinds of faults by representing their relationships to symptoms in the same network [37], and fault isolation of multiple faults takes place simultaneously. Furthermore, since they are probabilistic models, DBNs are robust to uncertainties and incomplete information [7], which are highly desirable features given the complexity of whole building HVAC systems. Finally, DBNs are expandable, as a model can be built from DBN submodels and can be set up using aggregated DBN models from a DBN model library [13]. This scalability can allow DBNs to be easily adapted as new systems are installed or if old systems are replaced.

Despite these advantages, limited work has been done on DBNs for whole building systems, as summarized in Table 1. Work on whole-HVAC DBNs often reduces the scale of the problem by focusing on specific system couples rather than on the whole HVAC system. For example, Chen et al. [7] developed a discrete Bayesian Network (DisBN) approach for diagnosing and isolating cross-level faults in a VAV HVAC system. Their work aimed to address the difficulty of obtaining prior probability distributions and conditional probability tables as required by Bayesian Networks. However, they specifically focused on nine different cross-system faults affecting the chiller and AHU systems rather than the whole HVAC. [6] proposed a multiple-model approach accounting for both HVAC subsystem interdependencies and changing modes of operation. In their approach, individual DBNs were developed for each mode of operation, using a combination of expert knowledge and historical data, and were implemented on an HVAC system consisting of a zone to be conditioned, a boiler, pump, radiator, and an AHU. By focusing on specific modes of operation, however, they focused only on specific system couples rather than accounting for all interdependencies. Furthermore, seasonal DBNs cannot detect simultaneous operation of the heating and cooling systems, which is one of the most common faults in commercial and educational buildings [38].

Another key difficulty is DBN evaluation, which is complicated by the lack of labeled fault data in highly integrated building systems. Some authors therefore evaluate their methods qualitatively by analyzing historical data and assessing the feasibility of the fault isolation results [16]. For example, Taal and Itard [11] and Taal and Itard [13] developed a whole-HVAC DBN for a case study building in the Netherlands and demonstrated their methodology by analyzing their method's capabilities for detecting various symptoms. These works demonstrate the potential for DBNs to be developed at different scales for different stages of fault isolation; however the performance of the DBN is not evaluated through quantitative metrics.

Other authors instead use simulations to introduce faults in a virtual test bed for quantitative assessment [14]. For example, Verbert et al. [6] demonstrated their approach through two simulated faults. Similarly, Huang et al. [5] present a fully data-driven method to develop DBNs for cross-level fault diagnosis and isolation and evaluate it on a simulated HVAC system. In their simulated evaluation, their method performed similarly to methods incorporating expert knowledge while decreasing network complexity. However, the authors acknowledge that their methodology, like most data-driven approaches, is influenced by factors such as data volume, data quality and data information. These issues are exacerbated in real-world applications.

In summary, the literature on DBNs for HVAC system FDD shows a lack of works focused on the whole building and its interactions, with some works focusing on specific system couples (e.g., heating and ventilation, cooling and ventilation) rather than on the whole HVAC system. Furthermore, there is a lack of works using real-world data for validation, with most works using simulation results due to the difficulty of accessing labeled fault data. Finally, while the use of purely data-driven approaches in FDD is promising, their applicability beyond the specific boundary conditions related to the training data used can be limited [17]. Combining knowledge-based and data-driven approaches in DBNs provides a viable approach to improve the robustness and accuracy of whole-HVAC system FDD.

3. Case study

In this paper, we developed a DBN for a whole HVAC system, evaluated it qualitatively and quantitatively, and tested it during operation in a case study seven-floor office building in the Netherlands (Fig. 2). The building mainly houses office space, with large atria and open hallways, as shown in a typical floor plan of the building. The building is equipped with two boilers, one chiller and two AHUs, both equipped with cooling coils and heat recovery wheels, but only one of which has a heating coil. The building's operation is split into three main zones: East, West and Atrium. Each zone corresponds to one hydronic distribution group, which supplies hot water to the radiators in each zone. Cooling is delivered by a chiller to the AHUs, which both supply all three zones.

According to the building's control documents, heating systems are available when the outdoor temperature is below 20 °C, while cooling systems are available when it is at least 14 °C. These overlapping operating temperatures can potentially lead to simultaneous operation of the heating and cooling systems in spring and autumn. The hydronic systems are controlled in order to maintain a desirable supply temperature as a function of the outdoor air temperature and the demand for heating in the building.

Air is supplied to each office by variable air volume (VAV) systems that are controlled based on CO_2 measurements at the outlet of each room. During normal operation, the air distribution systems provide the minimum amount of ventilation necessary as long as CO_2 measurements stay below 400 ppm. The volume of air supplied is increased as CO_2 increases, with maximum airflow provided when CO_2 levels exceed 1400 ppm. Cooling is supplied only by the ventilation systems and no thermostat is available in the offices. Building occupants do

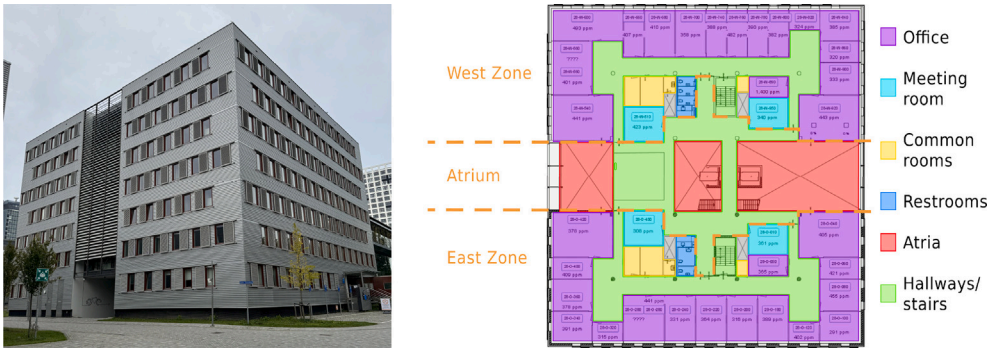

Fig. 2. Case study building (left) and a typical floor plan for the building (right).

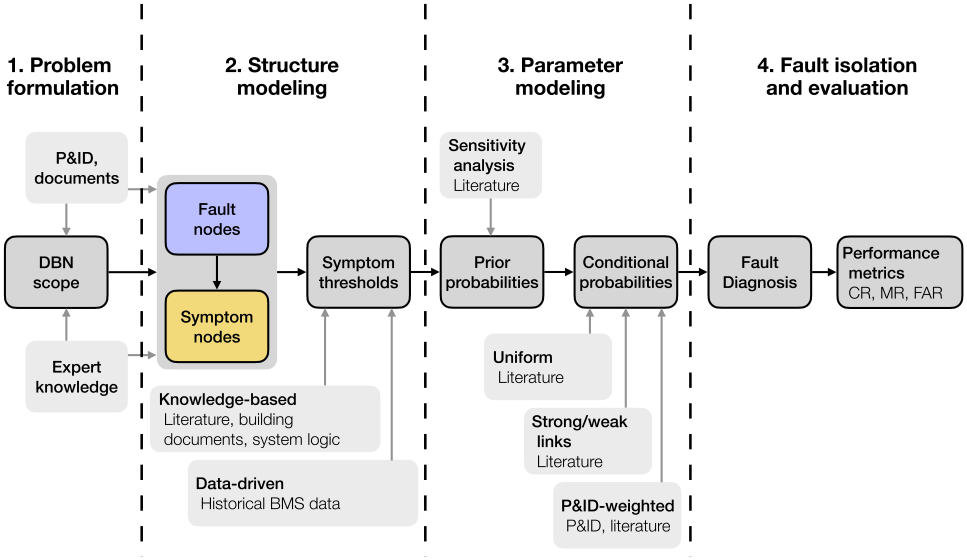

Fig. 3. Workflow used to create, test and evaluate the DBN. Each of the cases considered when defining the symptom thresholds, prior probabilities and conditional probabilities along with the data sources used is shown in a lighter color. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

have the possibility of opening their windows if they require additional ventilation or passive cooling.

The HVAC systems are equipped with temperature, pressure and moisture sensors, but there are no flow meters in any of the components. Mass flow and airflow sensors have higher noise content and measurement uncertainty [39] as well as higher cost [40] and are therefore often excluded from HVAC systems. For example, ASHRAE Guideline 36 [41], which focuses on the control and monitoring of air distribution systems in common HVAC configurations, does not include flow meters in its minimum recommended sensors. However, as a result, energy balance symptoms in the system cannot be detected. In this paper we therefore focus primarily on operational state and occupant comfort symptoms, which can be detected by the typical minimum sensor installation in buildings to support operational comfort.

4. Methodology

The procedure for developing the DBN is shown schematically in Fig. 3. The first step involves defining the scope of the problem and systems to be considered. In the second step, the dependencies between faults and symptoms in the system are defined and conceptualized graphically in a Bayesian Network. Subsequently, the parameters required by the model, namely the prior and conditional probabilities for fault detection, are defined. Finally, we introduce the fault isolation methods and the metrics used for evaluation.

4.1. Problem formulation

A DBN including all possible faults and symptoms in every subcomponent of the whole building HVAC system can quickly become impractically complex. In order to reduce the scale of the problem, we use aggregated system definitions (e.g., AHU, boiler, chiller), as has been done by other authors [6,20], without including faults in their sub-components. The advantage of such an approach is that it allows subsystem-specific methods, possibly developed by manufacturers, to be applied to trade components. Furthermore, it allows for subcomponent-level methods to be updated or replaced if changes in the building's subsystems are introduced over time. Fig. 4 shows the aggregated system boundaries for two generator modules (boiler and chiller) and two air modules (AHU1 and AHU2).

The resulting aggregated system P&ID for the case study building is shown in Fig. 5. The two boilers are shown as the Boiler group, which supplies the three distribution groups and AHU1. The chiller delivers cooling to the two AHUs, which both supply all three zones and are controlled by CO_2 sensors located in each room.

4.2. Structure modeling

In developing the DBN structure, we first define all possible faults in the HVAC system, and then extract the symptoms that would arise from each fault from the P&ID.

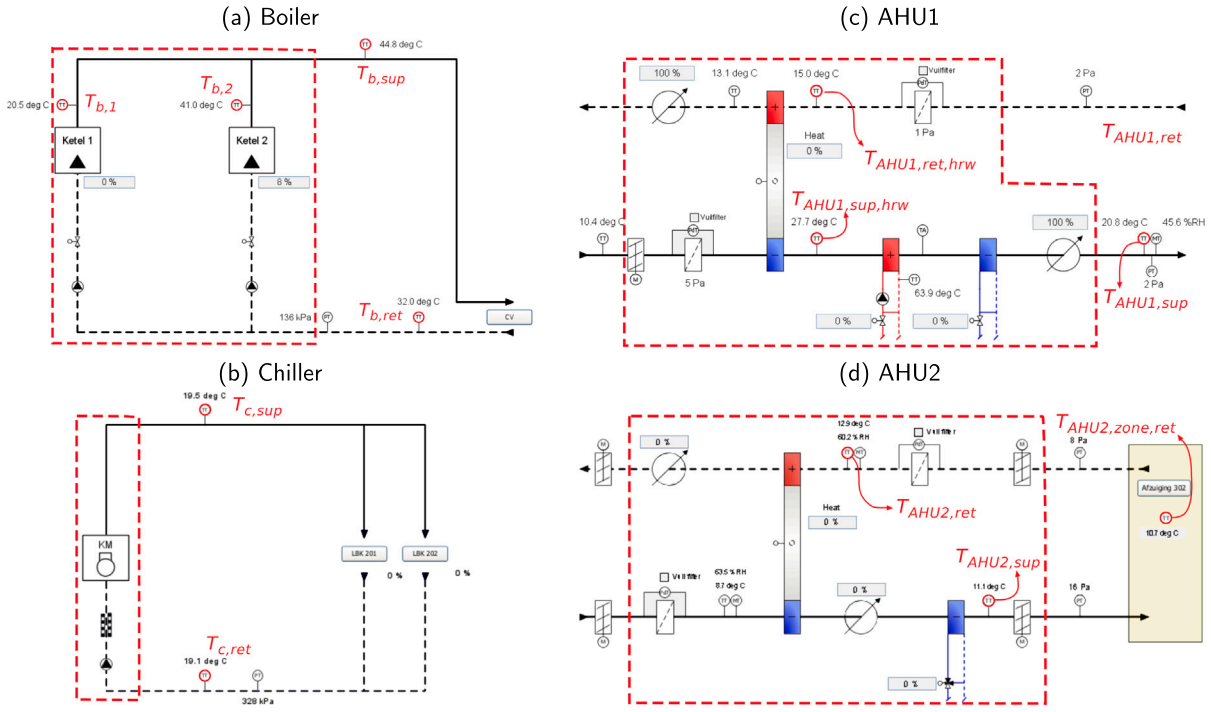

Fig. 4. Annotated representation of the main systems in the case study building's BMS software: (a) boiler, (b) chiller, (c) AHU1, and (d) AHU2. The aggregated system boundaries used in the DBN are shown by the dashed line, while the internal sensors used to detect potential sensor faults are shown individually.

