Project acronym	B4B
Project full name	Brains for Building's Energy Systems
Grant No	M00I32004
Project duration	May 1, 2021 – December 31, 2025

Deliverable 1.07

Determining sensors required, choosing Low-Cost Monitors and evaluating Machine Learning algorithms to detect Air Handling Unit faults

Lead authors: Srinivasan Gopalan (TU/e)

Co-authors: Rick Kramer (TU/e)

Work package	1
Result	2
Lead beneficiary	TU/e
Due Date	Month 48 (1 May 2025)
Deliverable Status	Final
File name	B4B WP1 D1.07_Evaluation Machine Learning algorithms, applying new sensors & possibilities dynamic air conditioning.docx
Reviewers	Kees Wisse (DWA) and Pieter Pauwels (TU/e)

This project received funding from Dutch Ministry of Economic Affairs and Climate Policy and Ministry of the Interior and Kingdom Relations under the MOOI program

EXECUTIVE SUMMARY

Buildings consume about 36% of the world's energy consumption of which up to 30% could be wasted due to poor maintenance & degradation of equipment, inadequate sensing, improper control and lack of utilization of the capabilities of Building Automation Systems (BASs). Faults occurring in HVAC systems, which form a subset of these causes, also lead to thermal discomfort of occupants in addition to energy waste. There is a need for Automated Fault Detection and Diagnosis (AFDD) tools that help in identifying these issues at a very early stage. The lack of sufficient sensors in buildings and their poor accuracy impede field testing and real-building deployment of AFDD tools. The availability of several Machine Learning algorithms requires an evaluation of their performance in detecting faults. The main objective of this report is three-fold:

(i) To determine locations within the HVAC system where sensors need to be installed to detect HVAC system faults,

(ii) To evaluate the performance of Low-Cost Sensors (LCSs) and Low-Cost Monitors (LCMs) in their ability to measure parameters such as Temperature (T), Carbon dioxide (CO_2) concentration and Relative Humidity (RH) with high precision, high accuracy and low bias, and

(iii) To evaluate the performance of Machine Learning (ML) algorithms to accurately forecast fault-free operating values of system/component-level sensors and consequently detect the presence of symptoms caused by faults.

Sensoring locations for symptom detection

The research presented in Chapter 2 of this report used a simulation-based approach to determine relationships between Air Handling Unit (AHU) faults- heating & cooling coil valves (stuck, leak), fan (failure & stuck), heat recovery wheel (failure), sensor biases (supply air & indoor air) and incorrect setpoint (supply air)- and their symptoms (recorded at sensors) to help HVAC engineers select a standardised set of sensors which will facilitate the uptake of generic AFDD tools. An EnergyPlus model of the Kropman office building in Breda was used to study the sensor values recorded in the absence and presence of the above-mentioned faults.

The results indicated that:

i. During the heating season, of the 14 faults simulated, the IAT North sensor can detect 12 faults, the heating coil mass flow rate sensor and the supply air exit temperature sensor at the heat recovery wheel can detect 11 faults each, the AHU electrical power meter and the return air entry temperature sensor at the heat recovery wheel can detect 10 faults each.

ii. During the cooling season, of the 13 faults simulated, the chiller electrical power meter can detect 10 faults; the return water temperature sensor and supply air temperature sensors can detect 8-9 faults each depending on the zoning of the studied building.

iii. The fan failure fault can be detected at the highest number of sensors during both the seasons.

It is also quite common to use knowledge of how HVAC system works and field knowledge to determine the sensors needed to detect faults. Readers are also referred to Deliverable 1.03 of the B4B project for the developed DBN libraries.

Evaluation of the performance of LCSs/LCMs for use in AFDD

The research presented in Chapter 3 of this report evaluated the performance of 16 LCMs available in the market and ranked them based on their suitability for use in FDD applications. The study was conducted in a controlled climate chamber to test multiple LCMs and evaluate their performance in measuring Temperature (T), Carbon dioxide (CO_2), and Relative Humidity (RH) under varying conditions. Metrics corresponding to Precision, Linearity, Bias, and Error were calculated and a weighted rating based on the metrics' values for each LCM was obtained using the rating scale of the Smart Readiness Indicator (SRI) framework and the importance of the parameter for FDD. It was concluded that Nuwave, Edimax and Cloudgarden LCMs performed the best and can be chosen for use in FDD.

Evaluation of the performance of ML algorithms in symptom detection

The research presented in chapter 4 of this report performed a comparison of the performance of supervised learning-based Machine Learning (ML) algorithms for time series-based fault detection in the Air Handling Unit (AHU). Additionally, the performance of algorithms with a lower feature set (without any symptomatic feature) was compared to the performance with a higher feature set (possibly including symptomatic features). The Kropman office building in Breda was used as a living lab where several low-cost monitors and mass flow rate meters were installed and data was logged at a 1-minute interval by the Building Management System (BMS) to obtain high-granularity data. Non-faulty, baseline data was used to train models based on seven algorithms: Support Vector Regression (SVR), Random Forest (RF), Decision Tree (DT), XGBoost, CatBoost, LightGBM and Multi-Layer Perceptron (MLP). The results showed that, across the six scenarios studied, XGBoost regression gave the best prediction for three scenarios; MLP, RF and SVR gave the best prediction for one scenario each. Across all the symptoms, the ML models predicted better and detected symptoms with a higher accuracy when using a greater number of features. The symptom detection results did not show any notable indication of wrong prediction of fault-free values of target variables when using features that could be symptomatic under the presence of faults. The dropping of “symptomatic features” led to a poorer fit of the developed ML models which in turn led to a reduced accuracy in symptom detection.

Readers are referred to Deliverable 1.08 of the B4B project for information about FDD software module prototyping and integration aspects.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

Definition of select terminologies used in Chapter 2

Fault- An event that causes the HVAC system to deviate from its normal operating pattern or behaviour.

Sensor- a device that is used to measure various operating parameters such as temperature, pressure, flow rates, energy consumption etc.

Symptom- a deviation in the measurement of a sensor observed under the presence of a fault

Outlier- an outlier is a datapoint (value measured by sensors) which lies at an abnormal distance from other datapoints[1].

[1] NIST: 7.1.6. What are outliers in the data?

ACRONYMS

Acronym	Full form
AFDD	Automated Fault Detection and Diagnosis
AHU	Air Handling Unit
APAR	Air Handling Unit Performance Assessment Rules
ASHRAE	American Society of Heating Refrigerating and Air-conditioning Engineers
BMS	Building Management System
CC	Cooling Coil
CCV	Cooling Coil Valve
CI	Confidence Interval
CV-RMSE	Coefficient of Variation of the Root Mean Square Error
DT	Decision Trees
EMS	Energy Management System
FCU	Fan Coil Unit
FDD	Fault Detection and Diagnosis
FN	False Negative
FP	False Positive
HC	Heating Coil
HC SWT	Heating Coil Supply Water Temperature
HCV	Heating Coil Valve
HRW	Heat Recovery Wheel
HVAC	Heating Ventilation and Air Conditioning
IAT	Indoor Air Temperature
IEQ	Indoor Environmental Quality
KNMI	Royal Netherlands Meteorological Institute
KPI	Key Performance Indicator
LCM	Low-Cost Monitor
LCS	Low-Cost Sensor
MAD	Median Absolute Deviation
MAE	Mean Absolute Error
MAT	Mixed Air Temperature
MFR	Mass Flow Rate
ML	Machine Learning
MLP	Multi-Layer Perceptron
MNB	Mean Normalised Bias
MPC	Model Predictive Control
MPE	Mean Percentage Error
MSE	Mean Squared Error
NHST	Null Hypothesis Statistical Testing
NMBE	Normalised Mean Bias Error
NN	Neural Networks
NRMSE	Normalised Root Mean Squared Error
OAFR	Outdoor Air Flow Ratio
PM	Particulate Matter
PPD	Predicted Percentage Dissatisfied
RAT	Return Air Temperature
RF	Random Forest
RGI	Research Grade Instrument

Acronym	Full form
RH	Relative Humidity
RMSE	Root Mean Squared Error
RWT	Return Water Temperature
SAT	Supply Air Temperature
SDCD	Symptom Daily Continuous Duration
SI	Sensitivity Index
SOP	Symptom Occurrence Probability
SRI	Smart Readiness Indicator
SVM	Support Vector Machine
SVR	Support Vector Regression
SWT	Supply Water Temperature
TN	True Negative
TP	True Positive
VAV	Variable Air Volume
WPM	Weather Pattern Matching

1 INTRODUCTION

Buildings are responsible for 36% of the energy consumption across the world as per the United Nations Environment Program (UNEP) [1]. Heating, Ventilation and Air Conditioning (HVAC) systems constitute an important part of a building's energy consumption. For example, as reported in J. Kim et al. [2], HVAC systems constitute 37% of the final energy consumption in small commercial buildings as per the Commercial Building Energy Consumption Survey (CBECS) survey of 2012. This share could, however, change for larger commercial buildings. It has been widely reported in literature that up to 30% of the energy consumed by buildings is wasted, which has been attributed to poor maintenance & degradation of equipment, inadequate sensing, improper control and lack of utilization of the capabilities of Building Automation Systems (BASs) [3], [4], [5], [6]. The faults occurring in HVAC systems, which form a subset of the causes leading to energy waste in buildings, also lead to thermal discomfort of occupants.

Fault Detection and Diagnosis (FDD) is a process that discovers the presence of these operational faults in a timely manner and helps prevent the wastage of energy and the thermal discomfort to occupants that accompanies these faults. FDD comprises of two steps, i.e. detection and diagnosis. Detection is the process of checking if the behaviour of the HVAC system has diverged from expected operating patterns and diagnosis is the process of determining the root cause of the divergence [7]. As explained by A. Taal et al. [8], the process of FDD is similar to the way a doctor diagnoses a disease based on symptoms. First, symptoms are detected and then the fault that caused the symptom is diagnosed. Hence fault detection is also known as symptom detection.

Despite an increase in the amount of research on FDD, the deployment of real-time FDD tools in buildings is still at a nascent stage. The lack of sufficient sensors in buildings and their poor accuracy (limited by initial cost considerations) are key, critical factors impeding field testing and real-building deployment of FDD tools [9], [10]. Sensors play a very important role in the detection of symptoms through comparison of values recorded by the sensor under the presence of the fault with e.g. setpoints or baseline operating values. For example, the absence of a Mixed Air Temperature (MAT) sensor in the Air Handling Unit (AHU) would lead to 9 out of 28 symptom detection rules not being evaluated in the AHU Performance Assessment Rules (APAR) [11] leading to the variation of the applied FDD method with number of sensors and location of sensors [12]. Therefore, it is imperative that buildings are equipped with a standardised set of sensors, which are also highly accurate, to allow the deployment of generic FDD tools. As a first step towards this goal, the following studies were performed within the Brains for Buildings' Energy Systems (B4B) project:

i. To identify the sensors needed to detect faults occurring in Air Handling Units using simulation

ii. To evaluate performance of Low-Cost Sensors (LCSs) and Low-Cost Monitors (LCMs)- devices that integrate multiple LCSs to measure multiple indoor air parameters with a strong focus on fit-for-purpose for FDD application.

The selection of sensors required to detect symptoms depends on the relationship between faults and symptoms. These relationships exist in terms of probabilities- one fault could cause multiple symptoms and the same symptom could be caused by multiple faults [13]. The knowledge of HVAC system experts is generally used to deduce the relationships between faults and symptoms. There are also instances of experiments being used to determine the trends of sensor measurements (symptoms) under faulty conditions such as in Breuker and Braun [14], Comstock et al. [15], and Cho et al. [16]. Simulation serves as a very useful tool to identify the symptom effects of faults. It offers the opportunity to virtually install many sensors measuring multiple variables (a luxury that cannot be afforded in reality outside of living labs) and perform simulations under faulty conditions for extended durations, under various severities, across various outdoor climate conditions and help make an informed decision on the choice of sensors needed for FDD. Hence, a simulation-based study was performed.

An economically viable way to address the issue of poor sensor accuracy is to use Low-Cost Sensors (LCSs) and Low-Cost Monitors (LCMs)- devices that integrate multiple low-cost sensors to measure multiple parameters. While there has been an increase in the LCS and LCM technologies available in the market [17], [18], it is very important that the performances of these LCS and LCM technologies are evaluated by quantifying the fit-for-purpose before using them. Various studies exist in the literature that have compared the performance

of LCSs and LCMs against Research Grade Instruments (RGIs) in laboratories and field studies as reviewed by Sá et al. [18]. Recent studies by I. Demanega et al. [19] and H. Zheng et al. [20] also focused on evaluating LCSs/LCMs. The focus of these studies has been on monitoring Indoor Environmental Quality (IEQ) with Particulate Matter (PM) being the most studied parameter. However, the performance accuracy of LCSs and LCMs for the purpose of FDD applications is yet to be investigated and is a gap in the literature which was addressed through the study performed in the B4B project. By evaluating the performance of the LCSs and LCMs for their suitability to FDD, the best performing candidates can be used by building owners to install missing sensors.

Once the required sensors have been installed, the next step is to use them for fault (symptom) detection. Y. Zhao et al. [9] classify the methods of symptom detection into knowledge-driven methods and data-driven methods. Knowledge-driven methods include first-principles models-based methods (white box or grey box models) and rule-based methods. Data-driven methods, which has been a key focus of the growing literature on FDD [4], can be further categorized into supervised learning-based methods such as regression or classification and unsupervised learning-based methods such as clustering or principal component analysis. Regression-based methods use a black box model, whose input is a set of features, to predict the fault-free value of the target variable and if this predicted value deviates from the actual value of this variable by a specific threshold, then a symptom is deemed to have been detected. Such a black box model would require fault-free historical data on which it can be trained so that it can predict fault-free values for the target variable in future. Machine Learning (ML) algorithms that can be used to train such a black-box model include Decision Trees (DTs), Support Vector Machine (SVM), Neural Networks (NN), Ensemble Learning which are also some of the commonly used algorithms in the literature [10]. Considering the wide range of choices available and the advantages and disadvantages that come with them, it is important to compare the performance of these algorithms for the application of time-series prediction with the specific focus on the use for symptom detection. Therefore, a study to evaluate the performance of multiple ML algorithms for symptom detection was also performed within the B4B project. The results of this study is also presented in this report.

2 SENSORS REQUIRED IN AIR HANDLING UNITS FOR FAULT DETECTION

This chapter has been written as a manuscript titled "Gopalan, S., & Kramer, R. P. Statistical significance testing to determine sensors needed for Air Handling Unit fault detection: A simulated case study using EnergyPlus" to be published in a peer-reviewed scientific journal.