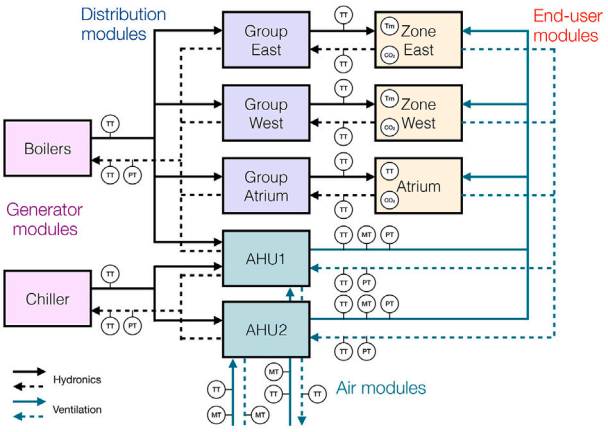

Fig. 5. Simplified schematic representing the main aggregated systems in the building, along with the existing temperature (TT), pressure (PT), moisture (MT) and CO_2 sensors in the case study building.

4.2.1. Fault definition

Each aggregated building system in Fig. 5 is assigned a generic component fault, such as "AHU1 fault". Each system also has its own associated control strategy, and thus a generic control fault is also assigned to each (e.g., "AHU1 Controls" fault). Every sensor used to define a symptom in the DBN is also connected to a sensor fault representing possible bias or drift in the building. Finally, "Zone control" faults are included to represent inadequate use of room-level manual controls, namely the ability of occupants to adjust the radiators and open the windows. These are included in order to allow the incorporation of occupant behavior into the diagnosis method in future studies. The resulting faults are shown as purple nodes in Fig. 6, while the complete list of faults along with their definitions and sample subcomponent faults is summarized in Appendix A.

4.2.2. Symptom definition

Symptoms are defined by assessing which sensors downstream from any given faulty piece of equipment would be affected by the fault. Every sensor shown in Fig. 5 is considered a potential source of evidence of a fault in the building system. The resulting list of symptoms along with their definitions is summarized in Appendix B.

A symptom is detected when a given sensor value deviates from its expected value by more than a threshold ε_i. For example, a boiler supply temperature symptom is detected if during a given timestep:

$$\left|T_{b,\mathrm{sup}}(t) - T_{b,set}(t)\right| > \varepsilon_{b,\mathrm{sup}} \tag{2}$$

where $T_{b,\mathrm{sup}}$ and $T_{b,set}$ are the boiler supply and setpoint temperatures and $\varepsilon_{b,\mathrm{sup}}$ is the symptom detection threshold.

In addition to these operational state symptoms, which are directly related to system controls, some symptoms are solely related to the building's comfort performance based on reference values in the building documentation, even if the systems are not directly controlled to meet these targets. For example, if the relative humidity in the AHU n return airflow $RH_{n,ret}$ is not within the expected range of values for indoor comfort while the relative humidity on the supply side $RH_{n,\mathrm{sup}}$ is within this range, a symptom is also detected.

The final type of symptom considered is designed to detect possible bias or drift in the sensors used for our DBN. Selected rooms in the building are equipped with LoRa sensors measuring air temperature, humidity and CO_2 concentration at occupant level, which we use to check whether the CO_2 concentrations being measured by the ventilation control system are correct. Additionally, some sensors within the aggregated system are used to detect certain sensor faults. For example, during normal operation the supply temperature from the boiler group $T_{b,\mathrm{sup}}$ should be the same as the supply temperature from the boiler j (boiler 1 or boiler 2) that is actually in operation $T_{b,j}$ (Fig. 4a). Therefore, a deviation in these readings is also evidence of a system malfunction. For systems without similar internal sensors (boiler return side, AHU2 supply side, chiller and hydronic groups), the deviation between supply and return temperatures when the systems are off is used to detect sensor symptoms. In

Table 2
Knowledge-based thresholds defined for the first DBN symptom detection scenario. Thresholds marked as [*] were defined based on the building documentation, including HVAC system control documents, electrical installation plans, etc. For each sensor symptom threshold $\varepsilon_{x,s} \in \{\varepsilon_{b,\text{sup},s}, \varepsilon_{b,ret,s}, \varepsilon_{c,s}, \varepsilon_{P,n,s}, \varepsilon_{n,\text{sup},s}, \varepsilon_{n,ret,s}, \varepsilon_{dist,i,s}\}$ was defined based on the corresponding component symptom $\varepsilon_x \in \{\varepsilon_{b,\text{sup}}, \varepsilon_{b,ret}, \varepsilon_c, \varepsilon_{P,n}, \varepsilon_{n,\text{sup}}, \varepsilon_{n,ret}, \varepsilon_{dist,i}\}$.

Threshold	Knowledge-based value	Source
$\varepsilon_{b,\text{sup}}$, $\varepsilon_{b,set}$	3 K	[11]
$\varepsilon_{b,ret}$	20 K	[*]
ε_{Pb}	90 kPa	[*]
$\varepsilon_{c,\text{sup}}$	1 K	[11]
$\varepsilon_{c,ret}$	6 K	[*]
ε_{Pc}	40 kPa	[*]
$\varepsilon_{dist,i,\text{sup}}$, $\varepsilon_{dist,i,ret}$	3 K	[11]
$\varepsilon_{n,\text{sup}}$, $\varepsilon_{n,set}$	1 K	[11]
$\varepsilon_{T,\text{min}}$, $\varepsilon_{T,\text{max}}$	19 °C, 25 °C (heating)	[*]
	22 °C, 27 °C (cooling)	[*]
$\varepsilon_{P,AHU1,\text{sup}}$	185 Pa, 200 Pa	[*]
$\varepsilon_{P,AHU1,ret}$		
$\varepsilon_{P,AHU2,\text{sup}}$	177 Pa	[*]
$\varepsilon_{P,AHU2,ret}$		
$\varepsilon_{rh,\text{min}}$, $\varepsilon_{rh,\text{max}}$	30%, 70%	[*]
$\varepsilon_{CO2,\text{min}}$, $\varepsilon_{CO2,\text{max}}$	360 ppm, 3000 ppm	[34]
$\varepsilon_{CO2,comfort}$	1200 ppm	[*]
$\varepsilon_{x,s}$	$0.5 \cdot \varepsilon_x$	[42]

order to avoid transient effects, any given system should be off for 3 hours before any sensor symptoms can be detected.

4.2.3. Symptom detection thresholds

Symptom detection thresholds are typically defined using domain knowledge and expert experience, but can also be derived from historical data [32]. In this paper, we considered both: on one hand, we tested fixed thresholds based on expert knowledge and building documentation; on the other, we assessed the use of thresholds based on historical data.

The fixed thresholds used in the first case, based on literature and building documentation, are summarized in Table 2. Cooling system temperature symptom thresholds are lower than those for the hot water circuits due to their stronger effect on cooling system performance [13]. Finally, due to the need for accurate measurements for all building controls, temperature sensor deviation thresholds were the most stringent. Here, we assumed temperature sensor deviation thresholds to be 50% lower than the corresponding system thresholds.

General knowledge-based assumptions can provide a good initial estimate and a comparable benchmark for DBNs in different buildings in the absence of further information. However, control strategies can change over time, such that these initial assumptions might not be representative of a system's current state.

In order to avoid these deviations between expected and actual building operation, thresholds can also be assigned based on historical data. For example, Li et al. [35] set thresholds equal to three times the standard deviation of the fault-free historical data. While the use of data-driven thresholds might better represent actual building operation, it can also hide the effects of sensor bias in historical data. Therefore, the performance of data-driven thresholds based on different standard deviations is compared to reference literature values to assess the effects of different thresholds on fault isolation performance. In our study, we tested three different data-driven thresholds and compared them to the fixed, knowledge-based reference values: $\varepsilon_i = 0.5 \cdot \sigma_i$, $\varepsilon_i = \sigma_i$ and $\varepsilon_i = 3 \cdot \sigma_i$. Therefore, for example, the boiler supply temperature symptom threshold in the 3-sigma case is calculated as follows:

$$\varepsilon_{b,\text{sup}} = 3 \cdot \sigma \left(|T_{b,\text{sup}} - T_{b,set}| \right) \tag{3}$$

4.2.4. Causal relationships

Having identified all possible faults and symptoms, their causal relationships are translated into a DBN, as shown in Fig. 6. Even though the faults are defined at the aggregated system level, the resulting DBN is complex and highly interconnected, with a total of 43 fault nodes, 50 symptom nodes and 177 connecting edges. For readability, sensor faults are grouped as one node in this figure, although in reality each sensor fault is an individual node. Similarly, zone-level faults are grouped for readability. The relationship between symptoms and control faults is defined based on the documentation for the building automation systems. For example, a CO_2 sensor fault will lead to an AHU control symptom, since the AHUs are controlled by the CO_2 sensors in each room in the building.

4.3. Parameter modeling

Once the DBN structure and relationships between fault and symptom nodes have been defined, the DBN parameters need to be assigned. These are, namely: the fault node prior probabilities, which define the frequency of a given fault in building systems ($P(F_i)$ in Eq. (1)); and the conditional probabilities that define the likelihood that a symptom is detected when a given fault occurs ($P(S_j|F_i)$). The model parameters considered for each of the evaluation cases for the DBN presented here are summarized in Table 3.

4.3.1. Prior probabilities

Fault node prior probabilities define the frequency of a given fault occurring in a building system. Most real-world BMS data do not include labels indicating faulty and fault-free building system operation. Therefore, assigning the probability of a given fault occurring in a data-driven way is difficult, so prior probabilities are more often estimated by experts [9].

In order to verify the effect of model parameter assumptions on the fault isolation performance of DBNs, we carried out a sensitivity analysis of prior probabilities. We considered all possible combinations of the following prior probabilities [34]:

- Component fault probability: 0.5%, 2%, 5%, 10%
- Control fault probability: 1%, 5%, 10%
- Sensor fault probability: 0.5%, 2%, 5%, 10%

Fault detection was carried out for each combination of the above probabilities and the effects on three performance metrics were assessed.

4.3.2. Conditional probabilities

Conditional probabilities are the likelihoods of a symptom being present given that a fault is present and are usually contained in conditional probability tables (CPT). As previously discussed regarding prior probabilities, however, there are generally insufficient reference data to define CPTs. Furthermore, a significant challenge is the exponential increase in the number of parameters in the CPTs as the number of parent nodes increases [43]. Studies have shown that it is unreliable to directly generate the conditional probabilities for each state of the evidence node when there are more than four fault nodes [7].

In order to overcome these issues, studies often use Noisy-OR and Noisy-MAX gates to reduce the number of parameters that have to be estimated [10]. By assuming symptom independence given the parent faults, Noisy-OR and Noisy-MAX gates reduce the number of parameters required to specify a CPT from logarithmic to linear [44]. In this paper, all faults and symptoms were defined as binary variables ("present" vs. "absent") and therefore the Noisy-OR gate was used to define all CPTs.

Here, again, three different scenarios were considered in order to assess the performance of the DBN under different assumptions. In the first case ("Uniform"), a uniform conditional probability that the child nodes will be false if a parent node is false of 95% is assumed [13]. In addition, a $LEAK$ probability (i.e., the probability that a symptom is detected when the corresponding fault is absent) of 0% was assumed.

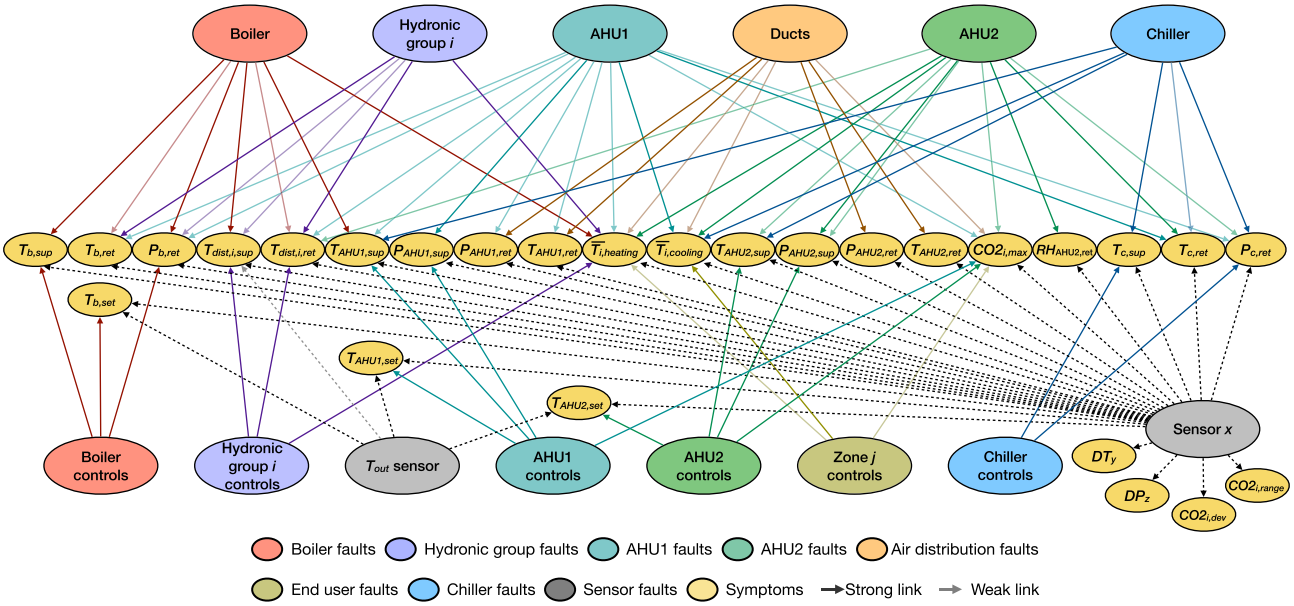

Fig. 6. DBN for the case study HVAC system. System and control faults are color-coded by the system group they represent. Sensor fault edges are shown as dotted lines to improve readability. Nodes including the subscript i indicate that there are actually three nodes: east, west, and atrium. Subscript j indicates two nodes: east and west. Subscript x replaces the name of every sensor used as a symptom. DT_y and DP_z replace the names of the temperature sensors ($DT_{b,\text{sup}}$, $DT_{b,ret}$, DT_c, $DT_{AHU1,\text{sup}}$, $DT_{AHU1,ret}$, $DT_{AHU2,\text{sup}}$, $DT_{AHU2,ret}$) and pressure sensors (DP_{AHU1} and DP_{AHU2}) that have dedicated deviation symptoms. Individual representations of the DBN structure for each system in the whole-HVAC DBN including all zones and sensor faults are shown in Appendix C. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Model parameters assumed for DBN evaluation based on historical and experimental data and operational testing. Knowledge-based parameters were assumed during historical evaluation, while all parameter assumptions were assessed as part of a sensitivity analysis during the experimental evaluation. Finally, the best parameters were selected for operational testing during a suspected fault.

Model parameter case	Assumption	Evaluation case		
		Historical	Experimental	Operational
Symptom detection threshold				
Knowledge-driven	Table 2	✓	✓	
Data-driven	$0.5 \cdot \sigma_i$		✓	
	σ_i		✓	✓
	$3 \cdot \sigma_i$		✓	
Prior probabilities				
Component and sensor	0.5%		✓	
fault probabilities	2%	✓	✓	✓
	5%		✓	
	10%		✓	
Control fault probability	1%		✓	
	5%	✓	✓	✓
	10%		✓	
Conditional probabilities				
Uniform	95%	✓	✓	✓
Weighted	Strong: 90%		✓	
	Weak: 30%		✓	
P&ID	Eq. (4)		✓	
Fault isolation timescale				
Subhourly		✓		
Hourly mode		✓	✓	✓
Hourly threshold		✓		
Daily		✓		

In the second case ("Weighted"), strong and weak links between faults and symptoms were defined [7], as shown in Fig. 6. For each symptom, the most likely component and/or control faults were selected as strong links, with the remaining relationships being defined as weak. Sensor faults were assumed to be strong in all cases. Strong conditional probabilities were set to 90% and weak conditional probabilities were set to 30% for all cases [45].

In the final case ("P&ID"), the conditional relationships between faults and symptoms were defined based on the P&ID in Fig. 5. For every symptom, the most likely fault was assumed to be the component directly preceding it in the P&ID, followed by the component before that, and so on. For any given symptom S_j, the conditional probability that it was caused by fault F_i was assumed to be:

$$P\left(S_j|F_i\right) = \left(P_{strong}\right)^i \tag{4}$$

where P_{strong} is the conditional probability for a strong link (90%) and $i \in [1, 2, \ldots]$ is the distance from the potential fault F_i to the symptom S_j. Therefore, for example, for a Zone East temperature symptom $T_{dist,east,\text{sup}}$, the component "Group East" was assumed to be the most likely cause (90% conditional probability), followed by the "Boiler" group (81%).

4.4. Fault isolation and DBN evaluation

The final step in the FDD process is the diagnosis of the faults leading to the observed symptoms. In order to do so, the DBN structure and model parameters are implemented using the Discrete Bayesian Network model in *pgmpy*, a Python package for causal inference and probabilistic inference using Directed Acyclic Graphs (DAGs) and Bayesian Networks [46]. Given the symptom detection evidence obtained from the BMS data, the resulting fault probabilities are calculated by exact inference using the Variable Elimination algorithm in *pgmpy*. A fault is assumed to be present if the posterior probability is at least 50%.

The data collection frequency in the BMS system changes from sensor to sensor. Before symptom detection could take place, these data were resampled to match the lowest temporal resolution in the dataset, that is, a 10-minute frequency was used for all measurements. For numerical variables, the data were resampled by taking the mean of higher frequency measurements, whereas for categorical variables the mode was used. Control signals were only recorded when their status changed;

therefore these were forward-filled to ensure the current signal was used. Sensors with constant readings over the three-year historical period were assumed to be faulty and were thus dropped from the dataset. Any remaining missing samples were excluded from the analysis.

4.4.1. Detection and diagnosis timescale

Symptom detection and fault isolation can be carried out at different timescales. Here, three different symptom detection timescales were considered: subhourly detection (i.e., using all samples in the dataset, at 10-minute frequency), hourly and daily. The use of all samples in the dataset can easily lead to false positives due to transient effects such as when systems are starting up and have not yet reached their setpoint. In order to avoid false positives, hourly and daily symptom detection strategies were defined. In the "hourly mode" case, the mode of all subhourly symptom states was taken as the hourly symptom state (i.e., for a given hour, if a symptom was present more often than it was absent, the symptom was considered present in that hour). For the second hourly detection case ("hourly thresholds"), symptom detection thresholds were selected, whereby a symptom was only considered present if it was detected for two timesteps in a row, at least three times during that hour [45]. The same thresholds were used for daily detection ("daily thresholds"). Similarly, fault isolation at hourly and daily timescales was tested. Fault isolation was ultimately carried out on an hourly basis, as it is a timescale short enough for building operators to take action in real-time applications, but also long enough to minimize transient effects for the faults considered.

4.4.2. DBN evaluation

The DBN's ability to isolate faults is evaluated, partially validated, and tested on a real-world application (Table 3). The first evaluation is carried out using historical data from the years 2022–2023 to detect long-term trends in the building's HVAC systems. This data was also used to define the data-driven thresholds presented in Section 4.2.3. In the absence of fault labels, it is impossible to verify the performance of the DBN using this data, but this analysis can demonstrate whether the model can diagnose different faults and shed light on the most frequent faults in the case study building.

The second evaluation is then carried out using data from experiments carried out in the building's AHUs in 2024. These data are used for partial model validation and to assess the performance of the whole-HVAC DBN in AHU fault isolation using labeled data. The experiments and performance metrics used are described in the following subsections.

Finally, in the operational testing, a real-world application of the DBN during a suspected fault is demonstrated. In early September 2024, a surge in occupant complaints prompted facility management to investigate a potential issue with the ventilation systems in the building. The DBN was run for the period 1–10 September 2024 in order to verify whether the model was able to detect this issue and isolate its root cause.

4.4.3. Fault experiments for performance assessment

The whole-HVAC system DBN's performance was tested using experimental data collected over six days in 2024. Different fault combinations were introduced into the controls for different subcomponents in each AHU module, as summarized in Table 4. In each of the fault cases shown, the normal operation of the AHUs was modified via the systems' controls to mimic a specific fault in the ventilation systems. However, since faults were injected via the control systems and control signals were used during symptom detection, all faults were detected as control faults.

Due to restrictions from the building operator and safety concerns, faults could only be introduced in the AHUs. Fault injection in other components, such as boilers and chillers, could therefore not be tested. Nevertheless, the experimental data are useful in verifying the DBN's ability to detect AHU faults in different subcomponents and to isolate the faults in individual AHUs. Furthermore, the experimental data can be used to assess the cascading effects of AHU malfunction on other systems and whether the DBN can successfully identify the AHUs as the root cause of symptoms detected in other systems.

4.4.4. Assessment metrics

In this research, we selected three of the most common performance metrics used to evaluate DBNs for building systems [47]. The first is the Classification Rate (CR), also known as accuracy [48–50], correct rate [51] and correct diagnosis rate [38]:

$$CR = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives. This metric gives an indication of the model's overall performance, combining its ability to correctly detect faulty and normal behavior. Therefore, the higher the CR, the better the network performs.

The second metric is the Miss Rate (MR), also known as missed detection rate [52] and false negative rate [38]:

$$MR = \frac{FN}{TP + FN} \tag{6}$$

This metric provides information on whether the DBN is unable to correctly diagnose faults in the system. A high MR would lead to faults in a building system not being addressed in a timely manner, leading to inadequate system performance.

Finally, we consider the False Alarm Rate (FAR), also called false alarm ratio [26] and false positive rate [38]:

$$FAR = \frac{FP}{FP + TN} \tag{7}$$

In contrast to the MR, this metric assesses whether the DBN raises nonexistent issues during normal behavior. A high FAR would therefore imply that maintenance might be sent to check systems during normal performance. Over the long run, this can lead to real detected faults being ignored under the assumption that a false alarm is being raised.

In addition to these metrics common in DBN evaluation, we have considered three further metrics commonly used in data-driven FDD applications. These are Precision (Pr), F1-score (F_1) and Balanced Accuracy (BA), which are calculated as follows:

$$Pr = \frac{TP}{TP + FP} \tag{8}$$

$$F_1 = \frac{2 \cdot Pr \cdot (1 - MR)}{Pr + (1 - MR)} \tag{9}$$

$$BA = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{10}$$

Another common metric in data-driven FDD is Recall, which is simply equal to $(1 - MR)$, so it is omitted from our discussion.

5. Results

5.1. Historical evaluation

BMS data for a two-year period (2022–2023) were used to detect symptoms in the whole HVAC system. For simplicity, at this stage the simplest form of the DBN, with uniform conditional probabilities and knowledge-based symptom thresholds and prior probabilities, was considered (Table 3).

5.1.1. Symptom detection using historical data

The frequency of detection for every symptom considered during this period at each of the symptom detection timescales is shown in Fig. 7. In all cases, the share of samples in which a given symptom was detected was highest for the daily scale. This is due to the detection threshold

Table 4
List of fault experiments conducted in the AHUs in the case study building. Faults were introduced in the control systems for the AHU fans, heat recovery wheels (HRW), heating coil valves (HCV) and in the pressure and temperature setpoints. Given that the experiments were injected into the control systems, all faults were detected as control faults.

Code	Date	Time	AHU1 fault type (Experimental setup)	AHU2 fault type (Experimental setup)
A	23/03/2024	8:21 – 11:14	Supply fan control fault (Fan speed fixed at 30%)	Supply fan control fault (Fan speed fixed at 30%)
B		11:14 – 14:32	Supply fan control fault (Fan speed fixed at 70%)	Supply fan control fault (Fan speed fixed at 70%)
C		14:33 – 17:00	HRW control fault (Rotation speed fixed at 0%)	HRW control fault (Rotation speed fixed at 0%)
D		17:00 – 20:00	HRW control fault (Rotation speed fixed at 30%)	HRW control fault (Rotation speed fixed at 30%)
E	24/03/2024	8:02 – 11:05	HRW control fault (Rotation speed fixed at 70%)	HRW control fault (Rotation speed fixed at 70%)
F		11:05 – 13:58	Air pressure setpoint too high (Setpoint increased by 50 Pa)	Air pressure setpoint too high (Setpoint increased by 50 Pa)
G		13:58 – 17:00	Air pressure setpoint too low (Setpoint decreased by 50 Pa)	Air pressure setpoint too low (Setpoint decreased by 50 Pa)
H		17:00 – 20:00	Air pressure setpoint constant (Air pressure set to 130 Pa)	Air pressure setpoint constant (Air pressure set to 120 Pa)
I	29/03/2024	9:25 – 11:24	HCV control fault (HCV fixed at 0% open)	HRW control fault (Rotation speed fixed at 0%)
J		11:24 – 14:00	HCV control fault (HCV fixed at 30% open)	HRW control fault (Rotation speed fixed at 100%)
K		14:00 – 17:12	HCV control fault (HCV fixed at 100% open)	HRW control fault (Rotation speed fixed at 50%)
L		17:12 – 20:00	Air pressure setpoint fixed (Air pressure set to 230 Pa)	Air pressure setpoint fixed (Air pressure set to 250 Pa)
M	30/03/2024	9:23 – 10:56	Air pressure setpoint constant (Air pressure set to 200 Pa)	Air pressure setpoint constant (Air pressure set to 250 Pa)
N		11:22 – 14:35	Supply temperature setpoint too high (Setpoint increased by 3 °C)	Supply temperature setpoint too high (Setpoint increased by 3 °C)
O		14:35 – 17:59	Supply temperature setpoint too low (Setpoint decreased by 3 °C)	Supply temperature setpoint too low (Setpoint decreased by 3 °C)
P		18:10 – 20:00	Supply temperature setpoint constant (Set to 17 °C)	Supply temperature setpoint constant (Set to 15 °C)
Q	31/03/2024	8:44 – 12:00	Simultaneous HCV and HRW control faults (HCV set to 10% and HRW set to 10%)	HRW control fault (Rotation speed set to 10%)
R		12:00 – 16:30	Simultaneous HCV and HRW control faults (HCV set to 90% and HRW set to 10%)	Return fan control fault (Fan speed set to 0%)
S		16:30 – 19:48	Simultaneous HCV and HRW control faults (HCV set to 10% and HRW set to 90%)	Return fan control fault (Fan speed set to 100%)
T	01/04/2024	9:46 – 12:30	Simultaneous HCV and HRW control faults (HCV set to 90% and HRW set to 90%)	Return fan control fault (Fan speed set to 50%)
U		12:30 – 18:30	Simultaneous HCV and HRW control faults (HCV set to 50% and HRW set to 50%)	Return fan control fault (Fan speed set to 50%)