2.1 Introduction

Y. Chen et al. (Chen et al., 2022) performed a study quantifying the symptom effects of Fan Coil Unit (FCU) faults through two novel Key Performance Indicators (KPIs) that they introduced: Symptom Occurrence Probability (SOP) - a weighted summation of the ratio of symptom occurrences to duration of simulation of different operating windows divided based on outdoor air temperature- and Symptom Daily Continuous Duration (SDCD)- which indicates the duration a symptom is continuously present throughout the day. They quantified the symptom effects observed at 17 sensors for 18 different faults of multiple severity levels pertaining to cooling & heating coils and valves, ventilation (dampers, fan, filter), sensor bias and control systems. The study was performed for a simulation duration of one year. Y. Chen et al. [22] followed up their first study with another study where they calculated a KPI called Sensitivity Index (SI) for faults pertaining to an Air Handling Unit-Variable Air Volume (AHU-VAV) system. The SI is calculated using the probability of fault occurring at a particular severity and the SOP values of the symptoms it causes. Here again, the positive and negative symptoms are treated separately and the SI values for the positive and negative symptoms were calculated separately. Subsequently, the aggregated SI for each fault can be calculated by summing up the positive and negative SI values. This study was performed for faults pertaining to valves, sensor bias, damper stuck, coil fouling and control systems. A total of 56 measurements across AHU & VAV were used. The probabilities for faults & their severity levels were obtained from expert knowledge. A recent study by J. Cha et al. [23] performed a fault impact analysis involving the categorization of faults into three levels based on the level of symptoms observed for each fault. They also analysed fault impacts at various fault intensity values and developed an FDD using classification algorithms. This study only employed six symptoms- electric energy consumption, gas energy consumption, peak electricity load, peak gas load, average Predicted Percentage Dissatisfied (PPD) & setpoint unmet hours while occupied. It is quite difficult to diagnose the specific fault that would lead to a symptom at these indicators because they are affected by any HVAC system fault. The use of temperature sensors, mass flow rate meters and other control signals, as employed in Y. Chen et al. (Chen et al., 2022) and Y. Chen et al. [22], would help in accurately diagnosing component level faults.

Based on the review of Y. Chen et al. (Chen et al., 2022) and Y. Chen et al. [22], two gaps emerge. First, they used threshold rules based on mean (μ) ± standard deviation (σ) to detect symptoms, with thresholds of σ and 3σ, respectively. However, these rules assume normal distribution of data and 1σ risks high false positives. Alternatives such as the Median (M) and Median Absolute Deviation (MAD) exist [24]. Second, it is common practice in data pre-processing to remove outliers using the 3σ rule. When applied to fault-free data, outliers can be deemed as symptoms. Symptoms caused by faults should therefore be compared against symptoms observed in fault-free operation to confirm the effects of faults, which the prior studies did not address.

Therefore, this study has the following objectives:

1. To identify a suitable rule that can be used to detect the presence of a symptom using descriptive statistics.
2. To perform statistical significance testing of the faults' effects on the occurrence of symptoms at each sensor. This is performed by simulating counterfactual scenarios of faulty and non-faulty conditions. This will conclusively identify if the applied fault affects the occurrence of symptoms, adding robustness to the method.
3. To use the results of the statistical significance tests for developing fault-symptom relationships. These relationships can be used by building owners to install the required sensors to make their building FDD-ready.

The novel aspects of the research presented in this chapter are:

(i) The calculation of symptom occurrences in the fault-free case to identify the suitable rule for symptom detection using descriptive statistics, and

(ii) The application of statistical significance tests to establish the effect of faults on symptom occurrences.

In addition to these novel aspects, this research also performed a replication of the work of Y. Chen et al. [21] by calculating the SOP and SDCD values for the faults implemented in the case study.

The chosen system for focus is the Air Handling Unit (AHU), because of its importance in conditioning the indoor air space [25] and the high potential for implementing FDD solutions due to the high cost of diagnosis for complex control strategies and the mismatch between the number of AHUs and maintenance personnel [12].

2.2 Methodology

In this section, the methodology followed to achieve the stated research objectives of obtaining minimum symptoms in the baseline simulation and determining the statistically significant relationship between faults and symptoms is presented and explained. Figure 1 illustrates the eight phases of this study.

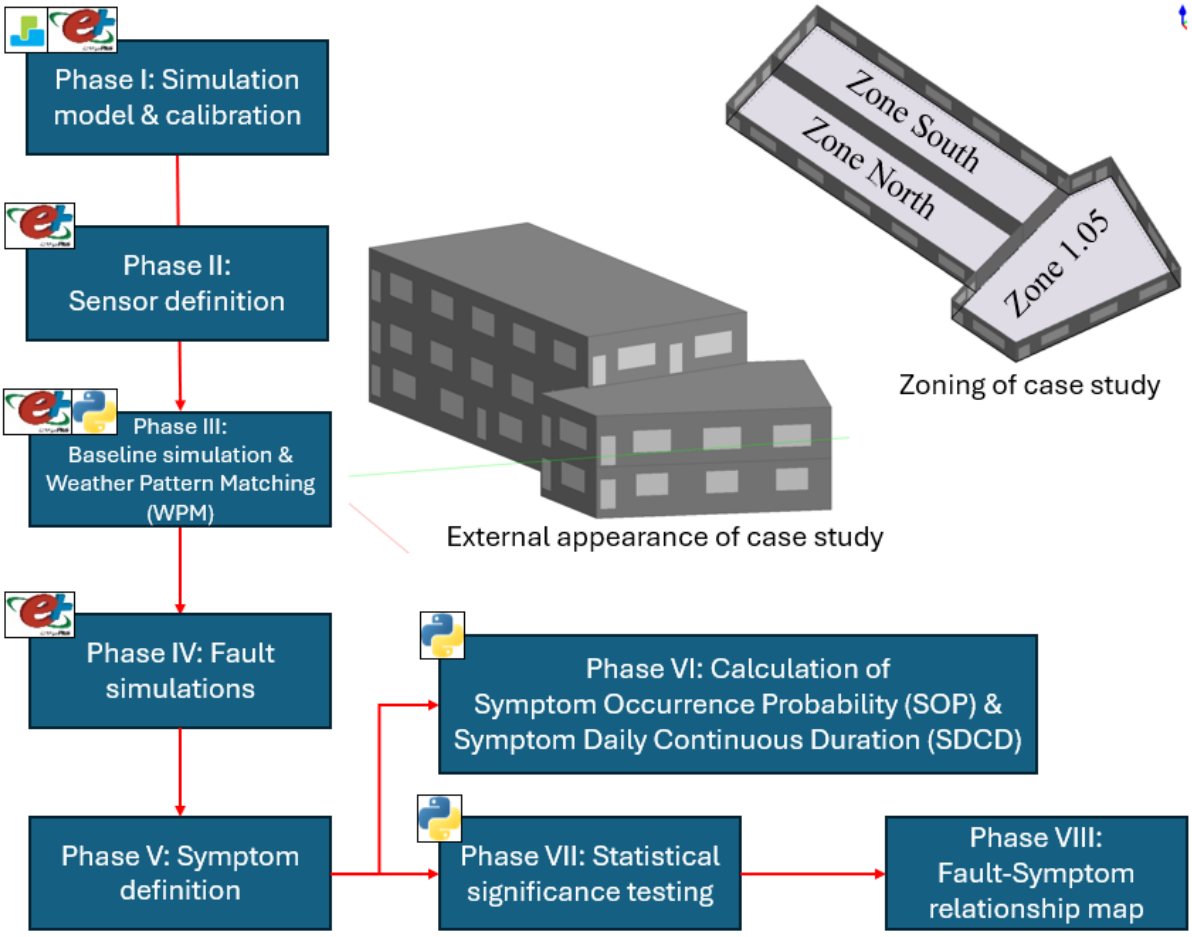

Figure 1: Overview of the methodology employed in this study; the external appearance and zoning of the case study, adapted from [26] has also been provided.

2.2.1 Phase I: Simulation model and calibration

The case study chosen for this research was the office of the building services firm Kropman B.V., situated in the city of Breda in The Netherlands. The external appearance of the case study and the zoning of its HVAC system is illustrated in Figure 1. Readers are referred to Gopalan et al. [27] for detailed information on the construction materials and their characteristics used for developing the simulation model. The HVAC system consists of an AHU which contains a supply and return fan, a supply and return filter, a Heat Recovery Wheel (HRW), and a heating coil. The cooling coils are present on the distribution side as after-coolers. Figure 2

illustrates the AHU of the case study. In addition to the heating coil, there are radiators present in the three zones. A detailed description of the zoning, control strategy and the HVAC system working of the case study can be found in Gopalan et al. [27]. The case study was modelled using DesignBuilder and its energy performance was simulated using EnergyPlus.

In Gopalan et al. [27], the authors presented the simulation and calibration results of the model for one day in the heating season and one day in the cooling season. In this research the model is simulated for a longer period: 6 months between October 1, 2022 to March 31, 2023 were chosen for simulating the heating season baseline and faults and a period of 3.5 months between June 1, 2023 to September 15, 2023 were chosen to simulate the cooling season baseline and faults. The weather file of Gilze-Rijen, a city located 15 km from the location of the building, was used for the simulation and it was obtained from the Royal Netherlands Meteorological Institute (KNMI).

The baseline models were calibrated according to the American Society of Heating Refrigerating and Air-conditioning Engineers (ASHRAE) Guideline-14 using the error metrics Normalised Mean Bias Error (NMBE) and Coefficient of Variation of the Root Mean Square Error (CV-RMSE) whose formulae, taken from Gestwick and Love [28] and ASHRAE [29], are provided in Eq. (1) and Eq. (2):

$$NMBE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{(n-p)\times\bar{y}} \times 100 \quad \text{Eq. (1)}$$

$$CV-RMSE = \frac{[\sum_{i=1}^{n}((y_i - \hat{y}_i)^2)/(n-p)]^{1/2}}{\bar{y}} \times 100 \quad \text{Eq. (2)}$$

in which

y_i is the measured data point,
$\hat{y}_i$ is the simulated data point,
$\bar{y}$ is the mean value of the measured data points,
n is the number of data points, and
p is the number of parameters in the model (0 for NMBE & 1 for CV-RMSE) [28].

The calibration in accordance with ASHRAE guideline 14-2014 was performed using monthly energy consumption data. The case study is a living lab office building supporting experimental activities for various smart building applications. Experiments were ongoing from February to March 2023 and from July to September 2023. Therefore, the heating season calibration of the baseline model was performed with data from October 2022 to January 2023 and the cooling season calibration of the baseline model was performed for the month of June 2023. Table 1 presents the error metrics for the heating and cooling season models, which is within the limits of ±5% for NMBE and 15% for CV-RMSE as defined in the guideline. It must be noted that since the cooling season model was calibrated only for the operation during one month, CV-RMSE could not be calculated because the value of p used for CV-RMSE calculation is 1 which would lead to a division by 0 in this case.

Table 1: Calibration results according to ASHRAE guideline 14-2014 for the Simulation model

Model	Error Metric	HVAC Electricity	Natural Gas (Boiler)	Chiller (Electricity)
Heating season Oct. 2022 - Jan. 2023	NMBE	-1.85%	1.03%	-
	CV-RMSE	5.97%	8.04%	-
Cooling season June 2023	NMBE	-	-	-4.77%
	CV-RMSE	-	-	-

2.2.2 Phase II: Sensor definition

EnergyPlus allows users to define multiple types of sensors & energy meters at several locations in the HVAC system. Figure 2 illustrates the sensoring locations and the types of sensors that were defined and used. Table 2 lists the full form of the measure variables and their short forms which are used to refer to them in the subsequent sections. The data from the simulation was recorded at a 1-minute interval. The mass flow rate through the heating/cooling coils, the supply water temperature and return water temperature values were used to calculate the energy exchanged at the coils which were also used as energy transfer sensors to detect symptoms. It must be noted that the use of the heating coil supply water temperature sensor to detect faults depends on the hydronic scheme. The results obtained in this research are case-specific. Similarly, the AHU in

the case study operates to meet the IAT requirements and hence IAT sensor is studied as a symptom. In cases where the AHU does not operate to meet IAT requirements, the IAT sensor is not needed to detect symptoms.

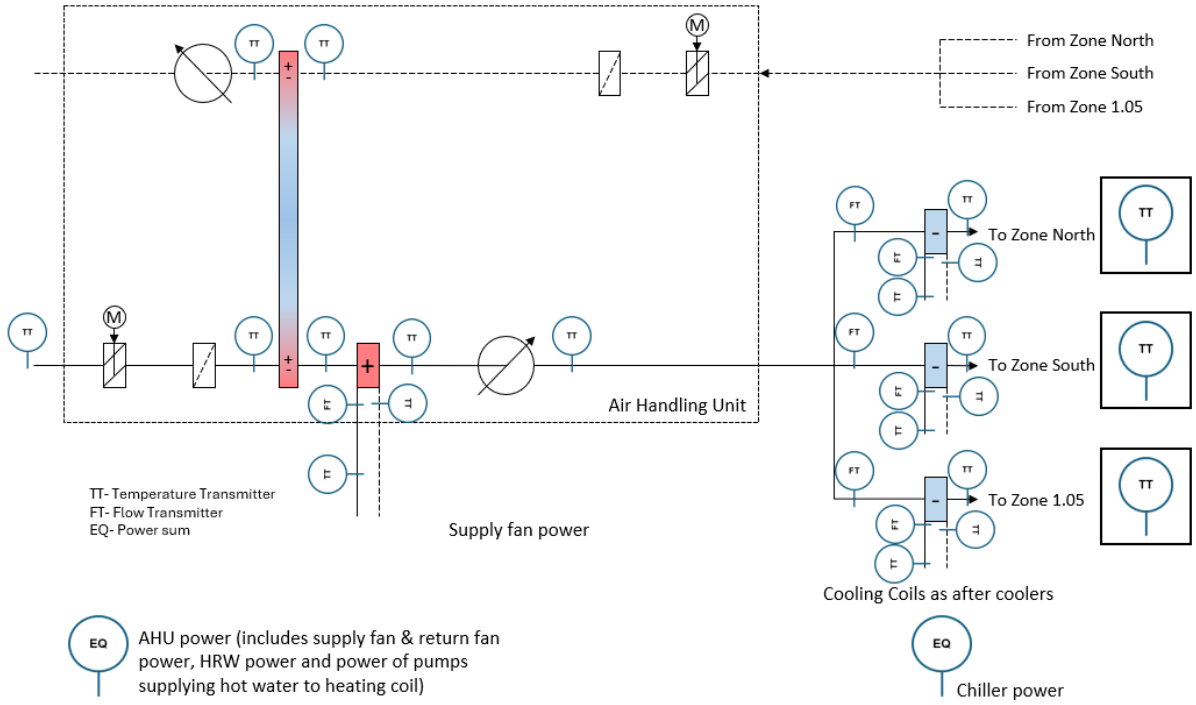

Figure 2: Sensors and energy meters available in case study building defined in the EnergyPlus environment.

Table 2: Sensors and energy meters list.

Heating season simulation Measured variable	Short form	Cooling season simulation Measured variable	Short form
HRW supply air entry Temperature	HRW supply air entry T	Air flow rate North	Airflow North
HRW supply air exit Temperature	HRW supply air exit T	Cooling Coil North Supply Water Temperature	CC North SWT
HRW return air entry Temperature	HRW return air entry T	Cooling Coil North Mass Flow Rate	CC North MFR
HRW return air exit Temperature	HRW return air exit T	Cooling Coil North Return Water Temperature	CC North RWT
Heating Coil Supply Water Temperature	HC SWT	Cooling Coil North Energy Transfer	CC North Energy Transfer
Heating Coil Mass Flow Rate	HC MFR	Supply Air Temperature North	SAT North
Heating Coil Return Water Temperature	HC RWT	Indoor Air Temperature North	IAT North
Heating Coil Energy Transfer	HC Energy Transfer	Air flow rate South	Airflow South
Heating Coil leaving air Temperature	HC leaving air T	Cooling Coil South Supply Water Temperature	CC South SWT
Supply Air Temperature AHU	SAT AHU	Cooling Coil South Mass Flow Rate	CC South MFR
Air flow rate North	Airflow North	Cooling Coil South Return Water Temperature	CC South RWT
Air flow rate South	Airflow South	Cooling Coil South Energy Transfer	CC South Energy Transfer
Air flow rate 1.05	Airflow 1.05	Supply Air Temperature South	SAT South
Supply Air Temperature North	SAT North	Indoor Air Temperature South	IAT South
Supply Air Temperature South	SAT South	Air flow rate 1.05	Airflow 1.05
Supply Air Temperature 1.05	SAT 1.05	Cooling Coil 1.05 Supply Water Temperature	CC 1.05 SWT
Indoor Air Temperature North	IAT North	Cooling Coil 1.05 Mass Flow Rate	CC 1.05 MFR
Indoor Air Temperature South	IAT South	Cooling Coil 1.05 Return Water Temperature	CC 1.05 RWT
Indoor Air Temperature 1.05	IAT 1.05	Cooling Coil 1.05 Energy Transfer	CC 1.05 Energy Transfer

Heating season simulation Measured variable	Short form	Cooling season simulation Measured variable	Short form
AHU Power	AHU Power	Supply Air Temperature 1.05	SAT 1.05
		Indoor Air Temperature 1.05	IAT 1.05
		Chiller Power	Chiller Power

2.2.3 Phase III: Baseline simulation and Weather Pattern Matching

Following the definition of the sensors in the simulation environment, the baseline models were simulated. The heating season model was simulated for 6 months between October 2022 – March 2023 and the cooling season model was simulated for 3.5 months between June 2023 – September 15, 2023. Once the simulations were completed, the WPM method was implemented. In this method the entire range of ambient air temperatures of the heating and cooling seasons were divided into smaller windows covering a small range of ambient air temperature each. Table 3 shows the range of ambient air temperatures in each window for the heating and cooling seasons. The heating season was divided into six windows and the cooling season was divided into five windows. All the output values recorded by the HVAC system sensors were divided and placed into the respective WPM-windows based on the recorded corresponding ambient air temperature. Subsequently, for each sensor, the mean and standard deviation of the values within each window were calculated, which were used to detect the presence of a symptom as explained subsequently.

Table 3: Weather Pattern Matching ambient air temperature windows for simulation model.

Heating Simulation (6 months)			Cooling simulation (3.5 months)		
Ambient air temperature range	Duration (minutes)	Operating Ratio (%)	Ambient air temperature range	Duration (minutes)	Operating Ratio (%)
-8.1°C to -3.9°C	753	1.05	15.2°C to 18.2°C	2,886	8.22
-3.8°C to 0.4°C	4,331	6.06	18.3°C to 21.3°C	11,339	32.30
0.5°C to 4.7°C	12,448	17.43	21.4°C to 24.4°C	9,601	27.35
4.8°C to 9.0°C	22,819	31.96	24.5°C to 27.5°C	7,637	21.76
9.1°C to 13.3°C	21,381	29.94	27.6°C to 30.8°C	3,637	10.24
13.4°C to 17.8°C	9,668	13.54			
Total	**71,400**			**35,100**	

2.2.4 Phase IV: Fault simulations

As mentioned in section 1, this study focused on the AHU, specifically on faults pertaining to heating coil, cooling coil, fans, heat recovery wheel, sensor bias (offset) and incorrect SAT setpoints. The studied faults and their severity levels, are listed in Table 4. These fault severities were adopted from Gopalan et al. [27] who in turn had chosen these severities based on a literature study. It must be noted that depending on whether the AHU operates to meet IAT requirements, the IAT sensor bias fault would be relevant. The results obtained for the IAT sensor bias fault in this study are case specific.

Table 4: AHU faults and their severities studied

Component	Fault	Fault settings studied	EnergyPlus feature used for fault modelling
Heating coil & valve actuator	Heating coil valve stuck	0%, 100%	Energy Management System (EMS): *Coil valve stuck*- the minimum, maximum & actual mass flow rate of water through the coil was set at a value corresponding to the valve stuck % *Coil valve leak*- the minimum mass flow rate of water through the coil was set at a value corresponding to the valve leak % *Fan failure/stuck*- the air mass flow rate set at a value corresponding to the airflow %
	Heating coil valve leak	0%, 40%	
Cooling coil & valve actuator	Cooling coil valve stuck	0%, 50%	
	Cooling coil valve leak	0%, 40%	
Fans	Fans failure	0% airflow	
	Fans stuck	50% airflow, 85% airflow	
Heat recovery wheel (only for heating season)	Wheel failure	0% rotation	Heat Recovery (availability schedule set to 'Off 24/7')
Sensor bias	Supply air temperature sensor bias	±2°C bias	Operational Faults (offset of ±2°C introduced)
	Indoor air temperature sensor (thermostat) bias	±2°C bias	
Control setpoint	Incorrect supply air temperature setpoint (high)	+2°C	Setpoint Managers: minimum & maximum setpoint values changed for the heating model;

Component	Fault	Fault settings studied	EnergyPlus feature used for fault modelling
	Incorrect supply air temperature setpoint (low)	-2°C	setpoint values at outdoor low and high temperatures changed for the cooling model

2.2.5 Phase V: Symptom definition

Following the fault simulations, symptom conditions were defined. If the value of a data point belonging to a particular WPM-window, and recorded by a specific sensor, deviates from the mean of the values recorded by the same sensor in the corresponding WPM-window of the baseline model by a specific factor of the standard deviation, then a symptom occurrence can be detected. In the introduction section it was mentioned that the use of mean and standard deviation suits normally distributed data as 99.7% of the data will fall within three times the standard deviation from the mean. However, the data recorded by HVAC system sensors may not be normally distributed. Therefore, in this research, median and Median Absolute Deviation (MAD) were also explored as an alternative rule to define the symptom and compare the number of symptoms obtained with the number of symptoms obtained when using mean and standard deviation.

Following this, the reference value recorded by each sensor within each operating window can be either chosen as the mean value or the median value. Subsequently, a symptom can be defined in statistical terms as in Eq. (3):

$$|y_i - \mu| > t.\sigma \quad \text{Eq. (3)}$$

where

y_i is a datapoint measured by a sensor belonging to a particular WPM-window,
μ is the mean measurement of the baseline data from the same sensor (belonging to the same operating window),
σ is the standard deviation of the baseline measurements made by the same sensor in the same operating window, and
t = 1, 2, 3…

Alternatively, a symptom can be detected using the Median (M) and Median Absolute Deviation (MAD) values, according to Eq. (4):

$$|y_i - M| > t.MAD \quad \text{Eq. (4)}$$

where

y_i is a datapoint measured by a sensor belonging to a particular operating window,
M is the median measurement of the baseline data from the same sensor belonging to the same WPM-window,
MAD is the Median Absolute Deviation of the measurements made by the same sensor in the same operating window,
t = 1, 2, 3…

The previous research by Y. Chen et al. (Chen et al., 2022) and Y. Chen et al. [22] used μ and σ for symptom definition. Y. Chen et al. (Chen et al., 2022) used t = 1 and Y. Chen et al. [22] used t = 3. In this research, a comparison of the number of symptoms obtained when applying the above two formulae for values of t = 1 and t = 3 for the baseline model measurement of each sensor was performed. When the above symptom definitions are applied to fault-free data, some data points (outliers) would fall outside the defined thresholds and be deemed symptoms. The aim is to identify the rule that generates the fewest symptoms across all the studied sensors in the fault-free case. This identified rule was used as the definition of symptom. The symptom occurrences under fault-free conditions (baseline) are either caused by faults outside the scope of this research (which inherently exist in the HVAC system) or they are regular outliers. In baseline model, no other fault was introduced. Therefore, it can be concluded that these symptoms must be outliers. Outliers are known as anomalies when detecting events of interest [30]. The probability of anomaly occurrence is expected to be only slightly greater than zero [31]. Hence, when choosing the suitable symptom detection rule, the one that resulted in the least occurrence of symptoms in the fault-free case was chosen.

2.2.6 Phase VI: SOP and SDCD calculations

Following the establishment of the causality between faults and symptoms, the KPIs SOP & SDCD, developed by Y. Chen et al. (Chen et al., 2022), were used to quantify the symptom effects of faults. SOP was defined as a two-step process by Y. Chen et al. (Chen et al., 2022) where they calculated the SOP within each WPM-window as the ratio of the duration the symptoms were observed to the duration of operation within each window in the first step and then used the average SOPs within all windows weighted by the operating ratio of the respective windows. However, as demonstrated in Appendix A, this method can be simplified to the ratio of total duration of symptoms observed to total operation duration. Therefore, the following simplified formula for SOP was used:

$$SOP = \frac{total\ number\ of\ symptoms\ observed}{total\ duration\ of\ operation} \quad \text{Eq. (5)}$$

The SDCD is the maximum continuous duration a symptom is observed for a day of the simulation. The KPI that is analysed is the mean of all SDCD values occurring over the duration of the simulation, according to:

$$mean\ SDCD = \frac{\sum_{i=1}^{n} SDCD_i}{n} \quad \text{Eq. (6)}$$

where

i is each individual day from the simulated duration
n is the today number of days the model is simulated
$SDCD_i$ is the SDCD of the ith simulated day

The mean of SDCD values observed for all days is shown as the result in this research. The SOP and SDCD can inform engineers, facility managers and maintenance technicians on which sensor should be given a priority when diagnosing a fault when that fault causes multiple symptoms.

2.2.7 Phase VII: Statistical significance testing

As discussed in section 2.1, a key consideration that was overlooked in the work of Y. Chen et al. (Chen et al., 2022) and Y. Chen et al. [22] is the occurrence of symptoms even in the fault-free case. This is different from the symptoms that would be observed when the model is simulated under the presence of a fault. The differences between the occurrence of symptoms observed in the fault-free cases and faulty cases can be conclusively established and attributed to the faults by using Null Hypothesis Statistical Testing (NHST), or hereafter simply referred to as statistical significance testing. In this test, the alternative hypothesis that was formulated is that the difference in symptom effects observed in the sensor measurements was due to the respective faults introduced. The null hypothesis is that the difference in symptom effects observed is not due to the faults. The sample set used for statistical testing were the number of minutes the symptoms were observed during each operating day of the simulation. In this way, the faults were treated as events that occurred during the operating hours of the building every day. There were 119 operating days in the heating season simulation and 65 operating days in the cooling season simulation, leading to 119 samples and 65 samples respectively for the heating and cooling seasons for each sensor for each fault. The operating duration for each day in the heating season totalled 600 minutes. During the cooling season, the ventilation operates for 10 hours a day between 07:00 h – 17:00 h, but the cooling system only operates for 9 hours a day between 08:00 h – 17:00 h. Therefore, data of only 540 minutes of operation were considered for the cooling season simulation. Consequently, the maximum number of symptom minutes that can be obtained for each day in the respective cases were 600 minutes and 540 minutes. Figure 3 illustrates the events as considered for the statistical significance testing.

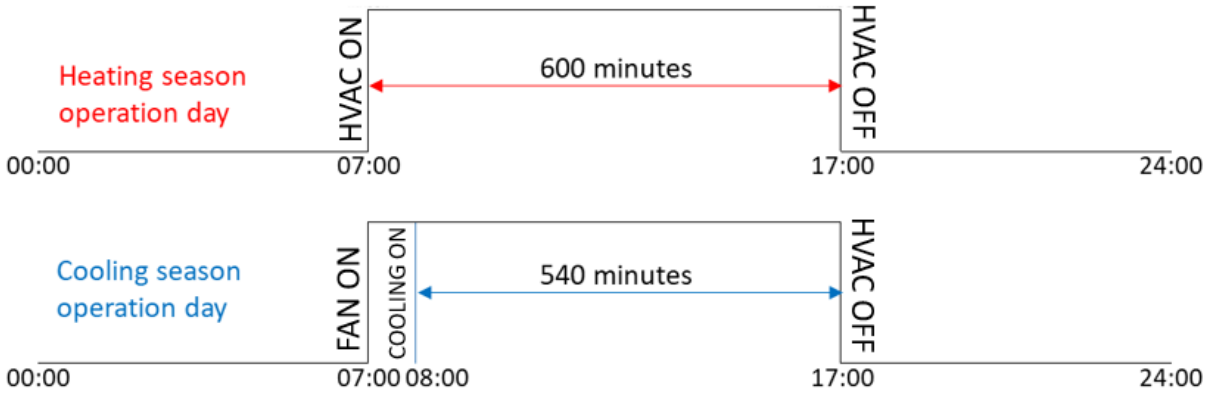

Figure 3: Event on heating and cooling day making a sample set consisting of symptom occurrences on each operating day.

This study employed the paired t-test to test the validity of the alternate hypothesis. Since the paired t-test was used for multiple hypothesis testing (for multiple faults across multiple sensors), it would lead to an increased chance of Type I error [32]. The Bonferroni correction post-hoc procedure was performed to overcome this and improve the robustness of the statistical significance testing results. The symptoms which returned a p-value lower than 0.05 were deemed to be statistically significant under the presence of the implemented faults. Similarly, the symptoms which returned a p-value higher than 0.05 were deemed to be statistically insignificant under the presence of the implemented faults. The effect size of the implemented fault on all the symptoms were studied using Cohen's *d*, according to Eq. (7):

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}} \quad \text{Eq. (7)}$$

where

$\bar{X}_1$ is the mean of the control group,
$\bar{X}_2$ is the mean of the test (faulty) group,
s_1 and s_2 are the standard deviations of control group and the test group,
n_1 and n_2 are the sample sizes of control group and the test group.