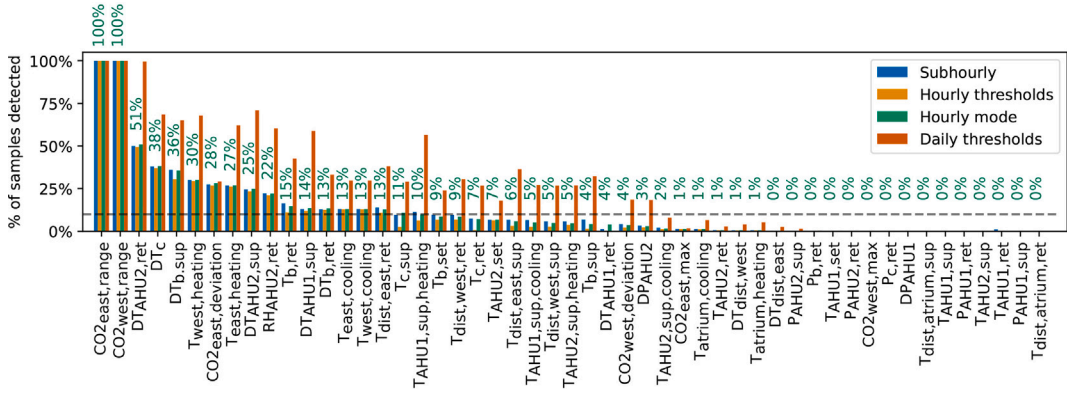

Fig. 7. Symptom detection frequency over two years for four symptom detection timescales. The percentage of samples where each symptom was detected for the chosen timescale ("Hourly mode") is shown in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

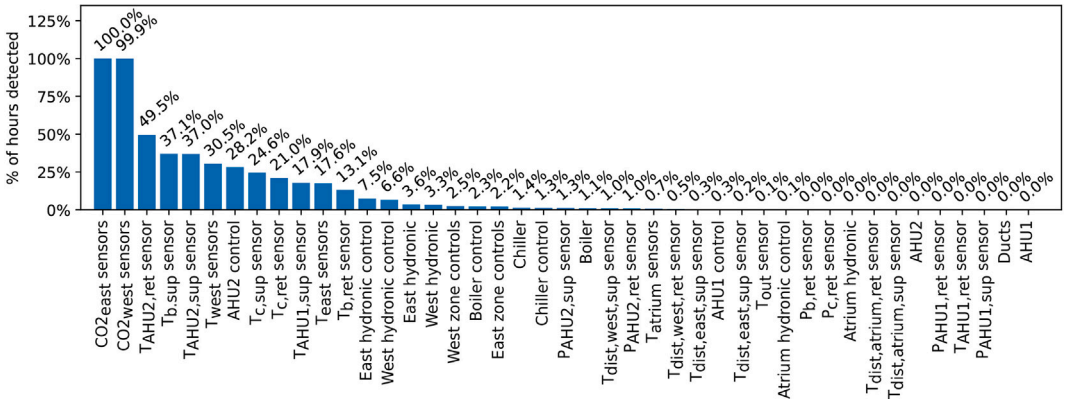

Fig. 8. Percentage of hourly samples where each fault was detected during the period 2022–2023. Symptom detection based on the hourly mode.

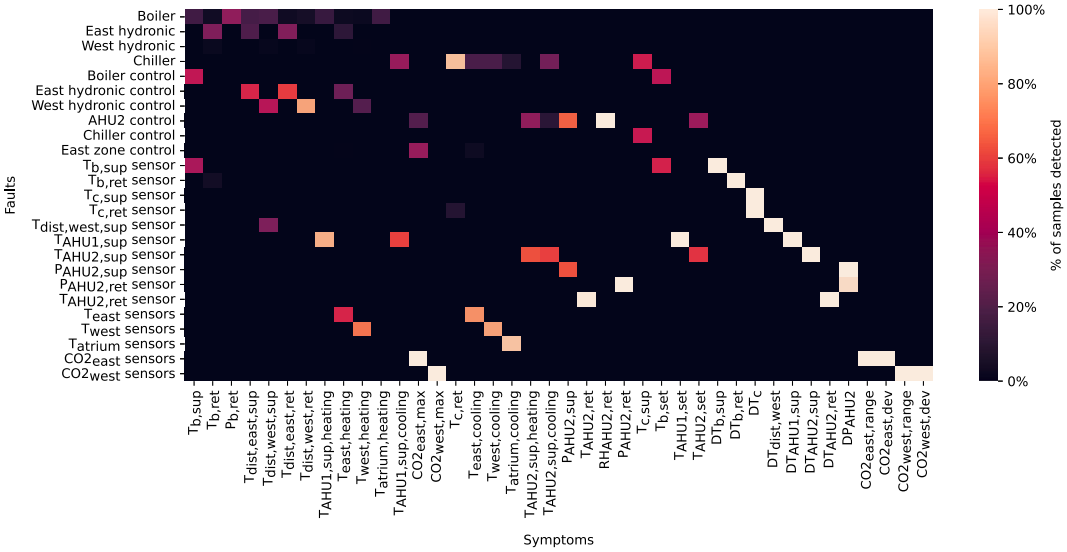

Fig. 9. Percentage of samples in which a fault was detected and each symptom associated with that fault was also detected.

used: it is more likely for a symptom to be detected for two timesteps in a row three times during a day than during an hour. Symptom detection frequency at subhourly and hourly scales was comparable, with the hourly mode generally giving a slightly higher detection rate than the hourly thresholds. Based on these results, the hourly scale was selected for symptom detection, since it achieved similar results to the subhourly scale while significantly reducing the computational expense.

The symptom detection results show that the CO_2 sensor symptoms were detected 100% of the time at all time scales. This is due to a number of sensors in the building giving readings below the lower threshold for CO_2 detection as defined in Table 2. Since the DBN considers aggregate systems, the CO_2 readings are considered at zone scale, so a fault in any sensor in the zone triggers a CO_2 sensor symptom.

The next most common symptoms are also sensor-related: the AHU2 return temperature sensor symptom ($DT_{AHU2,ret}$), chiller sensor symptom (DT_c) and boiler supply sensor symptom ($DT_{b,\mathrm{sup}}$). AHU2 supply temperature sensor symptoms and CO_2 sensor deviation symptoms in Zone East are also often detected. The frequency of these symptoms is indicative of building sensors requiring maintenance and calibration, as they are no longer providing reasonable values.

This is confirmed by analyzing the hourly data for the corresponding sensors, which show a number of inconsistencies that justify the detected sensor faults. A more detailed analysis of the aforementioned discrepancies found in the CO_2 and temperature measurements leading to the most common sensor symptoms is found in the Supplementary Material.

5.1.2. Fault detection using historical data

Fig. 8 shows the fault detection results for the period 2022–2023 based on symptom detection using the hourly mode. Similar to symptom detection, eleven of the top twelve most commonly-detected faults are sensor-related. In particular, incorrect readings in the CO_2 sensors in the East and West zones are detected 100% of the time. Given that the building's ventilation systems are controlled based on the CO_2 concentration in the rooms, it is not surprising that the only other top detected fault is related to the control of one of the AHUs. This might also explain why AHU2 return relative humidity symptoms ($RH_{AHU2,ret}$) are among the most-often detected. Since AHU2 return and supply temperature and boiler temperature sensor symptoms were also among the most common, these sensors are also detected at high frequency (49.5%, 37.1% and 37% of the time, respectively).

Among non-sensor related symptoms, deviations in the expected indoor temperatures in the heating season (and, to a lesser extent, in the cooling season) were also common, as were boiler and hydronic system return temperature symptoms. As a result, hydronic control faults and hydronic system faults were occasionally also encountered.

The large number of sensor faults detected in most hours can obscure faults occurring in the systems and controls at the same time. Fig. 9 shows every fault that was detected in the period 2022–2023 along with

Table 5
Performance metrics for detection of AHU faults for the three different conditional probability cases ("Uniform", "Weighted" and "P&ID") and for different symptom detection thresholds ϵ_i (literature-based fixed thresholds, "$0.5 \cdot \sigma$", "σ" and "$3 \cdot \sigma$") for a fault detection probability threshold of 50%. Values are shown as the mean (and standard deviation) of all runs of the model with different component, control and sensor probability assumptions. The best-performing version of the model for each metric is shown in bold, while the worst-performing is shown in italics.

Conditional probabilities	Symptom thresholds	TP	TN	FP	FN	CR [%]	MR [%]	FAR [%]	Pr [%]	F_1 [%]	BA [%]
Uniform	Fixed	57.2	150.0	10.0	22.8	86.3	28.5	*6.2*	*85.1*	77.7	82.6
		(0.4)	(0.0)	(0.0)	(0.4)	(0.2)	(0.5)	*(0.0)*	*(0.1)*	(0.3)	(0.2)
	$0.5 \cdot \sigma$	58.0	151.0	9.0	22.0	87.1	**27.5**	5.6	86.6	78.9	83.4
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	**(0.0)**	(0.0)	(0.0)	(0.0)	(0.0)
	σ	56.7	154.0	6.0	23.3	**87.8**	29.2	3.8	90.4	**79.4**	**83.5**
		(0.5)	(0.0)	(0.0)	(0.5)	**(0.2)**	(0.6)	(0.0)	(0.1)	**(0.4)**	**(0.3)**
	$3 \cdot \sigma$	40.0	157.0	3.0	40.0	82.1	50.0	**1.9**	**93.0**	65.0	74.1
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	**(0.0)**	**(0.0)**	(0.0)	(0.0)
Weighted	Fixed	54.8	151.3	8.7	25.2	85.9	31.6	5.4	86.3	76.3	81.5
		(1.4)	(0.5)	(0.5)	(1.4)	(0.4)	(1.8)	(0.3)	(0.4)	(1.0)	(0.8)
	$0.5 \cdot \sigma$	54.1	151.3	8.7	25.9	85.6	32.4	5.4	86.2	75.8	81.1
		(1.8)	(0.5)	(0.5)	(1.8)	(0.6)	(2.3)	(0.3)	(0.4)	(1.4)	(1.0)
	σ	52.4	154.0	6.0	27.6	86.0	34.5	3.8	89.7	75.7	80.9
		(1.9)	(0.0)	(0.0)	(1.9)	(0.8)	(2.4)	(0.0)	(0.3)	(1.7)	(1.2)
	$3 \cdot \sigma$	40.0	157.0	3.0	40.0	82.1	50.0	**1.9**	**93.0**	65.0	74.1
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	**(0.0)**	**(0.0)**	(0.0)	(0.0)
P&ID	Fixed	44.8	154.6	5.4	35.2	83.1	44.0	3.4	89.9	67.6	76.3
		(12.8)	(2.7)	(2.7)	(12.8)	(4.2)	(16.0)	(1.7)	(2.3)	(11.7)	(7.2)
	$0.5 \cdot \sigma$	44.1	154.6	5.4	35.9	82.8	44.9	3.4	89.7	66.9	75.9
		(12.4)	(2.7)	(2.7)	(12.4)	(4.0)	(15.5)	(1.7)	(2.4)	(11.4)	(6.9)
	σ	43.0	155.7	4.3	37.0	82.8	46.2	2.7	91.2	66.6	75.5
		(11.0)	(1.8)	(1.8)	(11.0)	(3.9)	(13.8)	(1.1)	(1.6)	(10.7)	(6.4)
	$3 \cdot \sigma$	36.5	157.0	3.0	43.5	*80.6*	*54.4*	**1.9**	92.3	*61.0*	*71.9*
		(3.7)	(0.0)	(0.0)	(3.7)	*(1.5)*	*(4.6)*	**(0.0)**	(0.7)	*(4.3)*	*(2.3)*

the percentage of the time steps in which each fault was detected and in which a given symptom was also detected. As seen in the previous results, the majority of faults encountered are sensor faults, and for most of those detected faults a corresponding sensor symptom was also detected at the same time. However, other non-sensor related symptoms were also often detected at times when sensor faults were detected. For example, a high share of the AHU2 return pressure and temperature sensor faults was also accompanied by the detection of AHU2 return pressure and temperature symptoms. These may actually be symptoms of a malfunctioning AHU or control system, but the DBN diagnoses a sensor fault due to the concurrent sensor symptoms.

5.2. Partial experimental evaluation

The DBN's performance was tested by carrying out fault detection on a dataset collected during fault experiments in the building's AHUs. Due to the aforementioned constant issues with sensors in the building, the majority of the detected faults are sensor-related. Therefore, in order to correctly assess the component and control faults in the DBN, the results presented in the rest of this section exclude sensor faults. If AHU faults can be accurately diagnosed, then DBN can be expected to perform acceptably in spite of the building's malfunctioning sensor infrastructure.

5.2.1. DBN performance evaluation

Table 5 shows the performance metrics for the different symptom threshold assumptions and conditional probability cases summarized in Table 3. Each possible combination of component, sensor and control fault prior probabilities was tested, and the resulting range of results for each performance metric is shown as the mean and standard deviation for each metric. Given that labeled fault data is only available for the AHUs, true positives and false negatives were only calculated for faults in the AHU controls themselves. Furthermore, simultaneous faults in different AHUs were simply counted as concurrent individual faults. Therefore, if during a simultaneous fault only one of the two faults is detected, one true positive and one false negative will be assigned. Given the level of aggregation of this DBN at system level, simultaneous faults in different subcomponents are considered as a single fault in the AHU.

In general, the choice of prior probabilities for component and control faults barely has any impact on the performance of the DBN. The effect is strongest for the P&ID-weighted scenario, which in any case is the worst performing of the three scenarios. Therefore, the exact values selected for the prior probabilities do not appear to have a strong influence on the fault isolation capabilities of the DBN in most cases. The reason for this is that although prior probability assumptions do affect the predicted posterior fault probabilities, the effect is not strong enough to change the ranking of the assigned faults. Therefore, despite the changes in posterior probabilities, the most likely fault in each case remains the same. Thus, the diagnosis accuracy of the DBN is not affected by the range of prior probabilities assumed, which is consistent with previous findings in the literature [7,33,34].

The uniform and weighted conditional probability cases perform similarly well in terms of CR, with the uniform case performing slightly better in terms of MR and slightly worse in terms of FAR. The P&ID-based case performs worst due to the assumption that the closest system to a given symptom is most likely to cause it. Therefore, symptoms caused by cascading effects from a fault in one system were incorrectly assigned to the system closest to the sensor that detected the symptom. In the final version of the DBN, the uniform conditional probability case was selected due to its better performance in terms of CR and MR while maintaining an adequately low FAR.

Regarding prior probabilities, the symptom threshold of three standard deviations appears to be too large for this DBN. This threshold achieves the lowest FAR but also the highest MR, indicating that the threshold is not stringent enough and therefore allows at least 50% of faults to go undetected. The other symptom thresholds perform similarly in terms of both CR and MR, with the fixed, uniform thresholds performing worst in terms of FAR.

All versions of the model achieve CRs of more than 80%, which is well above the 70% threshold presented by Wang et al. [48]. The DBN's diagnostic performance is better than reported 51–66% rates for an FDD

approach for AHUs [38] and is comparable to some of the models for chiller FDD by other authors [49,51]. While this CR is not as high as some of the best DBN models in the literature (e.g., [49,50]), these are typically single-system DBNs, which carry less uncertainty than whole HVAC system FDD.

Compared to data-driven methods with partially labeled data, the model exceeds the performance of seven different supervised learning methods reported by Li et al. [31] using 1% labeled fault data for rooftop units. Similarly, Fan et al. [28] used graph convolutional networks to detect faults in air-side systems (AHU, CAV and VAV) and achieved lower CRs for all cases except at the lowest imbalance ratio. While a number of fully data-driven approaches have been found to achieve CRs above 95% (e.g., [30,31]), these were achieved for individual component FDD with high shares of labeled data. Therefore, the DBN presented here achieves adequate performance with higher uncertainty and without the need for labeled data for training.

The MR for all cases is high, though the rate of less than 30% achieved in three of the cases is still comparable to single-system FDD approaches in the literature [38,52]. The MR is likely to be strongly affected by the fact that many of the sensors in the building appear to be faulty. On the other hand, the fact that over 70% of the AHU faults could still be detected in spite of the high degree of uncertainty in the sensor readings is an indication of the method's robustness to uncertainty in the input data. The performance of the DBN in this paper is expected to improve if maintenance and calibration of the sensor infrastructure in the building are carried out.

The FAR is fairly low for most scenarios, with two thirds of the tested cases achieving the 5% threshold specified by other authors [26,52]. This means that the DBN does not raise incorrect faults, which would otherwise lead to unnecessary control and maintenance efforts by facility management.

Since the metrics are calculated only for systems for which labeled data was available (AHU1 and AHU2), the experimental data was not dominated by true negatives from other building systems, and therefore Precision remains high for all cases. In particular, the version of the DBN with uniform conditional probabilities and data-driven symptom thresholds of σ achieves a Precision of 90% and F_1 close to 80%, with a balanced accuracy of 83.5%. This combination therefore achieves the most balanced performance in terms of correctly detecting faults and not raising false alarms, achieving the best performance in three of the six metrics considered.

5.2.2. Fault isolation by subcomponent fault type

Fig. 10 shows hourly fault detection results from the selected version of the model (Table 3) for each of the days when an experiment was carried out. Since the DBN uses control signals to detect whether a system is operational or not, the introduced faults are all detected as control faults. This is consistent with reality, as the faults introduced in the experiments were indeed introduced through the control systems, not by introducing physical damage into the subcomponents themselves.

Two main competing fault diagnoses are observed. During experiments F–H (air pressure setpoint faults), an "East hydronic control" fault is detected instead of AHU1 control. The same happens during experiments N–O (supply temperature setpoint faults), when an "East hydronic" fault is also detected. The second major competing fault is "Boiler", which is detected instead of "AHU1 control" both at the start of experiment R and during experiments T–U (simultaneous HCV and HRW control faults). This is due to an unexpected cascading effect on the building control systems, where the experiments carried out in the AHU caused the boiler temperature to increase (Fig. 11). Since AHU2 does not include a heating component, this effect is only observed for AHU1. The same happens during the aforementioned simultaneous fault experiments T–U, when a boiler fault is detected instead of an AHU1 control fault for a number of time steps.

These cascading effects cannot be directly explained based on the building's control logic and demonstrate some of the difficulties of FDD in highly coupled real-world HVAC systems. Furthermore, they cannot immediately be categorized as false positives, since the cause for these cascading effects is unknown and might indeed be related to a pre-existing fault in these systems and their controls. Nevertheless, in order to quantitatively assess DBN performance during each individual experiment, we consider any fault diagnosis made during a false negative as a false positive diagnosis in the other system. In other words, in order to calculate TP, FP, TN and FN, we consider three diagnoses: "AHU1 control" fault, "AHU2 control" fault or a fault in any other component. Therefore, in experiment E we have two correct diagnoses ("AHU1 control" and "AHU2 control"), but for fault F we have one correct diagnosis and one false positive ("AHU2 control" and "East hydronic control").

The resulting performance metrics by subcomponent fault experiment type are summarized in Table 6. While Table 5 evaluates detection at AHU level using the entire monitoring period as ground truth (experiment running vs. not), this is not directly transferable to individual subcomponents, since (i) attributing an AHU-level detection to a specific subcomponent outside its own experimental window is ambiguous; and (ii) using "no experiment" periods as the negative class for every subcomponent would repeatedly count the same true negatives for different experiments. Table 6 therefore restricts the evaluation window for each subcomponent fault to the runs of that specific experiment, and redefines the confusion matrix accordingly: a true positive is a run where an AHU control fault flag is raised (correct detection); a false negative is a run where no fault is flagged (missed detection); a false positive is a run where the anomaly is instead attributed to a different component (misattribution); and a true negative is a run where no other component is falsely flagged. Consequently, the metrics in Table 6 characterize both detection sensitivity and attribution correctness within each experiment, rather than overall false-alarm behavior during fault-free operation. By comparing the detection of "AHU1 control" and "AHU2 control" faults to any other fault in the system, we do not have a large number of consistently negative results for faults that are not detected, and therefore the number of positives and negatives is very balanced.

The DBN performs well for faults in the heating and cooling components, namely the HRW (faults C–E, AHU2 faults I–J and Q) and HCV (faults I–K in AHU1). For the selected parameter combination, these faults are consistently detected in every experiment, achieving a CR of 100%. On the other hand, the model struggles with supply fan faults, where fault A was detected only in AHU1 and fault B was detected in neither. Return fan faults in AHU2 (R–U), on the other hand, are correctly detected, with a CR of 87.5% due to a simultaneous "Atrium hydronic" diagnosis.

The DBN also performs very well for setpoint symptoms in AHU2 (faults F–H and M–P) but struggles to detect the same faults in AHU1. This is especially noticeable for air pressure setpoint symptoms (F–H), for which the CR is only 50%, with an "East hydronic control" fault detected instead of AHU1. As discussed previously, this is not necessarily an actual misdiagnosis, but it is considered one here for this comparison. For temperature setpoint faults, on the other hand, the DBN achieves a high CR of 83.3%, with an "AHU1 control" fault mostly detected, though "East hydronic control" is detected once more.

Regarding fault isolation during simultaneous faults, the model performs very well in detecting simultaneous faults in AHU1 on 31 March, but poorly on 1 April, resulting in a CR of 70%. Given that both AHUs supply the entire building, differentiating between faults in each of them can be challenging. In particular, when considering the aggregated system level, the effects of certain faults might cancel each other out, which might help to explain why the DBN performs better for certain simultaneous fault combinations than others.

5.3. Operational testing during a suspected fault

As a result of the above sensitivity analysis and performance assessment, we selected the best performing combination of parameters for operational testing (Table 3). Literature-based initial prior

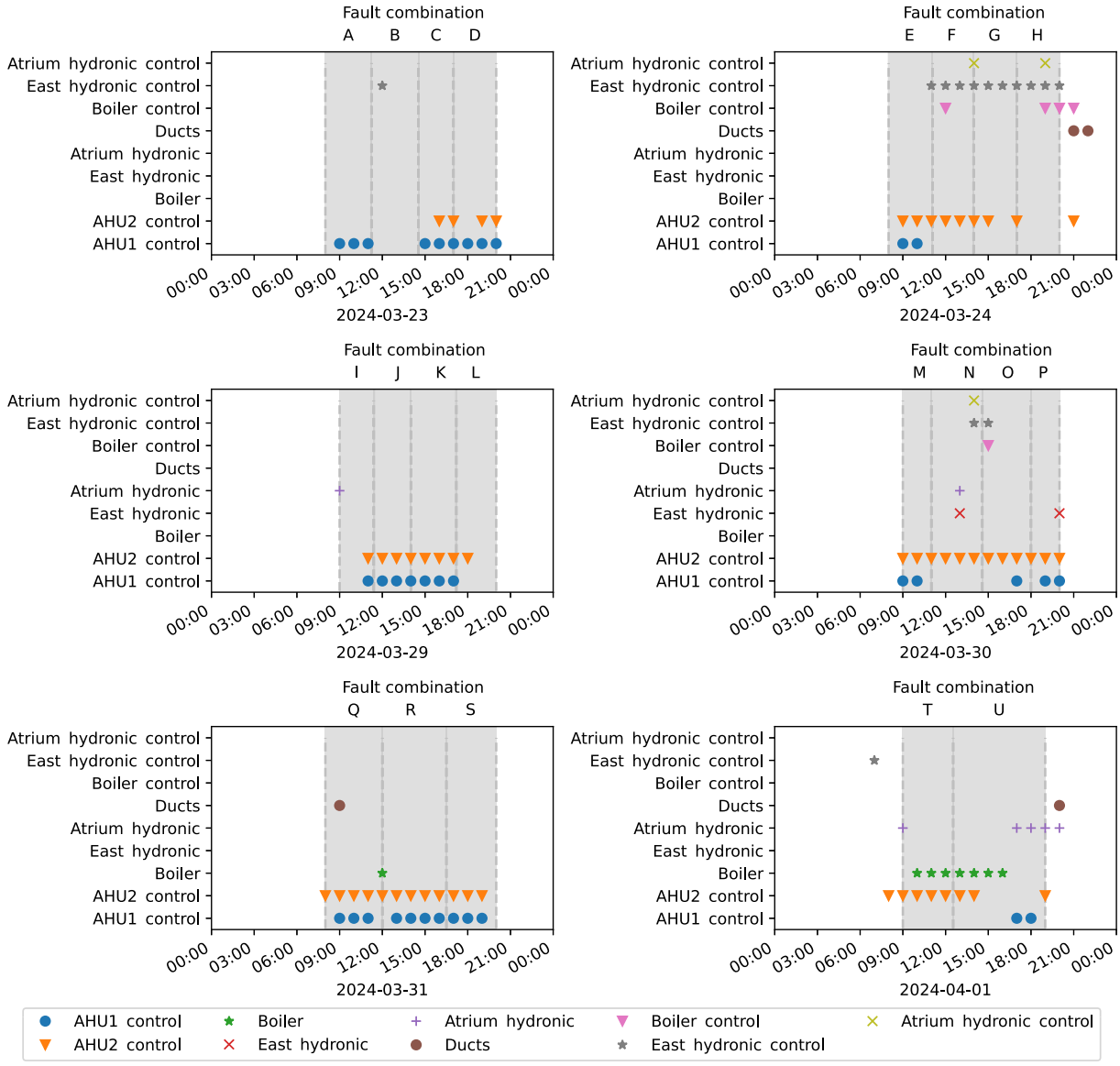

Fig. 10. Hourly fault detection results on each of the days when experiments were conducted for each injected fault combination (A–U, as shown in Table 4) using the best-performing version of the model. Fault-free periods are shown in white, while periods where faults were introduced are shown in gray. Since faults were injected via the control systems and control signals are used during symptom detection, all faults are detected as AHU control faults. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

probabilities were selected, since they had minimal effects on the best-performing models. Uniform conditional probabilities were selected for simplicity, since they require fewer explicit assumptions from modelers and perform best in terms of MR. Finally, data-driven symptom detection thresholds ($\epsilon_i = \sigma(i)$) were assumed due to the lower FAR and higher CR achieved.

The finalized DBN was then tested during a suspected fault in the building systems. Using the BMS data from that time period, we identified which faults were causing the issue to demonstrate how this DBN could be applied in practice. Fig. 12 shows the hourly fault detection results for the first ten days of September 2024. As shown in previous sections, CO_2 sensors are constantly detected as faulty given the fact that some sensors in both the East and West building zones are strongly biased. Given that the ventilation systems in the building are controlled based on the CO_2 concentration in the building, a fault in the sensors is likely to have a significant impact on the indoor environmental quality in the building.

The next most commonly detected fault is in the temperature sensors in AHU2. The fact that the sensors controlling the supply temperature are consistently found to be faulty is likely to have an impact on the building's environmental quality. This is consistent with the occupants' complaints about overheating in the building during this time period.

The next most common fault is with the AHU2 controls, which are found to be faulty on most days within the time period under study. While the supply air from both AHUs is mixed before delivery in all three zones of the building, occupant comfort issues appear to be caused by a problem with the controls of AHU2 rather than AHU1. Thus, aside from a complete overhaul of the building's sensing infrastructure, facilities management should prioritize the AHU2 controls as a likely cause of discomfort in the case study building.