2.2.8 Phase VIII: Fault-symptom relationships

The results of the paired t-tests provided information if the occurrence of symptoms observed under the presence of a fault was significantly different from the occurrence of symptoms in the fault-free case. This information was used to identify relationships between faults and symptoms which would inform building owners and managers as well as HVAC system engineers about the sensors that need to be installed to make buildings FDD-ready.

2.3 Results

In this section, the results of the study are presented. The suitable rule to detect symptoms, determined using empirical evidence, are first presented. Subsequently, the SOP & SDCD results are presented followed by the results of the statistical significance tests. Finally, the fault-symptom relationships, identified from the results of the statistical significance tests, are presented.

2.3.1 Suitable symptom detection rule

Tables 5 & 6 present the SOP values obtained for the fault-free case of the heating season when using different thresholds with mean and median. It can be observed that using the median and median absolute deviation leads to the classification of a higher proportion of operation duration as symptoms when compared to using mean and standard deviation for the corresponding thresholds. Since the combination of mean and standard deviation with the threshold of 3 gave the least SOP values (and therefore, the least number of symptoms),

this combination was chosen as the method to detect the presence of symptoms in the simulated AHU operation.

Table 5: SOP values obtained for the fault-free case of heating season when using different thresholds with mean and median.

Symptom	$\mu \pm \sigma$	$\mu \pm 3\sigma$	$M \pm MAD$	$M \pm 3MAD$
AHU Power	23.9%	0.9%	40.8%	10.5%
HRW supply air entry T	40.8%	0.0%	50.0%	0.8%
HRW supply air exit T	32.9%	0.8%	50.0%	3.5%
HRW return air entry T	16.7%	1.7%	50.0%	17.6%
HRW return air exit T	39.2%	0.0%	50.0%	1.1%
Airflow South	33.3%	0.3%	50.0%	2.8%
Airflow North	33.3%	0.3%	50.0%	2.8%
Airflow 1.05	33.3%	0.3%	50.0%	2.8%
HC SWT	33.8%	0.1%	49.4%	5.1%
HC MFR	21.6%	0.5%	35.5%	14.1%
HC RWT	24.1%	1.1%	49.7%	14.9%
HC Energy Transfer	21.2%	0.5%	30.0%	13.5%
HC leaving air T	34.3%	0.2%	50.0%	16.7%
SAT AHU	34.3%	0.2%	50.0%	17.7%
SAT South	18.3%	1.6%	50.0%	18.2%
SAT North	11.0%	2.0%	50.0%	19.7%
SAT 1.05	27.8%	1.5%	50.0%	9.2%
IAT South	18.3%	1.6%	50.0%	18.2%
IAT North	11.0%	2.0%	50.0%	19.6%
IAT 1.05	27.8%	1.5%	50.0%	9.2%

Table 6: SOP values obtained for the fault-free case of cooling season when using different thresholds with mean and median.

Symptom	$\mu \pm \sigma$	$\mu \pm 3\sigma$	$M \pm MAD$	$M \pm 3MAD$
Chiller Power	34.1%	0.6%	46.7%	14.5%
Airflow South	30.5%	1.0%	50.0%	9.8%
CC South SWT	26.2%	0.5%	39.1%	18.9%
CC South MFR	13.8%	2.5%	31.0%	17.9%
CC South RWT	25.2%	0.7%	49.0%	12.5%
CC South Energy Transfer	28.0%	0.4%	48.0%	11.2%
SAT South	18.8%	1.4%	50.0%	27.9%
IAT South	22.7%	1.3%	50.0%	7.2%
Airflow North	31.1%	1.0%	50.0%	9.1%
CC North SWT	26.2%	0.5%	39.1%	18.9%
CC North MFR	13.6%	0.2%	29.2%	18.9%
CC North RWT	30.5%	0.8%	49.0%	8.0%
CC North Energy Transfer	27.1%	0.7%	48.0%	13.7%
SAT North	21.9%	1.3%	50.0%	20.2%
IAT North	25.5%	1.0%	50.0%	6.4%
Airflow 1.05	30.9%	1.0%	50.0%	7.7%
CC 1.05 SWT	26.2%	0.5%	39.1%	18.9%
CC 1.05 MFR	13.6%	2.5%	31.9%	21.0%
CC 1.05 RWT	29.5%	0.6%	49.0%	11.2%
CC 1.05 Energy Transfer	29.5%	0.5%	48.0%	11.8%
SAT 1.05	22.1%	1.2%	50.0%	25.4%
IAT 1.05	27.9%	1.0%	50.0%	5.2%

2.3.2 SOP & SDCD: heating season

This section presents the results obtained for SOP & SDCD calculated according to Eq. (5) and Eq. (6). Figure 4 provides the SOP obtained for the heating season faults. The highest SOP value of 100% is observed about 13 times across four faults. The fan faults lead to 100% SOP values at the airflow rate and AHU power meters. These meters show symptoms for 100% of the fault duration, not only for a severe fault like fan failure, but also for a very mild fault like fan stuck at 85% airflow. HRW failure leads to 100% SOP values at sensors measuring exit air temperatures from HRW both on the supply air side and the return air side. Fan failure also causes symptoms at these two sensors to be present for 100% of the fault duration, because the stoppage of fan operation leads to a stoppage in HRW operation too.

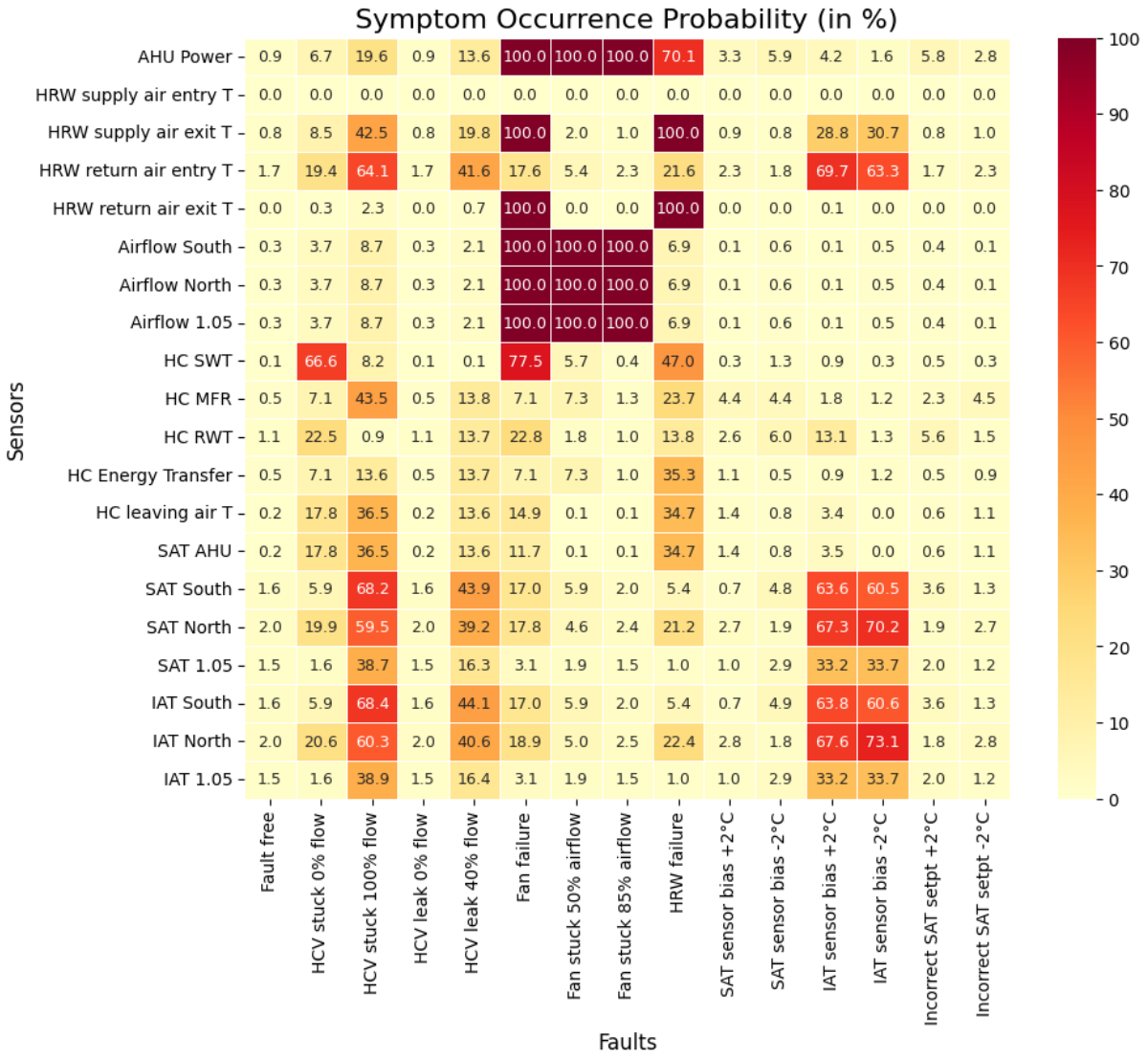

Figure 4: SOP values for heating season faults.

The next highest SOP values occur in the range of 59% to 78%. These can be observed for the SAT sensors and IAT sensors of the North and South zones, and HRW return air entry T sensor for the HCV stuck 100% flow fault and the IAT sensor bias fault (both positive and negative biases). Additionally, SOP values fall in this range for the HC SWT sensor for the HCV stuck 0% flow and fan failure faults, and the AHU power meter for the HRW failure fault. The next highest range of SOP values obtained are in the range of 28% to 47%. These can be observed at: (i) the SAT and IAT sensors of zone 1.05, and the HRW supply air exit T sensor for the HCV stuck 100% flow and the IAT sensor bias (both positive and negative) faults, (ii) the SAT and IAT sensors of North and South zones, and the HRW return air entry T sensor for the HCV leak 40% fault, (iii) the SAT AHU and the HC leaving air T sensor for the HCV stuck 100% flow and HRW failure faults, (iv) HC MFR and the HRW supply air exit T sensor for the HCV leak 100% flow fault, and (v) the HC energy transfer and HC SWT sensors for the HRW failure fault. There are also several symptoms whose values are in the range of 11% to 24%. The SOP values in the range of 0 to 10% occurs with the highest frequency.

The HCV leak 0% fault is the same as the fault-free case. Therefore, its SOP values across all sensors are the same. The sensor measuring temperature of entry air to HRW on the supply air side is placed before any of the heat recovery, heating or cooling activity. Therefore, this sensor does not show any symptom for any of the studied faults.

Figure 5 presents the mean SDCD values obtained for the heating season faults. The specific symptoms corresponding to the fan failure fault, fan stuck 50% airflow fault, fan stuck 85% airflow fault, and HRW failure fault, which have an SOP value of 100%, register mean SDCD values of 600 minutes as expected, i.e. these specific symptoms are observed during every minute of the HVAC system operation during the entire six

months of simulation performed. The symptoms with SOP values from 60% to 78% register mean SDCD values between 336 minutes and 453 minutes. The symptoms with SOP values from 28% to 47% register mean SDCD value between 127 and 257 minutes. The symptoms with SOP values from 11% to 24% register mean SDCD value between 66 and 118 minutes. The comparison of the SOP and SDCD values obtained for the heating season indicates that for symptoms having SOP of 11% or above, the mean SDCD value is more than 60 minutes.

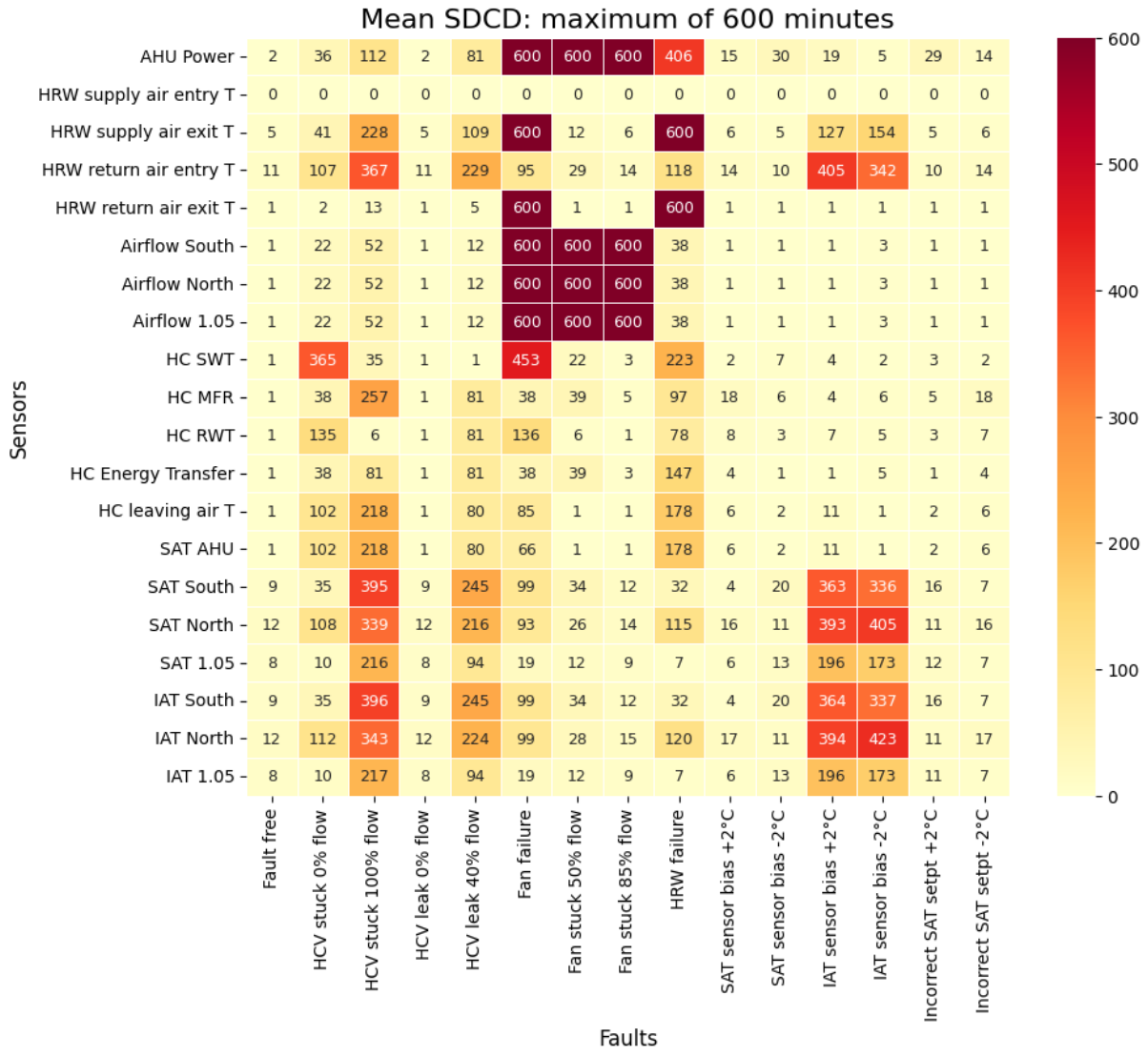

Figure 5: Mean SDCD values for heating season faults.

2.3.3 SOP & SDCD results: cooling season

Figure 6 presents the SOP values obtained for the cooling season faults. The highest SOP value of 100% is observed at the airflow meters across the three zones for the fan failure fault. The next highest SOP values are obtained in the range of 91.8% to 92.2% at the chiller power meter and the RWT sensor, CC energy transfer sensor and SAT sensor across all three zones. These symptoms are caused by the CCV stuck 0% flow and fan failure faults. The same two faults lead to: (i) SOP values of 83.9% and 84% at the IAT South sensor, (ii) SOP values of 76.2% and 77.7% at the IAT North sensor, and (iii) SOP values of 62.6% and 60.3% at the IAT 1.05 sensor. Three additional SOP values exceed 50%: the chiller power meter under the fan stuck 50% airflow fault, and the CC South RWT sensor under the SAT bias -2°C and incorrect SAT setpoint +2°C faults. There are 33 symptoms across various faults that have an SOP value between 30% and 41% and 23 symptoms across various faults that have an SOP value between 20% and 30%. There are eight symptoms having an SOP value between 10% and 20%. The remaining symptoms have an SOP value lower than 10%. No sensor

showed any notably different SOP values for the IAT sensor bias faults. The CCV leak 0% flow fault is the same case as fault-free case. Therefore, the obtained SOP values at all symptoms are equal for both cases.

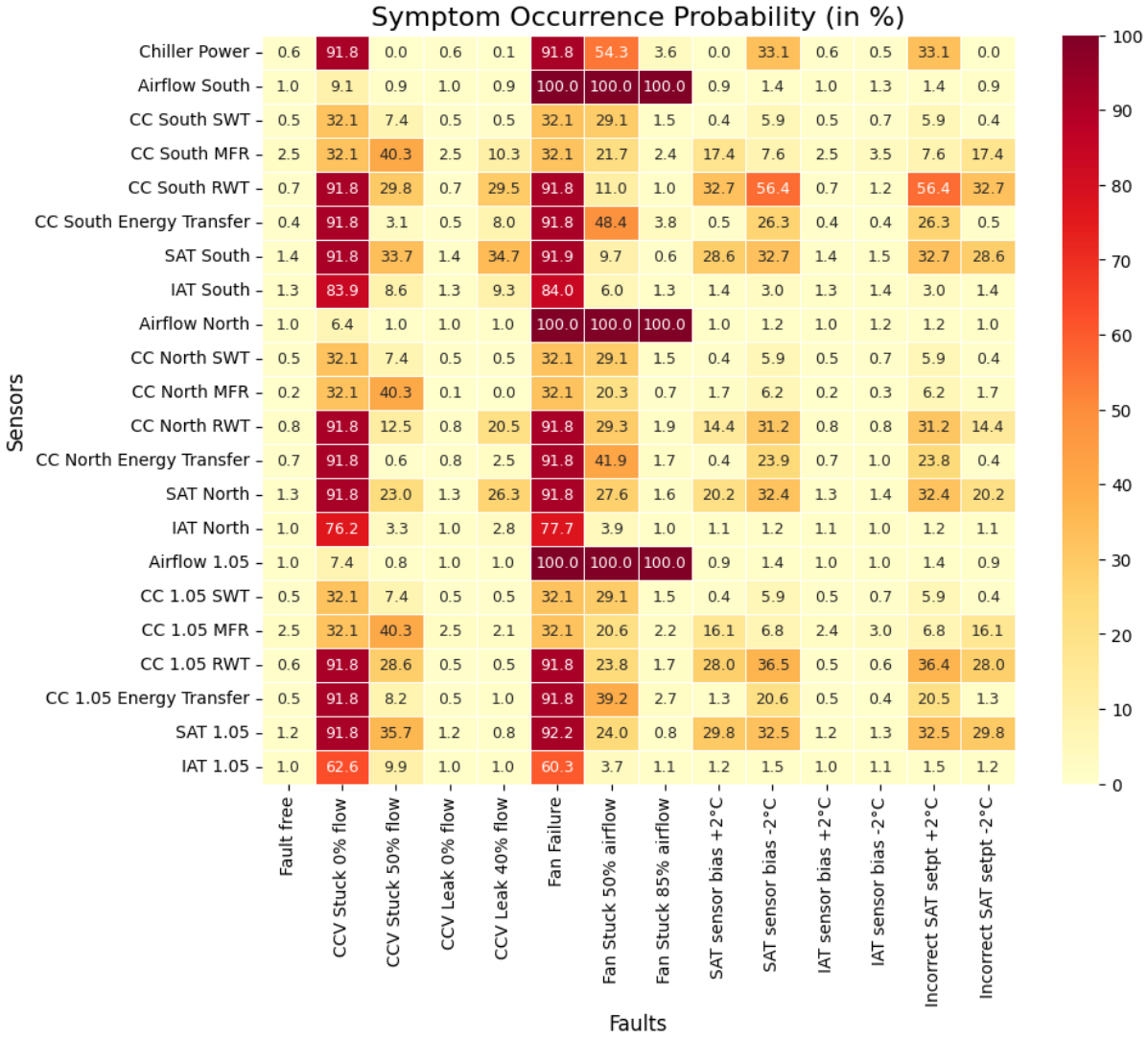

Figure 6: SOP results for cooling season faults.

Figure 7 provides the SDCD values obtained for the cooling season faults. The symptoms having 100% SOP in Figure 6 have a mean SDCD value of 540 minutes in Figure 7. The symptoms with SOP of greater than 90% have a mean SDCD value between 489 and 492 minutes. There are two symptoms with mean SDCD values between 350 and 370 minutes and eight symptoms with mean SDCD values between 200 and 300 minutes. There are 50 symptoms with a mean SDCD values in between 100 and 200 minutes. The remaining symptoms have a mean SDCD value of lower than 100 minutes. The comparison of the SOP and SDCD values obtained for the cooling season indicates that for symptoms having SOP of 14.4% or higher, the mean SDCD value is more than 60 minutes.

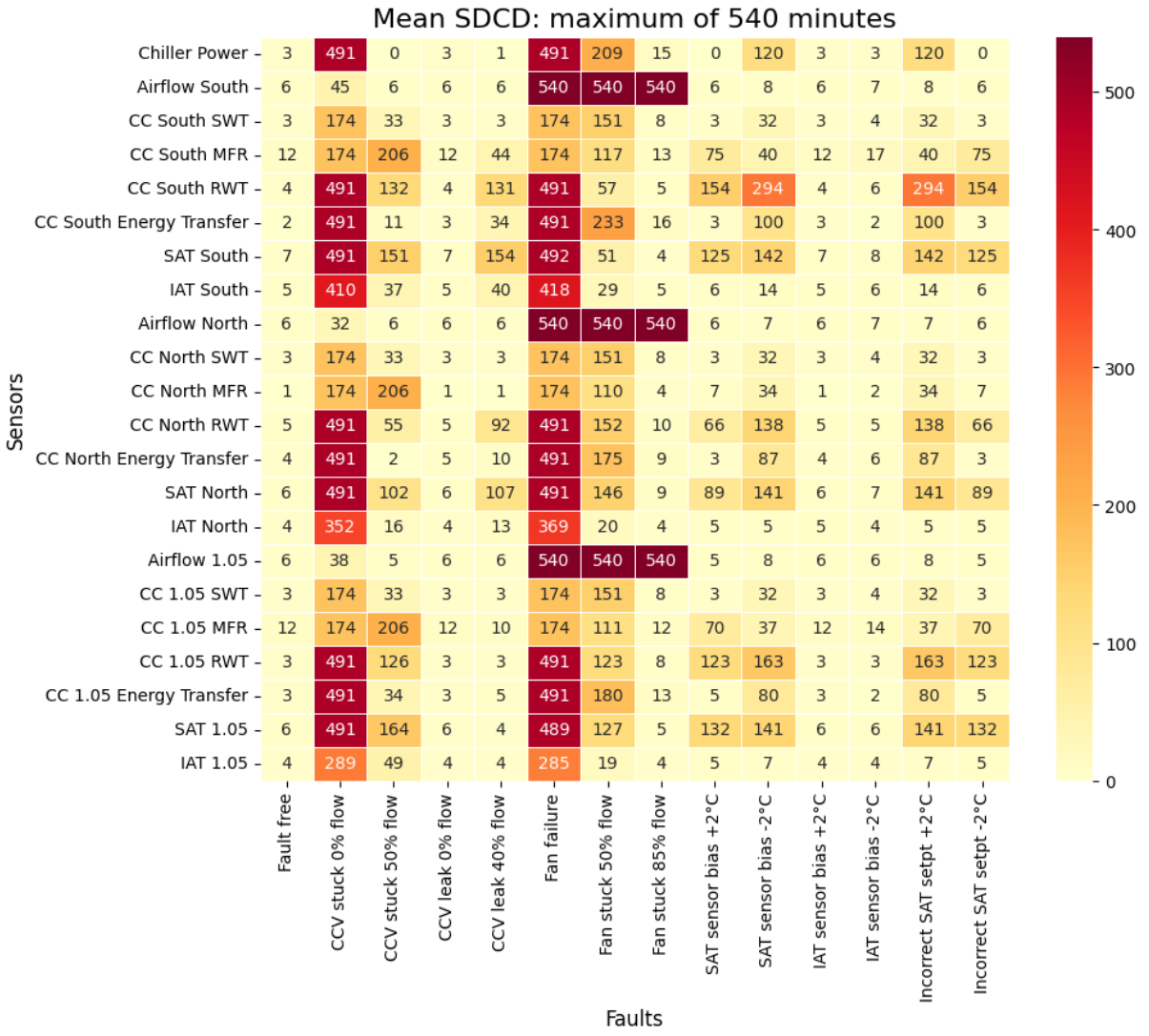

Figure 7: Mean SDCD results for cooling season faults.

2.3.4 Statistical significance testing

As explained in section 2.2.7, a paired t-test was performed followed by Bonferroni's correction to establish the statistical significance of the occurrence of symptoms attributed to the incorporation of faults. The effect size was calculated using Cohen's *d*, as explained in section 2.2.8.

Figure 8 illustrates the results of statistical significance testing and the Cohen's *d* values for the heating season faults. The green-coloured results indicate that the specific sensor under consideration shows a statistically significant difference in the occurrence of symptoms under the presence of the specific fault under consideration. This implies that this sensor can be used to detect the corresponding fault. The following are some of the key results:

i. The SAT AHU sensor & HC leaving air T sensor return the same result for the statistical significance testing for all the fault cases studied. They also have the same effect size values for all but one fault (fan failure). Any one of the two is sufficient for use in symptom detection.

ii. When considering the sensor bias and incorrect SAT setpoint faults (six cases in total), the HC MFR sensor and the HC RWT sensor show statistically significant differences in the occurrence of symptoms for more than one case each. However, the HC energy transfer sensor shows a statistically significant difference in the occurrence of symptoms only for one of these six fault cases. It must be noted that the HC SWT sensor also shows a statistically significant difference in the occurrence of symptoms only for one of these six fault cases.

iii. The results of the statistical significance testing obtained for the SAT AHU sensor did not translate directly to the corresponding results of the statistical significance testing obtained at the SAT sensors corresponding to the three different zones.

iv. The results obtained for the statistical significance testing at the three SAT sensors corresponding to the three different zones translated directly to the corresponding results obtained for the statistical significance testing at the three IAT sensors.

v. Among the IAT sensors, the sensor at North zone showed statistically significant differences in the occurrence of symptoms for the most of faults, followed by the sensor in the South zone and the sensor in zone 1.05.

vi. The AHU power symptom shows a statistically significant difference in the occurrence of symptoms for all but four faults- incorrect SAT setpoint -2 °C, IAT sensor bias -2 °C, SAT sensor bias +2 °C and HCV leak 0% flow. It must be noted that the power of the pumps delivering hot water to the heating coil is also included in this sensor. However, in practice, electricity meters for pumps are often lacking. Moreover, the electricity meters of the pumps (when available) could be separate meters.

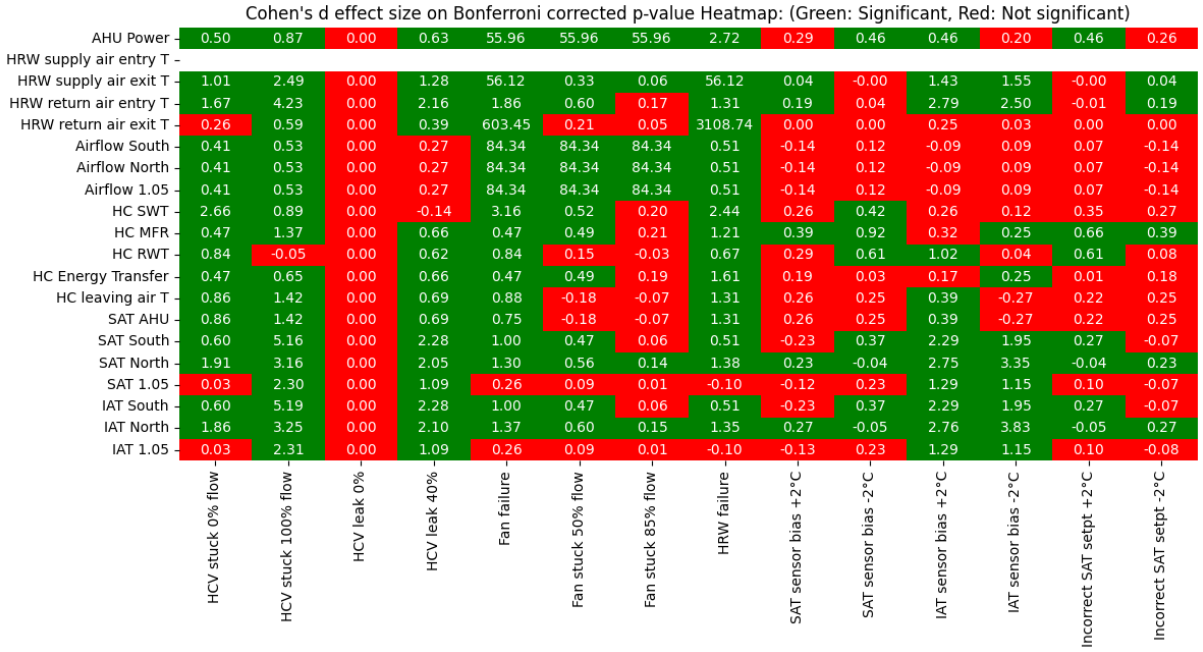

Figure 8: Result of statistical significance testing and effect size for heating season fault simulation. Significant results are marked green and insignificant results are marked red. The numeric values indicate the point estimate of the effect size (Cohen's d).