6. Discussion

The DBN presented here represents one of the few whole-HVAC system implementations in the literature. Unlike previous work (e.g., [6,7]), it includes heating, cooling and ventilation systems along with

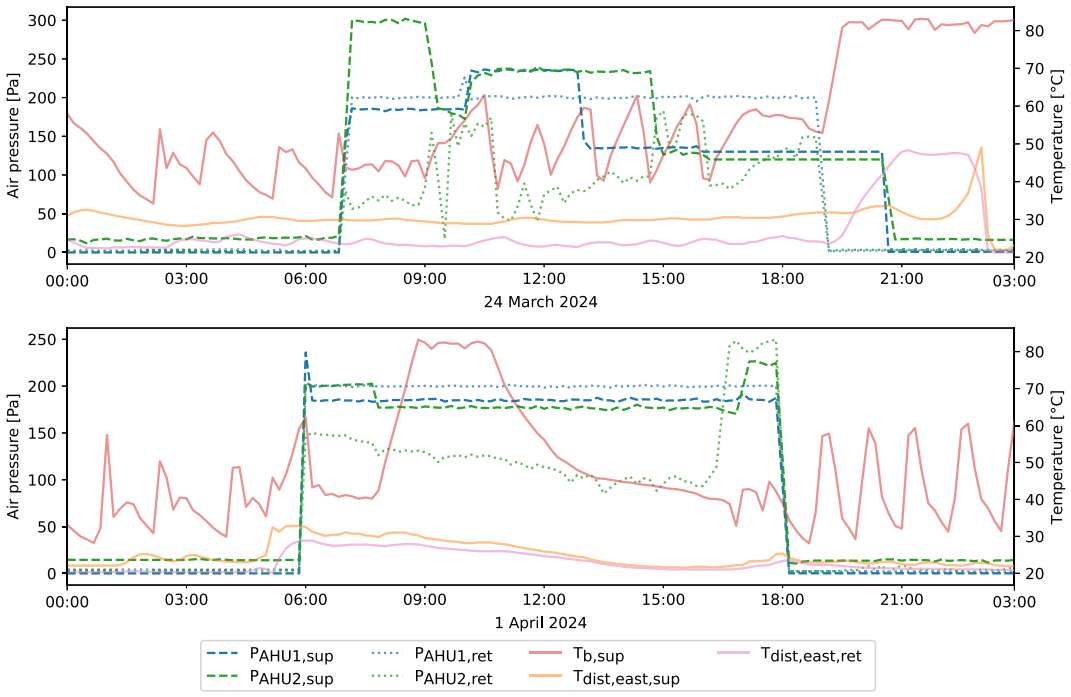

Fig. 11. Sensor data on the dates when boiler, boiler control and hydronic control faults were detected.AHU supply and return pressures (which were being changed in the experiments) are shown along boiler and hydronic group temperatures (which were not changed as part of the experiments).The air pressure setpoint faults introduced on 24 march 2024 (top) and the simultaneous HCV and HRW faults introduced in AHU1 on 1 April 2024 (bottom) lead to a sharp increase in boiler supply temperature in the evening and in the morning, respectively.

Table 6
Performance metrics for detection of AHU faults for the selected DBN input parameters by subcomponent fault experiment.

Subcomponent fault	TP	FP	FN	TN	*CR* [%]	*MR* [%]	*FAR* [%]	*Pr* [%]	F_1 [%]	*BA* [%]
Supply fan	1	1	3	3	50.0	75.0	25.0	50.0	33.3	50.0
Return fan	4	1	0	3	87.5	0.0	25.0	80.0	88.9	87.5
HRW	10	0	0	10	100.0	0.0	0.0	100.0	100.0	100.0
Pressure setpoint	4	4	6	6	50.0	60.0	40.0	50.0	44.4	50.0
HCV	3	0	0	3	100.0	0.0	0.0	100.0	100.0	100.0
Temperature setpoint	5	1	1	5	83.3	16.7	16.7	83.3	83.3	83.3
HCV + HRW	4	2	1	3	70.0	20.0	40.0	66.7	72.7	70.0

their corresponding zones, distribution systems and controls. Rather than utilizing purely qualitative evaluation of the results (e.g., [11]) or quantitative evaluation based on simulated data (e.g., [5,6]), the DBN was evaluated using both historical data and experimental data obtained by introducing faults into the building's AHUs.

The fault diagnosis results indicate that the network is able to detect the most likely faulty components at building scale. These include component degradation faults and biases in the building sensors (particularly CO_2 sensors), which are likely to be a cause of some of the control faults observed in this study. In spite of the high degree of uncertainty created by the faulty sensing infrastructure, the DBN is able to achieve a classification rate of more than 85% for AHU faults evaluated on-site. In particular, faults in energy-related subcomponents (heat recovery wheel, heating coil valve) were more consistently detected than air temperature and pressure setpoint- and sensor-related faults. Therefore, while the

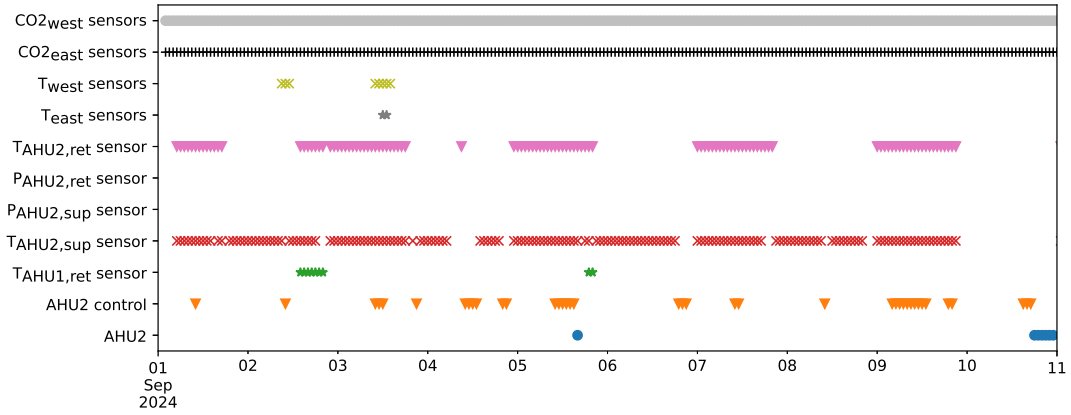

Fig. 12. Hourly fault detection results during the period when an unknown AHU fault.

present version of the DBN is unable to detect energy performance symptoms due to insufficient sensing infrastructure in the building, the DBN might be a useful tool to improve building energy performance if the required sensing infrastructure is retrofitted into the building systems in the future.

The DBN was able to correctly diagnose AHU faults when different simultaneous subcomponent control faults were introduced. However, the DBN performed differently for different simultaneous fault combinations. This highlights the need for further research into FDD during simultaneous faults, which is especially challenging due to inadequate sensors, coupled faults and unclear interactions [53]. This was indeed the case here, where sensors were consistently inaccurate and the causes of the interactions between AHU pressure and boiler temperature setpoints are not clear from the building documentation.

Similarly, the large number of control faults detected in both the historical data and the operational testing highlights the issue with insufficient documentation of present control strategies. While building documents provide a useful guideline for how building system operation was intended at the time of installation, changes over time are not typically documented, leading to the need to determine current control strategies based on the operational data. This underlines how the construction of DBNs is strongly influenced by the system being investigated and by the challenge of transferability.

A key aspect of the development of this DBN was the incorporation of zone-level effects and comfort symptoms into the diagnosis. This allowed the DBN to focus on indoor environmental quality-related symptoms to ensure faults affecting occupant comfort are addressed in time. Furthermore, the real-world example application to a suspected fault shows the potential for occupant feedback to be incorporated into FDD. Given the large number of faulty sensors encountered in the building, occupant complaints provided a key source of information for facilities management to carry out maintenance on the ventilation systems in the real world. Incorporating humans as sensors in whole building FDD can therefore help to ensure desired IAQ conditions are maintained [54]. Such feedback can provide valuable complementary information for DBNs, but while its implementation has been proposed before [55], no such implementation can be found in the literature on whole-HVAC system FDD.

6.1. Limitations and ongoing work

The DBN presented here covers the whole HVAC system, including all major components and their operation. Historical evaluation of the DBN results allowed us to assess general DBN performance and highlight the most frequent faults in the building systems over time. However, in the absence of labeled data, no performance metrics could be calculated during the period 2022–2023. We partially addressed this issue by introducing fault experiments in the building's systems in order to assess DBN performance, though faults could only be introduced in the AHUs due to restrictions from the building operator and safety concerns. Further experimental work should therefore be carried out on the various components of the HVAC systems in order to provide holistic validation of the DBN performance.

Another limitation of this approach is its reliance on building documentation, P&ID diagrams and control documents. The construction of whole-HVAC DBNs is time-consuming, requiring domain knowledge and detailed analysis of the existing systems. Although the overall framework used in this paper to define faults, symptoms, and their relationships is applicable to a wide range of HVAC systems, its implementation in a new HVAC system still requires a moderate level of manual effort. First of all, the specific list of symptoms and faults considered depends on the available sensor infrastructure and the specific HVAC configuration. While the use of data-driven and literature-based parameters can help minimize the case study-specific inputs needed, the expected operating ranges and overall control logic depend on the target HVAC system, which means the specific fault-symptom relationships, probability parameters, and thresholds must be adjusted to reflect the characteristics of the target HVAC system. Therefore, the required effort depends on the similarity between the source and target HVAC systems, with adaptation being considerably simpler for systems sharing comparable configurations. In order to overcome this limitation, DBN libraries containing reusable fault, symptom and causal relationship modules have been developed to support practitioners in adapting DBNs to new case studies, thereby improving the practical transferability of these approaches [42]. Additionally, further work should be carried out on the development of automated methods for DBN construction. Previous work on the use of large language models [56], computer vision [57] and deep learning approaches [58] has shown promising results in recognizing P&IDs automatically.

Analyzing the historical data demonstrated a large number of sensor faults in the case study building. Such consistent issues in the building sensing infrastructure raise significant uncertainty about any FDD method's ability to detect faults in the building. For example, the use of data-driven thresholds based on sensor measurements that show unusual behavior can lead to biased symptom detection thresholds. Given the absence of fault-free historical data to define baseline thresholds, we compared the performance of the DBN using different data-driven thresholds based on different standard deviations in the historical data as well as literature-based reference values. The fault isolation results nevertheless carry additional uncertainty due to the faulty sensors used for symptom detection.

Despite this uncertainty, the DBN presented here was able to achieve a CR of 87.8% for the experimental faults introduced into the AHU controls. This finding confirms DBNs' ability to deal with contradictory information and uncertainty. Nevertheless, this uncertainty might have contributed to the DBN's difficulty in isolating fan faults and pressure setpoint faults. As a result, the MR remained high, with a minimum of 27.5% achieved during the sensitivity analysis. For the specific case study presented here, significant maintenance work should be carried out on the sensors in the building. Such an overhaul could be coupled with an upgrade of the building's sensing infrastructure to incorporate the types of flow rate sensors needed to detect energy performance symptoms in the building.

In the absence of reliable sensors, the case study presented here stressed the importance of occupant feedback as a valuable complementary source of information in FDD. In subsequent work, the incorporation of building occupant feedback as a symptom and behavior as a potential fault in this whole-building DBN is being explored. This paper therefore represents an initial step towards FDD methods that can support both building system performance and occupant comfort.

7. Conclusions

This paper presented the development of a DBN for a whole building HVAC system and its testing and partial validation in a real-world case study building in the Netherlands. The DBN focuses on cross-level fault diagnosis at the aggregated system level, accounting for all generator, distribution and end-use modules in the building.

The results of fault detection based on the long-term historical data indicate that the network is able to detect the most likely faulty components at whole building HVAC scale. In particular, significant issues with the building's sensing infrastructure, and CO_2 sensors in particular, were detected. The evaluation of the DBN using data from AHU fault experiments conducted on-site demonstrated that the DBN was able to achieve a classification rate of more than 85% in spite of the high degree of uncertainty created by the sensing infrastructure. While false alarms were not a significant issue (FAR less than 5%), the miss rate was above 25% for all cases. As in other works in the literature, the choice of prior probabilities proved to have a relatively minor impact on the results, while the "uniform" conditional probability case and data-driven symptom detection thresholds were found to perform best. The implementation of the

final version of the DBN during an undiagnosed fault identified the most likely root cause for occupant complaints, which should be investigated further by building operators.

This study represents one of the few DBN implementations in the literature that include all major HVAC systems in the building. Furthermore, rather than utilizing purely qualitative evaluations of the results or quantitative evaluations based on simulated data, the DBN was tested using both historical and experimental data. The model actively diagnoses all major HVAC systems simultaneously, even though experimental validation was limited to AHU controls due to safety concerns. In spite of the significant faults found in the building's sensor infrastructure, the DBN was able to achieve an adequate fault detection performance and diagnose multiple simultaneous faults in the AHUs. Thus, this work advances the development of validated whole-building HVAC system FDD using DBNs.

CRediT authorship contribution statement

Martín Mosteiro-Romero: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ziao Wang:** Writing – review & editing, Methodology, Investigation. **Arie Taal:** Methodology, Conceptualization. **Chujie Lu:** Writing – review & editing, Investigation. **Laure Itard:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization.

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.