A closer inspection of Figure 8 shows some very interesting observations. These are enlisted below:

i. The HRW supply air exit T sensor shows a statistically significant difference in the occurrence of symptoms due to the fan stuck 85% airflow, SAT sensor bias +2 °C and incorrect SAT setpoint -2 °C faults. However, the SOP values are 1.0%, 0.9% and 1.0% respectively for these three faults when compared to an SOP value of 0.8% obtained for the fault-free case. Similarly, the effect sizes are only 0.06, 0.04 and 0.04 respectively which are quite small when compared to the effect sizes caused by the other faults at this sensor. There are two key factors to note here:
a. For a given statistical power, to detect a small effect, a larger sample size is needed [33]. The chosen sample size is powerful enough to detect such a small effect.
b. Only the point estimates of Cohen's *d* are shown in Figure 8. However, these values also have a 95% Confidence Interval (CI). A significant result implies that the 95% CI of Cohen's *d* value does not contain zero [34].

ii. ii. At the HRW return air entry T sensor, the SAT sensor bias +2 °C and incorrect SAT setpoint -2 °C faults show significant differences with identical effect sizes (0.19) and SOP values (2.3%). The fan stuck 85% flow fault yields the same SOP (2.3%) and a comparable effect size (0.17) but does not show a significant difference. This reflects the converse of the two factors explained in

the previous result: the sample size is not sufficient to detect this smaller effect and when the statistical significance test (paired t-test in this case) does not return a statistically significant result, the 95% CI of Cohen's *d* value contains zero .

iii. The HRW failure fault leads to a statistically significant difference in the occurrence of symptoms at the following sensors: HC Air leaving T sensor, SAT AHU sensor, SAT and IAT sensors at both North and South zones. In practice, it happens only when the heating coil is undersized i.e. when it cannot provide the required energy to meet the SAT setpoint requirements.

Figure 9 illustrates the results of statistical significance testing and Cohen's *d* for the cooling season faults. The green-coloured results indicate that the specific sensor under consideration shows a statistically significant difference in the occurrence of symptoms under the presence of the specific fault under consideration. This implies that this sensor can be used to detect the corresponding fault.

Figure 9: Result of statistical significance testing and effect size for cooling season fault simulation. Significant results are marked green and insignificant results are marked red. The numeric values indicate the point estimate of the effect size (Cohen's d).

The following are some of the key results:

i. The fan failure fault leads to a statistically significant difference in the occurrence of symptoms at all the sensors studied. The fan stuck at 50% flow fault leads to a statistically significant difference in the occurrence of symptoms at all but two of the sensors studied.

ii. The SAT sensor bias +2 °C and the incorrect SAT setpoint -2 °C faults lead to the same pattern of statistical significance in the occurrence of symptoms at the sensors studied. They also lead to the same effect size at the sensors studied. Likewise, the SAT sensor bias -2 °C and the incorrect SAT setpoint +2 °C faults lead to the same pattern of statistical significance in the occurrence of symptoms and the same effect size at the sensors studied.

iii. The IAT sensors in the three different zones respond differently to the different faults. The IAT South sensor shows a statistically significant difference in the occurrence of symptoms for five faults; the IAT 1.05 sensor shows a statistically significant difference in the occurrence of symptoms for three faults and the IAT North sensor shows a statistically significant difference in the occurrence of symptoms for two faults.

2.3.5 Fault-symptom relationships

In this section the fault-symptom relationships identified based on the paired t-tests conducted are presented. Some of the results obtained are contrary to the knowledge of how HVAC systems work in practice or are not

applicable to all HVAC systems. Hence, those results are left as blank in the results presented in this section and explained in detail in Appendix B.

Figure 10 shows the fault-symptom relationships for the heating season. The figure shows that the IAT North sensor can be used to detect 12 faults. The HC MFR sensor can be used to detect 11 faults. The HRW supply air exit T sensor can also be used to detect 11 faults. The AHU power meter and the HRW return air entry T sensor can be used to detect 10 faults. It must be noted that while the SAT sensors at North and South zones can be used to detect 10 faults, during the heating season, the SAT AHU sensor is the target variable that the heating coil controller aims to meet. This sensor can be used to detect 5 faults (possibly the HRW failure also if the coil is undersized). The SAT sensors installed in the downstream (distribution side) are purely for measurement during the heating season.

	HCV stuck 0% flow	HCV stuck 100% flow	HCV leak 0%	HCV leak 40%	Fan failure	Fan stuck 50% flow	Fan stuck 85% flow	HRW failure	SAT sensor bias +2°C	SAT sensor bias -2°C	IAT sensor bias +2°C	IAT sensor bias -2°C	Incorrect SAT setpt +2°C	Incorrect SAT setpt -2°C
AHU Power														
HRW supply air entry T														
HRW supply air exit T														
HRW return air entry T														
HRW return air exit T														
Airflow South														
Airflow North														
Airflow 1.05														
HC SWT														
HC MFR														
HC RWT														
HC Energy Transfer														
HC leaving air T														
SAT AHU														
SAT South														
SAT North														
SAT 1.05														
IAT South														
IAT North														
IAT 1.05														

	Statistically significant difference in the occurrence of symptoms- can detect the fault
	No statistically significant difference in the occurrence of symptoms- cannot detect the fault
	Results contrary to the knowledge of how HVAC systems work (or) not applicable to all systems

Figure 10: Fault-symptom relationships for heating season faults.

Figure 11 shows the fault-symptom relationships for the cooling season. The chiller power meter is the only sensor common for all the three zones and it can be used to detect 10 faults. The CC RWT and SAT sensors in the North and South zones can detect 9 faults and the corresponding sensors in zone 1.05 can detect 8 faults. The energy transfer at the cooling coils of zones 1.05, South and North can be used to detect 8, 7 and 6 faults respectively.

	CCV stuck 0% flow	CCV stuck 50% flow	CCV leak 0% flow	CCV leak 40% flow	Fan failure	Fan stuck 50% airflow	Fan stuck 85% airflow	SAT sensor bias +2°C	SAT sensor bias -2°C	IAT sensor bias +2°C	IAT sensor bias -2°C	Incorrect SAT setpt +2°C	Incorrect SAT setpoint -2°C
Chiller Power													
Airflow South													
CC South SWT													
CC South MFR													
CC South RWT													
CC South Energy Transfer													
SAT South													
IAT South													
Airflow North													
CC North SWT													
CC North MFR													
CC North RWT													
CC North Energy Transfer													
SAT North													
IAT North													
Airflow 1.05													
CC 1.05 SWT													
CC 1.05 MFR													
CC 1.05 RWT													
CC 1.05 Energy Transfer													
SAT 1.05													
IAT 1.05													

	Statistically significant difference in the occurrence of symptoms- can detect the fault
	No statistically significant difference in the occurrence of symptoms- cannot detect the fault
	Results contrary to the knowledge of how HVAC systems work (or) not applicable to all systems

Figure 11: Fault-symptom relationships for cooling season faults.

2.4 Discussion

The main objectives of this research were: (i) to identify the suitable symptom detection rule when using measures of central tendency, (ii) to establish the statistical significance of the difference in symptom occurrences observed at various sensors for an AHU during faulty operation, and (iii) to identify relationships between faults and symptoms using the results of the statistical significance testing. A comparison of symptoms observed for the non-faulty (baseline) case using mean-standard deviation and median-median absolute deviation resulted in the choice of mean and three times standard deviation as the appropriate definition to detect the presence of a symptom. The paired t-test (followed by Bonferroni's correction) resulted in identification of fault-symptom relationships that can aid the selection of sensors for detecting AHU faults.

2.4.1 Reflection on results

The $\mu \pm 3\sigma$ rule was selected as the suitable symptom detection rule because it produced the fewest symptoms in the fault-free case. The M-MAD combination deemed a higher share of the fault-free operating points as symptoms due to the robustness of the median to outliers. Since statistical significance testing determines whether fault-induced differences in symptom occurrence are meaningful, future work could evaluate M–MAD within the same framework. Cousineau and Chartier [35] note the drawback of using μ to detect outliers in small sample sizes; the smallest window in this research contained 753 datapoints. At lower data-recording frequencies, these limitations may become relevant and would require a separate investigation.

The simulation recorded output values of all the defined sensors at 1-minute intervals. However, in practice, sensor data is not recorded as frequently. For example, the Priva BMS system used in The Netherlands records data at 8-minute intervals. In such a situation, the SDCD values obtained in practice would also be different. An SDCD value of 60 minutes obtained in this simulation translate to at least 7 consecutive BMS measurements which would show a symptom. In such a case, FDD algorithms must be able to detect the symptom and

diagnose the fault that caused it using these 7 measurements. While certain faults (depending on component & severity) can cause symptoms that exist for longer durations, other faults could lead to situations where a symptom can exist depending on factors such as time of day, outside air temperature etc. In such cases, the ability to swiftly detect and diagnose faults is key.

The paired t-tests used daily symptom counts as samples. The sample sizes were 119 for the heating season and 65 for the cooling season. Sample size affects statistical power- a larger sample size results in a larger statistical power and will likely lead to the rejection of a false null hypothesis [36]. A smaller sample size is sufficient to detect a large effect [33]. For a given statistical power, a larger sample size is needed to detect a small effect [33].

Sensors with high Cohen's *d* values should be prioritised for use in fault detection. Threshold values of 0.2, 0.5 and 0.8 can be used to classify effect sizes as small, medium and large respectively. However, these values have their origin in the field of behavioural sciences and are context-dependent [33]. Similarly, sensors can be characterised using effect size, SOP, and SDCD to form fault-specific symptom signatures. For example, the IAT sensor bias +2 °C results in an SOP of 13.1% and a mean SDCD of 7 minutes with a Cohen's *d* value of 1.02 at the HC RWT sensor. In contrast, the Fan failure fault results in an SOP of 7.1% and a mean SDCD of 38 minutes with a Cohen's *d* value of 0.47 at the HC MFR sensor. Since the paired t-test is performed for each sensor, a single scale of Cohen's *d* values mapped to SOP and SDCD values is not applicable for all sensors. This results in the use of different reference values of SOP, SDCD & Cohen's *d* when selecting sensors for fault detection.

In this simulation study, several sensors were defined in the EnergyPlus environment and checked for their ability to detect the presence of faults. In practice, not all of the studied sensors are installed. For example, the airflow meters and mass flow rate measurements of hot and chilled water across the heating and cooling coils are not installed. Instead, the fan speed control signal or supply channel pressure sensor and HCV/CCV opening signals available from the BMS can be used respectively. For the coils, a virtual mass flow rate meter and consequently, a virtual energy flow rate meter can be developed based on available sensors in buildings. Based on the obtained results, the RWT sensors can detect several faults in the cooling season. It is worthwhile to install them for the cooling coils. The HRW supply side air exit T sensor also detects several faults. It is recommended to install this sensor too. While the HRW return side air entry T sensor also detects several faults, this sensor measures nothing but the air returning from the indoor zones. Therefore, the IAT sensors must be sufficient to detect the faults that it detects. The obtained results are based on the HVAC system layout, the zoning and the hydronic distribution of this specific case study.

The case specificity of the results must be considered. The study in this research focused on AHUs that condition air for indoor spaces. The study of the IAT sensor bias fault and the use of IAT sensor to detect faults are a consequent of using this AHU. For example, there exist make-up AHUs that precondition outside air and provide it to other AHUs. Such make-up AHUs are often controlled by SAT setpoints. Similarly, the supply temperature is often defined by the central heating circuit and/or the boiler or heat pump. The consideration of specific pump energy meters in the AHU power meters are also case-specific. In reality, the energy meters of pumps would not be available separately.

It must be noted that there are two software-specific factors that could have also influenced the results obtained in this research: the control strategy employed by EnergyPlus and the modelling of the faults in the software. EnergyPlus uses a 'predictive system energy balance' method where the amount of energy needed to be provided by the HVAC system to maintain the zone temperature is calculated and then the zone temperature value is updated based on the actual energy provided [37]. In the real world, other control strategies are being used e.g. PID control. There has been an increasing amount of research on Model Predictive Control (MPC) strategies. With different control strategies, the response of the HVAC system to the implemented faults would be different which may in turn lead to varying patterns of symptoms. While the implementation of the faults pertaining to valves, fan, HRW & incorrect SAT setpoint were done through Energy Management System (EMS), Heat Recovery object and setpoint manager, the sensor biases made use of the fault objects available in EnergyPlus. The results indicated that the IAT sensor bias fault did not lead to any statistically significant difference in the occurrence of symptoms in the HVAC system during the cooling season. In Gopalan et al. [27], it was observed that the effect of this fault on the chiller power was very minimal.

In Gopalan et al. [27], a calculation of the deviation in energy transferred at the heating and cooling coils for the simulated day, in response to the implemented faults, were presented. A comparison of the relative ranking of energy exchanged at the coils was performed against the relative ranking of the Cohen's *d* value obtained for HC Energy Transfer symptom. For the heating season, the deviation in the energy transferred at the heating coil was highest for fan failure and HCV stuck 0% flow fault whereas the results of Cohen's *d* showed that HRW failure led to the highest value. Also, the HCV leak 40% did not show any deviation in the energy transferred at the heating coil whereas, in this research, the fault led to a statistically significant change in the occurrence of symptoms and a Cohen's *d* value of 0.66. When the sensor bias and incorrect SAT setpoint faults were considered, the relative comparison of Cohen's *d* values for the SAT sensor bias and incorrect SAT setpoint faults matches the energy deviation at the heating coil presented in [27]. However, the IAT sensor bias faults had a lower energy deviation than the SAT sensor biases and incorrect SAT setpoint fault -2 °C whereas, the IAT sensor bias gave a high Cohen's *d* values. For the cooling season, the trends of the Cohen's *d* values matches the trends of the energy transfer at the coils for the cooling day simulation in Gopalan et al. [27]. It must be noted that the energy transfer at the coils were analysed only for one heating day and one cooling day in Gopalan et al. [27] whereas the occurrences of symptoms were calculated for the full duration of the heating and cooling seasons and the paired t-test was performed this data of symptoms. Additionally, the underlying data being compared are different- energy transferred across the coils and their deviations under faults against the distribution of number of daily symptom occurrences over entire seasons.

With respect to the fault-symptom relationships, the SAT AHU sensor is the target variable that the heating coil controller aims to meet during the heating season. The SAT sensors installed in the downstream (distribution side) is purely for measurement during the heating season. The generalisability of fault-symptom relationship results obtained for the IAT sensors during both heating and cooling seasons (and the individual SAT sensors during the cooling season) depend on the zoning of the case study building, air leakage, thermal properties of facade, surface area and thermal properties of the windows, orientation towards sunlight etc.