Acknowledgements

This research was carried out as part of the Brains for Buildings project, sponsored by the Dutch grant program for Mission-Driven Research, Development and Innovation (MOOI), which was established by the Dutch Ministry of Economic Affairs and Climate Policy and the Ministry of the Interior and Kingdom Relations and executed by RVO Netherlands Enterprise Agency.

Appendix A. Complete list of faults

Table A.7 summarizes the component, control and sensor faults considered in this study. An example of a specific subsystem fault is included for each aggregated system fault.

Table A.7
List of faults defined for the DBN. Subscript i represents the zones (East, West and Atrium), subscript j represents the zones where there are CO_2 sensors (East and West) and n represents the AHUs (AHU1 and AHU2).

Fault	Fault description	Example
Component faults		
Boiler	Fault in a subcomponent in the boiler assembly	Boiler 2 valve fault
Chiller	Fault in a subcomponent in the chiller assembly	Chiller pump fault
Hydronic group i	Fault in a subcomponent in the hydronic group i	Distribution pipe leak
AHU1	Fault in a subcomponent in AHU1	Heating coil valve fault
AHU2	Fault in a subcomponent in AHU2	Heat recovery wheel fault
Ducts	Fault in a subcomponent in the air distribution	Supply duct leak
Control faults		
Boiler controls	Fault in a control in the boiler assembly	Incorrect setpoint temperature
Chiller controls	Fault in a control in the chiller assembly	Chiller setpoint temperature
Hydronic group i controls	Fault in a control in the hydronic group i	Incorrect supply temperature
AHU1 controls	Fault in a control in AHU1	Incorrect setpoint temperature
AHU2 controls	Fault in a control in AHU2	Incorrect supply pressure
Zone j controls	Fault in a room-level control in zone j	Window open when the room is empty
Sensor faults		
$T_{b,\mathrm{sup}}$ sensor	Boiler supply temperature sensor fault	Sensor fault
$T_{b,ret}$ sensor	Boiler return temperature sensor fault	Sensor fault
$P_{b,ret}$ sensor	Boiler return pressure sensor fault	Sensor fault
$T_{c,\mathrm{sup}}$ sensor	Chiller supply temperature sensor fault	Sensor fault
$T_{c,ret}$ sensor	Chiller return temperature sensor fault	Sensor fault
$P_{c,ret}$ sensor	Chiller return pressure sensor fault	Sensor fault
T_i sensors	Fault in an air temperature sensor in Zone i	Floor 3 sensor fault
$CO2_j$ sensor	Fault in a CO_2 sensor in Zone j	Room *6-O-450* sensor fault
$T_{dist,i,\mathrm{sup}}$ sensor	Hydronic group i supply temperature sensor fault	Sensor fault
$T_{dist,i,ret}$ sensor	Hydronic group i return temperature sensor fault	Sensor fault
$P_{n,\mathrm{sup}}$ sensor	AHU n supply pressure sensor	Sensor fault
$P_{n,ret}$ sensor	AHU n return pressure sensor	Sensor fault
$T_{n,\mathrm{sup}}$ sensor	AHU n supply temperature sensor	Sensor fault
$T_{n,ret}$ sensor	AHU n return temperature sensor	Sensor fault
T_{out} sensor	Outdoor air temperature sensor fault	Sensor fault

Appendix B. Complete list of symptoms

Table B.8 summarizes the symptoms considered in this study. The threshold ϵ_i indicated for each symptom S_i corresponds to the thresholds summarized in Table 2.

Table B.8
List of symptoms defined for the DBN. Subscripts b, c, $dist$, i, n and j represent the boiler, chiller, distribution system i (East, west, atrium), AHU n (AHU1 and AHU2) and boiler number j (1 or 2), respectively. sup, ret, set and $calc$ represent the supply, return, setpoint and calculated setpoint. w refers to seasons when heating ($T_{out} \leq 20$ °C) or cooling ($T_{out} \geq 14$ °C) is available according to the control documents. s refers to the corresponding sensor. $comfort$ is the reference value to maintain indoor comfort. hrw refers to the heat recovery wheel (HRW) in an AHU, such that $T_{n,\text{sup},hrw}$ and $T_{n,ret,hrw}$ refer to the supply temperature right after the HRW and the return temperature right before the HRW, respectively. Subscript $LoRa$ refers to the sensor reading from one of the LoRa sensors in the building, whereas r refers to the room in which the individual sensor was located. For all symptoms that are only detected when the systems are off, a minimum of 3 hours off before symptom detection was implemented.

Symptom	Symptom description	Symptom presence definition
Operational state symptoms		
$T_{b,\text{sup}}$	Boiler supply temperature deviation	$\lvert T_{b,\text{sup}} - T_{b,set}\rvert > \epsilon_{b,\text{sup}}$
$T_{b,set}$	Boiler setpoint temperature deviation	$\lvert T_{b,set} - T_{b,calc}\rvert > \epsilon_{b,set}$
$T_{b,ret}$	Boiler return temperature deviation	$\lvert T_{b,\text{sup}} - T_{b,ret}\rvert > \epsilon_{b,ret}$
$P_{b,ret}$	Boiler return pressure too low	$P_{b,ret} \leq \epsilon_{P_b}$
$T_{c,\text{sup}}$	Chiller supply temperature deviation	$\lvert T_{c,\text{sup}} - T_{c,\text{sup},set}\rvert > \epsilon_{c,\text{sup}}$
$T_{c,ret}$	Chiller return temperature deviation	$\lvert T_{c,ret} - T_{c,ret,set}\rvert > \epsilon_{c,ret}$
$P_{c,ret}$	Chiller return pressure too low	$P_{c,ret} \leq \epsilon_{P_c}$
$T_{dist,i,\text{sup}}$	Hydronic group i supply temperature deviation	$\lvert T_{dist,i,\text{sup}} - T_{dist,i,set}\rvert > \epsilon_{dist,i,\text{sup}}$
$T_{dist,i,ret}$	Hydronic group i return temperature deviation	$\lvert T_{dist,i,\text{sup}} - T_{dist,i,ret}\rvert > \epsilon_{dist,i,ret}$
$T_{n,\text{sup},w}$	AHU n supply temperature deviation during season w (heating or cooling)	$\lvert T_{n,\text{sup}} - T_{n,set}\rvert > \epsilon_{n,\text{sup}}$ and season is w
$T_{n,set}$	AHU n setpoint temperature deviation	$\lvert T_{n,set} - T_{n,calc}\rvert > \epsilon_{n,set}$
$T_{n,ret}$	AHU n return temperature deviation	$T_{n,ret} \notin [\epsilon_{T,\min}, \epsilon_{T,\max}]$
$P_{n,\text{sup}}$	AHU n supply pressure too low	$P_{n,\text{sup}} \leq \epsilon_{Pn,\text{sup}}$
$P_{n,ret}$	AHU n return pressure too low	$P_{n,ret} \leq \epsilon_{Pn,ret}$
Comfort range symptoms		
RH_{AHU2}	AHU2 return relative humidity outside of comfort range	$RH_{AHU2,ret} \notin [\epsilon_{rh,\min}, \epsilon_{rh,\max}]$ and $RH_{n,\text{sup}} \in [\epsilon_{rh,\min}, \epsilon_{rh,\max}]$
$\overline{T}_{i,w}$	Zone i temperature outside of comfort range during season w (heating or cooling)	$\overline{T_i} \notin [\epsilon_{T,\min}, \epsilon_{T,\max}]$ and season is w
$CO2_{i,\max}$	Zone i maximum CO_2 too high	$CO2_{i,\max} > \epsilon_{CO2,comfort}$
Sensor deviation symptoms		
$CO2_{i,range}$	Zone i CO_2 sensor out of range	$\overline{CO2_i} \notin [\epsilon_{CO2,\min}, \epsilon_{CO2,\max}]$
$CO2_{i,dev}$	Zone i CO_2 sensor deviation	$\lvert CO2_{i,r} - CO2_{r,LoRa}\rvert > \epsilon_{CO2,dev}$
$DT_{b,\text{sup}}$	Boiler supply temperature sensor deviation	$\min(\lvert T_{b,\text{sup}} - T_{b,j}\rvert) > \epsilon_{b,\text{sup},s}$
$DT_{b,ret}$	Boiler return temperature sensor deviation	$\lvert T_{b,\text{sup}} - T_{b,ret}\rvert > \epsilon_{b,ret,s}$ (*)
DT_c	Chiller temperature sensor deviation	$\lvert T_{c,\text{sup}} - T_{c,ret}\rvert > \epsilon_{c,s}$ (*)
DP_n	AHU n pressure sensor deviation	$\lvert P_{n,\text{sup}} - P_{n,ret}\rvert > \epsilon_{Pn,s}$ (*)
$DT_{AHU1,\text{sup}}$	AHU1 supply temperature sensor symptom	$\lvert T_{AHU1,\text{sup}} - T_{AHU1,\text{sup},hrw}\rvert > \epsilon_{AHU1,\text{sup},s}$ (*)
$DT_{AHU1,ret}$	AHU1 return temperature sensor symptom	$\lvert T_{AHU1,ret} - T_{AHU1,\text{sup},hrw}\rvert > \epsilon_{AHU1,ret,s}$ (*)
$DT_{AHU2,\text{sup}}$	AHU2 supply temperature sensor deviation	$\lvert T_{AHU2,\text{sup}} - T_{AHU2,zone,ret}\rvert > \epsilon_{AHU2,\text{sup},s}$ (*)
$DT_{AHU2,ret}$	AHU2 return temperature sensor deviation	$\lvert T_{AHU2,\text{sup}} - T_{AHU2,zone,ret}\rvert > \epsilon_{AHU2,ret,s}$ (*)
$DT_{dist,i}$	Hydronic group i temperature sensor deviation	$\lvert T_{dist,i,\text{sup}} - T_{dist,i,ret}\rvert > \epsilon_{dist,i,s}$ (*)

(*) Checked only after system has been off for 3 hours.

Appendix C. Individual representations of the DBN structure for each system in the whole-HVAC DBN

The DBN structure for each individual system including all symptoms and faults is shown in Fig. C.13. Component and control faults in each aggregated system are color-coded for readability. Individual sensor faults are shown in gray. Symptoms that are related to multiple system faults (and their corresponding sensor faults) are repeated in each individual DBN, but are interconnected in the whole-HVAC DBN. A machine-readable version of the whole-HVAC DBN is also included in the supplementary material.

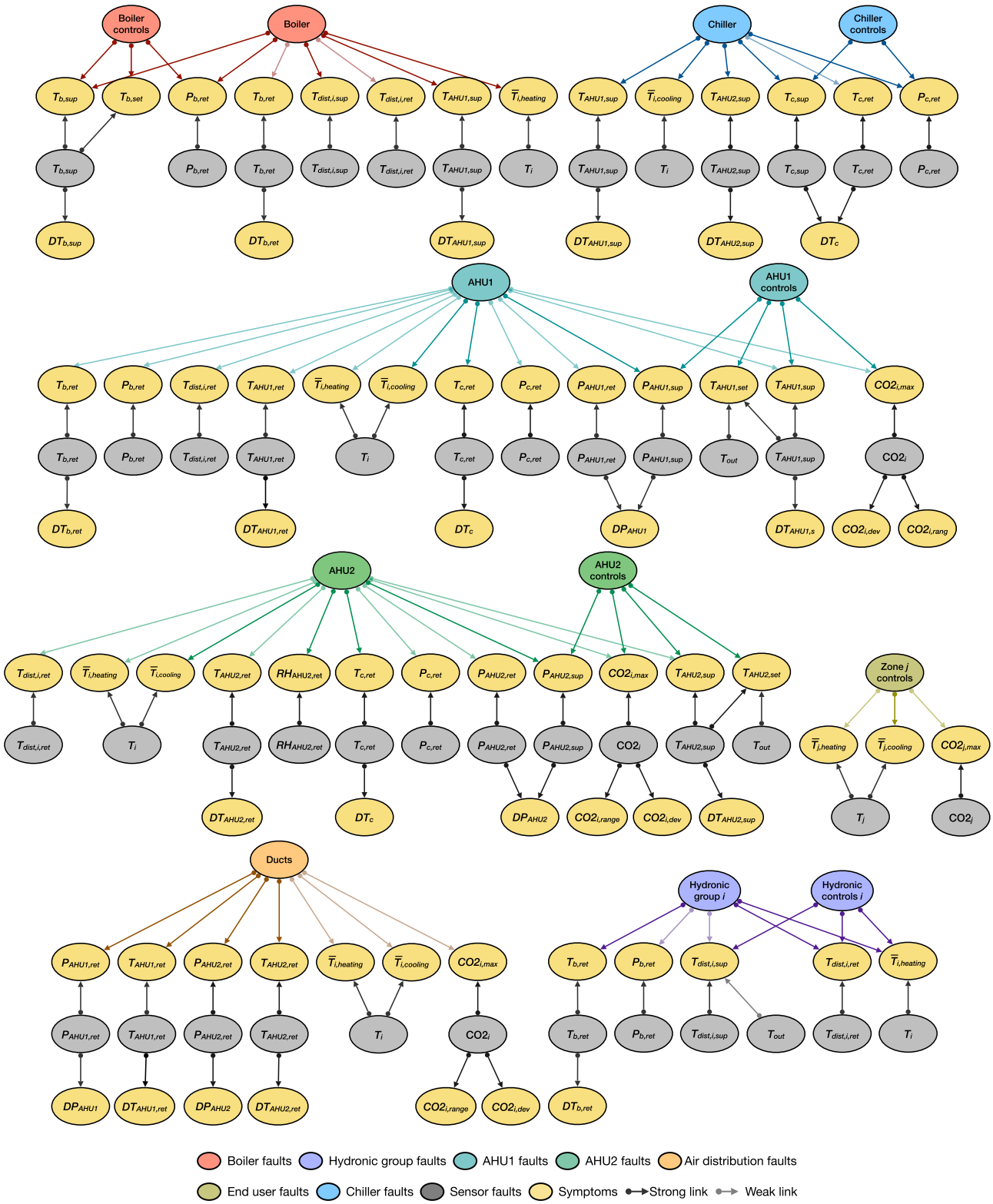

Fig. C.13. Individual representation of the DBN structure for each system in the whole HVAC DBN, explicitly including all sensor symptoms and faults. Nodes including the subscript i represent the three building zones: East, West, and Atrium. Subscript j indicates zones east and west.