2.4.2 Comparison of results with literature

In this section, the results obtained in this research are compared with the results of similar studies in the literature. The faults common to the work of Y. Chen et al. [17] and this research are the valve stuck fault, the valve leak fault and IAT sensor bias fault (each applicable for both the heating season and the cooling season). The symptoms common to Y. Chen et al. (Chen et al., 2022) and this research are the IAT sensor, SAT sensor, CCV MFR, CCV RWT, HC MFR, HC RWT and air flow rate. It must be noted that for Y. Chen et al. (Chen et al., 2022) all the six faults were studied for all six symptoms for one year of simulation whereas in this research, HC MFR & HC RWT were studied only for the heating season faults and CC MFR & CC RWT were studied only for the cooling season faults. It was checked if there are overlapping ranges of SOP values for the compared faults at the common symptoms. The results of the valve stuck faults had overlapping ranges for the corresponding four symptoms. The valve leak faults led to overlapping ranges of SOP values only at the IAT sensor. IAT sensor biases during the heating season had one overlapping SOP range each- HC MFR for the positive bias and IAT sensor for the negative bias. IAT sensor biases (both + & -) during cooling season did not have overlapping SOP ranges. It is important to note that there are some additional differences in the methodologies of Y. Chen et al. (Chen et al., 2022) and this research such as HVAC system layout, hydronic typology and control strategy employed, differences in fault severities and the distinction made between positive and negative symptoms in Y. Chen et al. (Chen et al., 2022).

The fault-symptom relationships identified in this research was compared with research from the literature such as Y. Zhao et al. [13], Y. Zhao et al. [38], Y. Chen et al. [39], T. Li et al. [40] and Dey and Dong [41]. Across these researches, the study of valve stuck fault, valve leak fault, fan faults, and SAT sensor faults matched with the faults presented in this report. The research cited above focused on RAT sensor faults whereas this research studied IAT sensor bias. The RAT sensor faults results were taken as synonymous with the IAT sensor results and also compared. The use of RAT symptoms in the literature was compared with whether symptoms could be observed at IAT in this research. The results for sensors which are unique to AHUs recirculating return air such as Mixed Air Temperature (MAT) sensor, Outdoor Air Flow Ratio (OAFR) could not be compared. Addi-

tionally, in this research, the use of HCV during cooling season and CCV during heating season to detect symptoms were not performed. Therefore, in cases where such symptoms were used in the literature, a comparison was not performed. Taking into account all these considerations, sensors such as SAT sensor, SWT sensor, HCV opening (or MFR), CCV opening (or MFR), energy transfer at the coils, and IAT sensor were used for detecting the aforementioned faults. The literature used fan speed and power (both at supply and return fans) for fan faults. Fan power was included in the AHU power symptom presented in this study. As mentioned earlier in this section, the fan speed values from BMS can be used instead of air flow meters. Thus, results obtained in this research agrees with the symptoms used in the literature to diagnose faults.

2.4.3 Applications

The applications of the results of this research are as follows:

i. As explained in Y. Chen et al. [21], the SOP values can be used as conditional probabilities when constructing Diagnostic Bayesian Networks (DBNs). It must be noted that the control strategies used in simulation software can also impact the occurrences of symptoms and consequently, the SOP values.

ii. The fault-symptom relationships can provide information about how many sensors can detect a fault. The choice of sensors to detect the fault can be based on the Cohen's *d* value, SOP & SDCD values. These results can be used to identify the optimal combination of sensors required to detect a set of prioritised faults.

iii. Poor accuracy of sensors due to initial cost considerations is another factor challenging FDD in buildings [9]. The research presented in Chapter 3 of this report compares the performances of 16 Low-Cost Sensors (LCSs) and Low-Cost Monitors (LCMs) with Research-Grade Instruments (RGIs) to identify the best performing sensors and monitors. These LCSs and LCMs can in turn be used to install missing sensors in HVAC systems for FDD. Combining the results of the LCS and LCM selection with the fault-symptom relationships obtained in this research can help building owners and managers in ensuring that their buildings are FDD-ready at a comparatively lower cost.

2.4.4 Limitations and future research

The recommendations for future research include:

i. This work lacks a verification of the obtained results through experimental means or otherwise. It is proposed to validate the results of this research as follows:
   (a) fault-symptom relationships obtained in this research can be validated through interviews with industrial experts, practioners and maintenance engineers, technicians based on their field experience. Fault experiments can be performed in buildings to check if the symptoms occur as per the relationships obtained in this map.
   (b) The SOP and SDCD values can be validated through an experimental case study. These values can also be obtained and validated through interviews with industrial experts.

ii. The obtained SOP results are for a specific HVAC control strategy of EnergyPlus and needs to be modified before applied to HVAC systems of real buildings. Also, the study was performed only for heating and cooling seasons separately. The shoulder season-faults were not studied. A study must be performed to compare the results obtained in this study with the results obtained for the same case study simulated using a PID-based feedback control. Also, the proposed study must include faults simulated during the shoulder season.

iii. The choice of window bins in WPM (section 2.2.3) could affect the obtained results. Similarly, binning of windows could be different for each of the studied sensor in the HVAC system. The sensitivity of the obtained SOP and SDCD results to the choice of window size and number of windows must be performed. Additionally, a comparison of the SOP and SDCD results obtained through this method against methods that use machine learning model-based time series forecasting for baseline prediction and consequently symptom detection will help in benchmarking the performance of this WPM-based symptom detection method that uses rules based on mean and standard deviation.

2.5 Conclusions

A simulation study was performed to identify the symptom effects of AHU faults across various sensors installed in the HVAC system. A heating season model and cooling season model of a case study were developed and simulated under baseline and faulty conditions. KPIs such as SOP & SDCD were calculated for each of the 33 analysed sensors. The paired t-test was performed to determine which sensors showed a statistically significant difference in the occurrence of symptoms for various faults in the heating and cooling seasons thereby leading to the identification of fault-symptom relationships.

The novel aspects of this research are:

i. The calculation of symptom occurrences in the fault-free case to identify the suitable rule for symptom detection using descriptive statistics, and
ii. The application of statistical significance tests to establish the effect of faults on symptom occurrences.

The main contributions of this research are:

i. An empirical proof that the combination of μ and 3σ is the suitable rule for symptom detection when using descriptive statistics,
ii. Fault-symptom relationships obtained from statistical significance testing which can be used to select sensors fault detection in combination with Cohen's *d*, SOP and SDCD values.

The main conclusions are as follows:

i. The combination of μ and 3σ should be used to detect symptoms when using a WPM-based symptom detection approach using descriptive statistics. This is because this combination gives the least amount of symptoms (outliers) in the fault-free case, which is the expectation in practice.
ii. Excluding the HCV leak 0% fault, for the heating season, out of 260 patterns of symptom occurrences caused by 13 faults across 20 sensors, 134 were deemed as statistically significant when comparing the difference of symptom occurrences between the faulty and fault-free cases for the heating season. A further 15 patterns of symptom occurrences, across three different faults, were also deemed statistically significant even though they were either contrary to the knowledge of how HVAC systems work or lack generalisability. Excluding the CCV leak 0% fault, for the cooling season, out of 264 patterns of symptom occurrences caused by 12 faults across 22 sensors, 139 were deemed as statistically significant. A further 3 patterns of symptom occurrences were also deemed statistically significant even though they were contrary to the knowledge of how HVAC systems work.
iii. The results of the statistical significance tests were used to determine fault-symptom relationships which in turn could be used to decide on the necessary sensors required to detect AHU faults. The IAT North sensor can detect 12 heating season faults. Chiller power meter can detect 10 cooling season faults. Fan failure fault leads to detectable symptoms at the highest number of sensors during both heating and cooling season.

3 EVALUATION OF THE PERFORMANCE OF LOW-COST MONITORS FOR THEIR USE IN FAULT DETECTION AND DIAGNOSIS

The results presented in this chapter were published as Gopalan, S., Zheng, H., Walker, S. S. W., Kramer, R. P., & Zeiler, W. (2023). Evaluation of the performance of low-cost monitors for their use in fault detection and diagnosis. In M. Schweiker, C. Van Treeck, D. Mueller, J. Fels, T. Kraus, & H. Pallubinsky (editors), Proceedings of Healthy Buildings 2023 Europe: Beyond Disciplinary Boundaries Artikel 1149 ISIAQ International.

3.1 Introduction

This chapter presents the results of the evaluation of Low-Cost Sensors (LCSs) and Low-Cost Monitors (LCMs) with a strong focus on fit-for-purpose for FDD application. A weighted rating criterion using the 0 - 3 rating scale of the Smart Readiness Indicator (SRI) framework was used. The weights were based on the relative importance of the studied parameters for FDD. The scope of this study was limited to the parameters of Temperature (T), Carbon dioxide (CO_2) & Relative Humidity (RH) because of their wide usage for FDD.

3.2 Methods

The method of this study is similar to the method of H. Zheng et al. [20] which in turn followed the method of I. Demanega et al. [19]. The experiments were carried out in a controlled climate chamber of capacity 52.5 m^3 situated in the Building Physics and Services laboratory at Eindhoven University of Technology, whose design has been described in H. Schellen et al. [42]. A view of the climate chamber is shown in Figure 12.

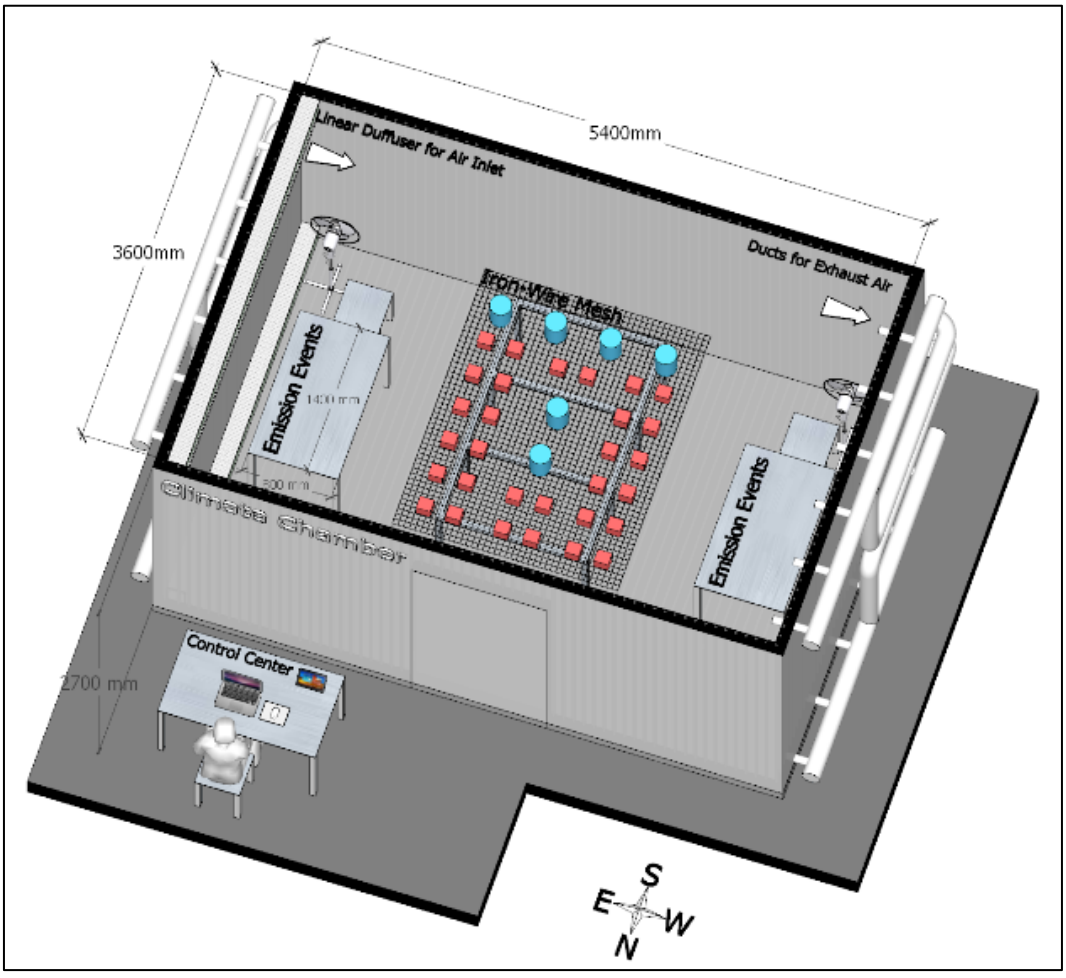

Figure 12: A view of the climate chamber used for the evaluation of sensors [8].

The indoor climate conditions were regulated by a HVAC system using LabVIEW, BPSCC, Version 1.0. Outdoor air is filtered twice (coarse air filter in the first stage and HEPA box air filter in the second stage) before being supplied to the chamber containing a wire-mesh table in its centre, two tables (one on each side of the wire-mesh table) where the events took place and four stools for the occupants to be seated. All the LCMs & RGIs were placed on the wire-mesh table and kept at a distance of at least 1 m from the emission sources. For each LCM, at least two units (devices) were placed at multiple locations within the chamber to measure the parameters. The study employed 16 LCMs, which were available in the Dutch market at the time of the study, and 2 RGIs. The accuracies of the sensors measuring the different parameters are listed in Table 7. The specifications of the RGIs are provided in Table 8.

Table 7: List of LCMs used in the experiment and the accuracies of the sensors measuring the individual parameters.

LCM Name	Accuracy of T sensor	Accuracy of CO_2 sensor	Accuracy of RH sensor
Foobot SAT	±1°C	±(50ppm + 5% of reading)	±5% in the range of 15-45°C
Atmotube PRO	±0.5°C	-	±3% RH
NuWave WB55 AR10-B-112-A	±1 °C (max) (at –10 to 85 °C)	±3% of reading	± 5% RH (max) (at 0-90% RH)
Cloudgarden	±0.3°C	±(50ppm + 3% of reading)	±2%
Edimax AI-2003W	±1°C	±30ppm	±5%
ClimateBuddy	±0.3°C (-40 to 90°C); ±0.6°C (90 to 125°C)	±(30ppm + 3% of reading)	±2% at 0-80%; ±3% at 80-100%
Netatmo (in-door unit)	±0.3°C	±50 ppm from 0 to 1000 ppm; ±5% from 1000 to 5000 ppm	±3%
Netatmo (out-door unit)	±0.3°C	-	±3%
Atal AirZenZ Pro	±0.5°C	±50 ppm ±3% of measured value	±3% at 20-80%
Elsys ERS CO2	±0.2°C	±50 ppm, or ±3% of reading (0-2000 ppm); ±10% of reading (0-10,000 ppm)	±2%
Aranet4, SenseAir	±0.4°C	±50 ppm or ±3% of reading (0-2000 ppm); ±10% of reading (2001-9999 ppm)	±3%
AirTeq	±0.3°C	±40ppm + 2% of reading at 22°C	±2%
BRControls	±0.3°C	±40 ppm + 3 % of measured value	±3%
Sensirion Combined	±0.8°C at 15 to 35 °C; ±1.5°C at -10 to 60 °C	±(40 ppm +5% of reading) at 400-5000 ppm	±6% at 15 to 35°C, 20-65% RH; ±9% at -10 to 60°C, 0-100% RH
Chess Normal	±(0.4°C+0.023x(T[°C]-25°C)) at 0-50°C	±(30 ppm + 3% of reading) at 400-10,000 ppm	±3%
Chess Infra	±0.2°C	-	2%
Foobot SAT	±1°C	±(50ppm + 5% of reading)	±5% in the range of 15-45°C

Table 8: Specifications of RGIs used in the experiments.