Appendix D. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.buildenv.2026.115265.

Data availability

Data will be made available on request.

References

[1] 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

[2] S. Katipamula, M.R. Brambley, Methods for fault detection, diagnostics, and prognostics for building systems–a review, Part I, HVACR Res. 11 (1) (2005) 3–25, https://doi.org/10.1080/10789669.2005.10391123

[3] 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

[4] M.S. Mirnaghi, F. Haghighat, Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: a comprehensive review, Energy Build. 229 (2020) 110492, https://doi.org/10.1016/j.enbuild.2020.110492

[5] J. Huang, N. Ghalamsiah, A. Patharkar, O. Pradhan, M. Chu, T. Wu, J. Wen, Z. O'neill, K.S. Candan, An entropy-based causality framework for cross-level faults diagnosis and isolation in building HVAC systems, Energy Build. 317 (2024) 114378, https://doi.org/10.1016/j.enbuild.2024.114378

[6] K. Verbert, R. Babuška, B. De Schutter, Combining knowledge and historical data for system-level fault diagnosis of HVAC systems, Eng. Appl. Artif. Intell. 59 (2017) 260–273, https://doi.org/10.1016/j.engappai.2016.12.021

[7] 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

[8] W. Kim, S. Katipamula, A review of fault detection and diagnostics methods for building systems, Sci. Technol. Built Environ. (2017) 1–18, https://doi.org/10.1080/23744731.2017.1318008

[9] 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

[10] Y. Chen, J. Wen, T. Chen, O. Pradhan, Bayesian networks for whole building level fault diagnosis and isolation, in: Proceedings of the International High Performance Buildings Conference, West Lafayette, IN, USA, 2018, pp. 266.

[11] 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

[12] O. Pradhan, J. Wen, Y. Chen, X. Lu, M. Chu, Y. Fu, Z. O'neill, T. Wu, K.S. Candan, Dynamic Bayesian network-based fault diagnosis for ASHRAE guideline 36: high performance sequence of operation for HVAC systems, in: Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys), Coimbra, Portugal, 2021, pp. 365–368.

[13] 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

[14] T. Gao, S. Marié, P. Béguery, S. Thebault, S. Lecoeuche, Integrated building fault detection and diagnosis using data modeling and Bayesian networks, Energy Build. 306 (2024) 113889, https://doi.org/10.1016/j.enbuild.2024.113889

[15] R. Bortolini, N. Forcada, A probabilistic-based approach to support the comfort performance assessment of existing buildings, J. Clean. Prod. 237 (2019) 117720, https://doi.org/10.1016/j.jclepro.2019.117720

[16] H.H. Hosamo, H.K. Nielsen, D. Kraniotis, P.R. Svennevig, K. Svidt, Digital twin framework for automated fault source detection and prediction for comfort performance evaluation of existing non-residential Norwegian buildings, Energy Build. 281 (2023) 112732, https://doi.org/10.1016/j.enbuild.2022.112732

[17] Z. Chen, Z. O'neill, J. Wen, O. Pradhan, T. Yang, X. Lu, G. Lin, S. Miyata, S. Lee, C. Shen, R. Chiosa, M.S. Piscitelli, A. Capozzoli, F. Hengel, A. Kührer, M. Pritoni, W. Liu, J. Clauß, Y. Chen, T. Herr, A review of data-driven fault detection and diagnostics for building HVAC systems, Appl. Energy 339 (2023) 121030, https://doi.org/10.1016/j.apenergy.2023.121030

[18] C. Miller, Z. Nagy, A. Schlueter, Automated daily pattern filtering of measured building performance data, Autom. Constr. 49 (A) (2015) 1–17, https://doi.org/10.1016/j.autcon.2014.09.004

[19] A. Capozzoli, F. Lauro, I. Khan, Fault detection analysis using data mining techniques for a cluster of smart office buildings, Expert Syst. Appl. 42 (2015) 4324–4338, https://doi.org/10.1016/j.eswa.2015.01.010

[20] S. Wu, J.-Q. Sun, Cross-level fault detection and diagnosis of building HVAC systems, Build. Environ. 46 (2011) 1558–1566, https://doi.org/10.1016/j.buildenv.2011.01.017

[21] S. Wu, J.-Q. Sun, A top-down strategy with temporal and spatial partition for fault detection and diagnosis of building HVAC systems, Energy Build. 43 (2011) 2134–2139, https://doi.org/10.1016/j.enbuild.2011.04.020

[22] F. Magoulés, H. Xiang Zhao, D. Elizondo, Development of an RDP neural network for building energy consumption fault detection and diagnosis, Energy Build. 62 (2013) 133–138, https://doi.org/10.1016/j.enbuild.2013.02.050

[23] Y. Chen, J. Wen, A whole building fault detection using weather based pattern matching and feature based PCA method, in: Proceedings of the 2017 IEEE International Conference on Big Data (BIGDATA), Boston, MA, USA, 2017, pp. 4050–4057, https://doi.org/10.1109/BigData.2017.8258421

[24] Y. Chen, J. Wen, J. Lo, Using weather and schedule-based pattern matching and feature-based principal component analysis for whole building fault Detection–Part I development of the method, ASME J. Eng. Sustain. Build. Cities 3 (1) (2022) 011001, https://doi.org/10.1115/1.4052729

[25] B. Sun, P.B. Luh, Q.-S. Jia, Z. O'neill, F. Song, Building energy doctors: an SPC and Kalman filter-based method for system-level fault detection in HVAC systems, IEEE Trans. Autom. Sci. Eng. 11 (1) (2014) 215–229, https://doi.org/10.1109/TASE.2012.2226155

[26] Z. Wang, Z. Wang, S. He, X. Gu, Z.F. Yan, Fault detection and diagnosis of chillers using Bayesian network merged distance rejection and multi-source non-sensor information, Appl. Energy 188 (2017) 200–214, https://doi.org/10.1016/j.apenergy.2016.11.130

[27] 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

[28] C. Fan, Y. Lin, M.S. Piscitelli, R. Chiosa, H. Wang, A. Capozzoli, Y. Ma, Leveraging graph convolutional networks for semi-supervised fault diagnosis of HVAC systems in data-scarce contexts, Build. Simul. 16 (2023) 1499–1517, https://doi.org/10.1007/s12273-023-1041-1

[29] 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

[30] J. Zhang, Z. Li, H. Chen, H. Cheng, L. Xing, Y. Wang, L. Zhang, Integrated generative networks embedded with ensemble classifiers for fault detection and diagnosis under small and imbalanced data of building air condition system, Energy Build. 268 (2022) 112207, https://doi.org/10.1016/j.enbuild.2022.112207

[31] B. Li, F. Cheng, H. Cai, X. Zhang, W. Cai, A semi-supervised approach to fault detection and diagnosis for building HVAC systems based on the modified generative adversarial network, Energy Build. 246 (2021) 111044, https://doi.org/10.1016/j.enbuild.2021.111044

[32] 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

[33] 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

[34] A. Taal, L. Itard, Fault detection and diagnosis for indoor air quality in DCV systems: application of 4S3F method and effects of DBN probabilities, Build. Environ. 174 (2020) 106632, https://doi.org/10.1016/j.buildenv.2019.106632

[35] 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

[36] C. Fan, F. Xiao, C. Yan, C. Liu, Z. Li, J. Wang, A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning, Appl. Energy 235 (2019) 1551–1560, https://doi.org/10.1016/j.apenergy.2018.11.081

[37] Y. Zhao, F. Xiao, S. Wang, An intelligent chiller fault detection and diagnosis methodology using Bayesian belief network, Energy Build. 57 (2013) 278–288, https://doi.org/10.1016/j.enbuild.2012.11.007

[38] G. Lin, H. Kramer, J. Granderson, Building fault detection and diagnostics: achieved savings, and methods to evaluate algorithm performance, Build. Environ. 168 (2020) 106505, https://doi.org/10.1016/j.buildenv.2019.106505

[39] B. Zhang, J. Wang, L. Shangguan, Y. Rezgui, A. Ghoroghi, X. Cao, C. Song, M. Ding, T. Zhao, Analysis of HVAC sensor characteristics for operation and maintenance of the indoor environment: a case research on public building HVAC, Build. Simul. 18 (2025) 1205–1228, https://doi.org/10.1007/s12273-025-1254-6

[40] A. Rogers, F. Guo, B. Rasmussen, Uncertainty analysis and field implementation of a fault detection method for residential HVAC systems, Sci. Technol. Built Environ. 26 (2020) 320–333, https://doi.org/10.1080/23744731.2019.1676093

[41] ASHRAE, Guideline 36-2018: High Performance Sequences of Operation for HVAC Systems, American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), 2018.

[42] A. Taal, Report D 1.03 a/B: Open Library of Standardized DBN Models for the Most Common Components & Sensors, Models and Controls in HVAC Installations. Technical report, Brains4Buildings, 2025. https://brains4buildings.org/results.

[43] B. Cai, L. Huang, M. Xie, Bayesian networks in fault diagnosis, IEEE Trans. Ind. Inform. 13 (5) (2017) 2227–2240, https://doi.org/10.1109/TII.2017.2695583

[44] A. Zagorecki, M.J. Druzdzel, Knowledge engineering for Bayesian networks: how common are Noisy-MAX distributions in practice? IEEE Trans. Syst. Man Cybern. Syst. 43 (1) (2013) 186–195, https://doi.org/10.1109/TSMCA.2012.2189880

[45] 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

[46] A. Ankan, J. Textor, pgmpy: a Python toolkit for Bayesian networks, J. Mach. Learn. Res. 25 (265) (2024) 1–8, http://jmlr.org/papers/v25/23-0487.html.

[47] S.P. Melgaard, K.H. Andersen, A. Marszal-Pomianowska, R.L. Jensen, P.K. Heiselberg, Fault detection and diagnosis encyclopedia for building systems: a systematic review, Energies 15 (2022) 4366, https://doi.org/10.3390/en15124366

[48] Y. Wang, Z. Wang, S. He, Z. Wang, A practical chiller fault diagnosis method based on discrete Bayesian network, Int. J. Refrig. 102 (2019) 159–167, https://doi.org/10.1016/j.ijrefrig.2019.03.008

[49] Z. Wang, L. Wang, Y. Tan, J. Yuan, X. Li, Fault diagnosis using fused reference model and Bayesian network for building energy systems, J. Build. Eng. 34 (2021) 101957, https://doi.org/10.1016/j.jobe.2020.101957
[50] Z. Wang, B. Liang, J. Guo, L. Wang, Y. Tan, X. Li, Fault diagnosis based on residual–knowledge–data jointly driven method for chillers, Eng. Appl. Artif. Intell. 125 (2023) 106768, https://doi.org/10.1016/j.engappai.2023.106768
[51] 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. Therm. Eng. 129 (2018) 674–683, https://doi.org/10.1016/j.applthermaleng.2017.10.079
[52] S. He, Z. Wang, Z. Wang, X. Gu, Z. Yan, Fault detection and diagnosis of chiller using Bayesian network classifier with probabilistic boundary, Appl. Therm. Eng. 107 (2016) 37–47, https://doi.org/10.1016/j.applthermaleng.2016.06.153
[53] B. Wu, W. Cai, H. Chen, X. Zhang, A hybrid data-driven simultaneous fault diagnosis model for air handling units, Energy Build. 245 (2021) 111069, https://doi.org/10.1016/j.enbuild.2021.111069
[54] K. Mohammed, W. Luo, S. Walker, R. Kramer, Developing key performance indicators for occupant-centric buildings using occupant feedback: a scoping review and methodological framework, Energy Build. 347 (2025) 116228, https://doi.org/10.1016/j.enbuild.2025.116228
[55] A. Heidtmann, Exploring the Integration of Occupant Behaviour in Fault Detection and Diagnosis in Smart Buildings (Master's thesis), Delft University of Technology, Delft, Netherlands, 2022.
[56] C. Lu, S. Walker, C. Struck, L. Itard, D. Saelens, LLM-assisted digitalization of HVAC diagrams, in: BuildSys '25: Proceedings of t12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Golden, USA, 2025, https://doi.org/10.1145/3736425.3772104
[57] M. Gupta, C. Wei, T. Czerniawski, Semi-supervised symbol detection for piping and instrumentation drawings, Autom. Constr. 159 (2024) 105260, https://doi.org/10.1016/j.autcon.2023.105260
[58] M.F. Theisen, K.N. Flores, L.S. Balhorn, A.M. Schweidtmann, Digitization of chemical process flow diagrams using deep convolutional neural networks, Digit. Chem. Eng. 6 (2023) 100072, https://doi.org/10.1016/j.dche.2022.100072