RGI Name	Parameter	Range	Accuracy
SBA 1-second regular interval	CO_2	0-5,000 ppm	<1% of range
EE08 series transmitters (with a data logger Squirrel 2040) 1-minute regular interval	T	-55°C - 80°C	±0.05°C from 0-50°C
	RH	0-100%	±2% between 0-90%;

Both the RGIs and LCMs were deployed in the chamber to measure the changes in the values of T, RH & CO_2 during the experiments. Table 9 lists the 14 events that were performed as a part of the experiment to evaluate the performance of the LCMs. It must be noted that certain lower T & RH values could not be achieved in the chamber using LabVIEW. However, since the objective was to track changes in T, RH & CO_2 values over time the essence of the experiments was still captured in the data recorded by the RGIs & LCMs.

Following the experiments, the data collected from the LCMs and RGIs were analysed using Python (Spyder Software). The data measured by the LCM units were resampled for a 5-min data interval before proceeding with the analysis. To analyse the performance of the LCMs, the following steps were followed:

(i) Time-series graphs, that showed the variation of the chosen parameter for each recorded event, were plotted to check for the detection of the trend by the LCMs and its conformity to the trend shown by RGIs.

(ii) Metrics, as suggested by the US Environmental Protection Agency (EPA) to evaluate the performances of the LCMs [43], [44], were calculated for every sensor for each event separately. The metrics calculated are introduced and described below in Table 10. The average of the measured values of the deployed LCM units were used to plot the time series graphs and to calculate R^2. For the remaining metrics, the values measured by the individual LCM units were directly used. The formulae used to calculate the rest of the metrics are described in Equations 8-11.

$$RMSE = \sqrt{\frac{1}{(N\times M)}\sum_{j=1}^{M}\left[\sum_{t=1}^{N}(x_{tj}-R_t)^2\right]} \quad \text{Eq. (8)}$$

$$NRMSE = \frac{RMSE}{\overline{R_t}}\times 100 \quad \text{Eq. (9)}$$

$$CV = \frac{\sqrt{\frac{1}{(N\times M)-1}\sum_{j=1}^{M}\left[\sum_{t=1}^{N}(x_{tj}-\bar{x}_t)^2\right]}}{\bar{x}}\times 100 \quad \text{Eq. (10)}$$

$$MNB = \frac{1}{(N\times M)}\sum_{j=1}^{M}\left[\sum_{t=1}^{N}\left(\frac{x_{tj}-R_t}{R_t}\right)\right]\times 100 \quad \text{Eq. (11)}$$

where:

N is the number of time instances for which data is recorded

M is the number of identical sensor units used

x_{tj} is the average LCM measurement for unit j & time instance t

R_t is the average RGI measurement for time instance t

$\overline{R_t}$ is the averaged value of RGI measurements for time instance t

$\bar{x}_t$ is the average value of all LCM unit-measurements for time instance t

$\bar{x}$ is the averaged value of all LCM units measured over the test period

Table 9: List of events in the experiment to evaluate the performance of LCMs.

Parameter	Event Description	Event Timespan	Chamber T	Chamber RH
T	T increased in the chamber at a constant RH	2h 00m	From 10˚C to 30˚C	40%
		1h 10m	From 14˚C to 20˚C	70%
		0h 50m	From 20˚C to 26˚C	40%
	T decreased in the chamber at a constant RH	3h 00m	From 30˚C to 14˚C	80%
CO_2	CO_2 Administering- CO_2 released into the chamber until a level of 4,500 ppm, followed by a period of decay	1h 30m	26˚C	70%
		1h 30m	20˚C	40%
	Human Occupancy- Volunteers present in the chamber perform activities & generate CO_2	6h 00m	20˚C	40%
		1h 45m	26˚C	70%
		1h 30m	26˚C	70%
RH	RH increased in the chamber at a constant T	1h 30m	30˚C	From 40% to 80%
		1h 00m	14˚C	
		1h 00m	26˚C	
	RH decreased in the chamber at a constant T	3h 30m	14˚C	From 75% to 40%
		1h 30m	20˚C	From 80% to 40%

Table 10: Evaluation Metrics as described in Duvall et al. [43] and Williams et al. [44].

Aspect	Definition	Metric
Linearity	The extent to which LCM measurements can explain the measurements of the RGI.	R^2 (Coefficient of Determination)
Error	The disagreement between the measurements of the LCMs & RGIs.	RMSE: Root Mean Square Error
		NRMSE: Normalised Root Mean Square Error
Precision	Variations among the concurrent measurements of LCMs	CV: Coefficient of Variation
Bias	Systematic distortion of measurements of LCMs when compared against the RGIs	MNB: Mean Normalised Bias

(iii) Each metric is averaged out over all the events conducted per IEQ parameter to obtain an average metric value per LCM per parameter.

(iv) The performance of each LCM in all of the evaluated metrics is rated as per Table 11. These rating criteria use the 0-3 scale used by the Smart Readiness Indicator (SRI) framework to evaluate the functionality level of various services offered by a building while assessing its SRI. Subsequently, an average rating for each LCM is obtained by averaging out the ratings provided for each metric.

(v) In the final step, a weighted average rating is obtained for each LCM by applying the weights of 0.50, 0.33 & 0.17 to their respective T, CO_2 and RH ratings. The weights for each parameter have been provided based on their importance to FDD in commercial buildings: (a) since Temperature sensors are the most common in buildings and widely used to determine the presence of a fault, Temperature gets the highest weight; (b) CO_2 is an important parameter that can be used as an indicator to detect faults with ventilation systems and therefore, it gets the second highest weight and (c) Relative Humidity, as a parameter, may not be as important an input for FDD in commercial buildings; hence it gets the lowest weight. After a weighted average rating has been obtained for each LCM, the best performing LCMs are recommended for use in FDD applications in commercial buildings.

Table 11: Rating Criteria for LCM performance in metrics.

Parameter	Metric	3 (Very good)	2 (Good)	1 (Moderate)	0 (Poor)
T	CV (%)	CV ≤10%	10% <CV ≤20%	20% <CV ≤30%	CV > 30%
	R^2 (-)	$R^2 \geq 0.8$	0.8> $R^2 \geq 0.6$	0.6 > $R^2 \geq 0.4$	$R^2 < 0.4$
	RMSE (°C)	RMSE ≤ 0.5°C	0.5°C <RMSE ≤1°C	1°C <RMSE ≤1.5°C	RMSE > 1.5°C
	NRMSE (%)	NRMSE≤±10%	10%<NRMSE≤20%	20% < NRMSE ≤ 30%	NRMSE > 30%
	MNB (%)	MNB≤±10%	10%<MNB≤20%	20%<MNB≤30%	MNB>30%
CO_2	CV (%)	CV≤10%	10%<CV≤20%	20%<CV≤30%	CV>30%
	R^2 (-)	$R^2 \geq 0.8$	0.8> $R^2 \geq 0.6$	0.6> $R^2 \geq 0.4$	$R^2 < 0.4$
	RMSE (ppm)	RMSE ≤ 100ppm	100ppm <RMSE ≤200ppm	200ppm<RMSE≤ 300 ppm	RMSE>300ppm
	NRMSE (%)	NRMSE≤±10%	10%<NRMSE≤20%	20%<NRMSE≤30%	NRMSE>30%
	MNB (%)	MNB≤±10%	10%<MNB≤20%	20%<MNB≤30%	MNB>30%
RH	CV (%)	CV≤10%	10%<CV≤20%	20%<CV≤30%	CV>30%
	R^2 (-)	$R^2 \geq 0.8$	0.8> $R^2 \geq 0.6$	0.6> $R^2 \geq 0.4$	$R^2 < 0.4$
	RMSE (%)	RMSE ≤ 2%	2%<RMSE≤4%	4%<RMSE≤6%	RMSE > 6%
	NRMSE (%)	NRMSE≤±10%	10%<NRMSE≤20%	20%<NRMSE≤30%	NRMSE>30%
	MNB (%)	MNB≤±10%	10%<MNB≤20%	20%<MNB≤30%	MNB>30%

3.3 Results

In this section, the results of the analysis are presented. Figures 13, 14, and 15 show the time series plots of all the LCMs and the corresponding reference sensors for one sample event corresponding to each of the three parameters studied: T, CO_2 & RH. It can be observed that the measurements of the LCMs followed the trend of the measurement by the RGIs.

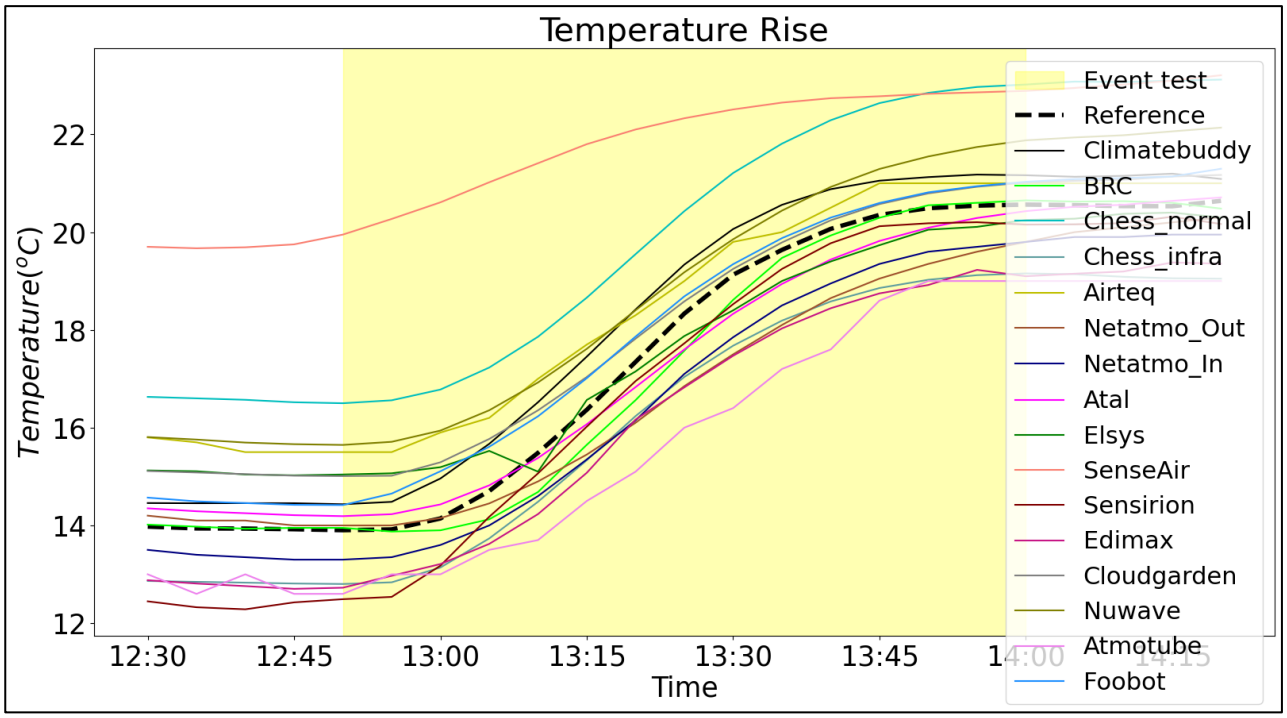

Figure 13: Time series plot: T rise from 14 °C to 20 °C at 70% RH.

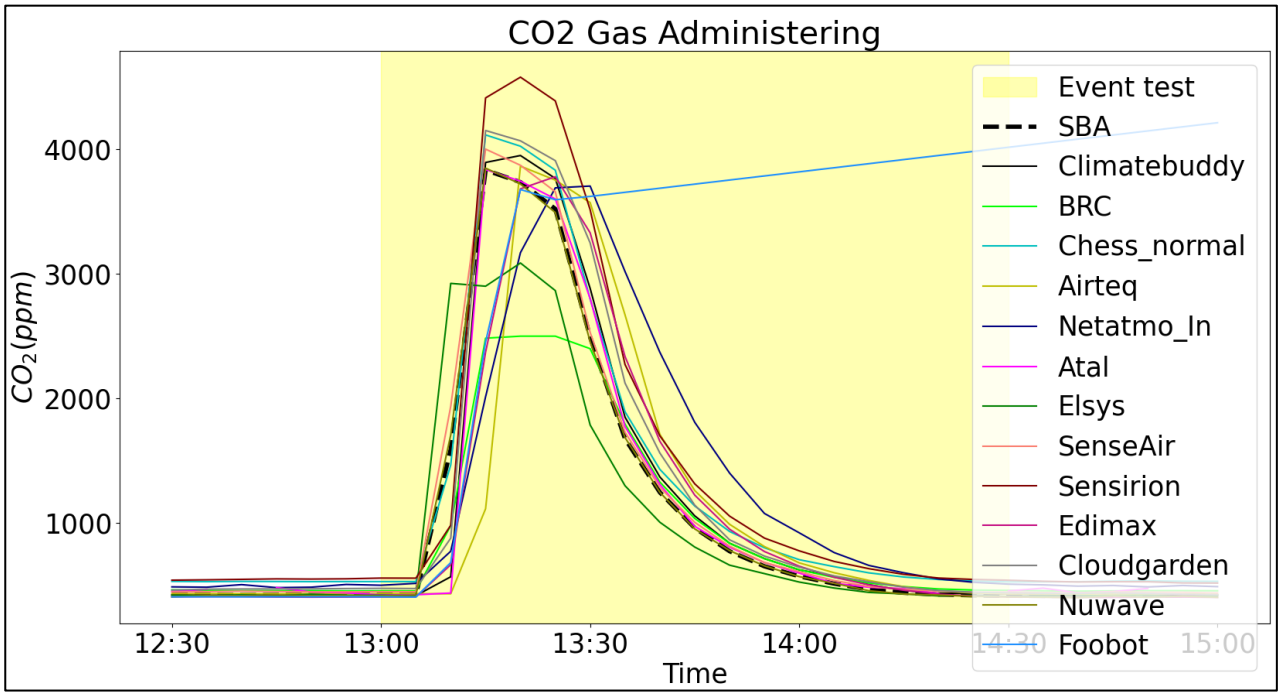

Figure 14: Time series plot: CO_2 Administering at 26 °C & 70% RH.

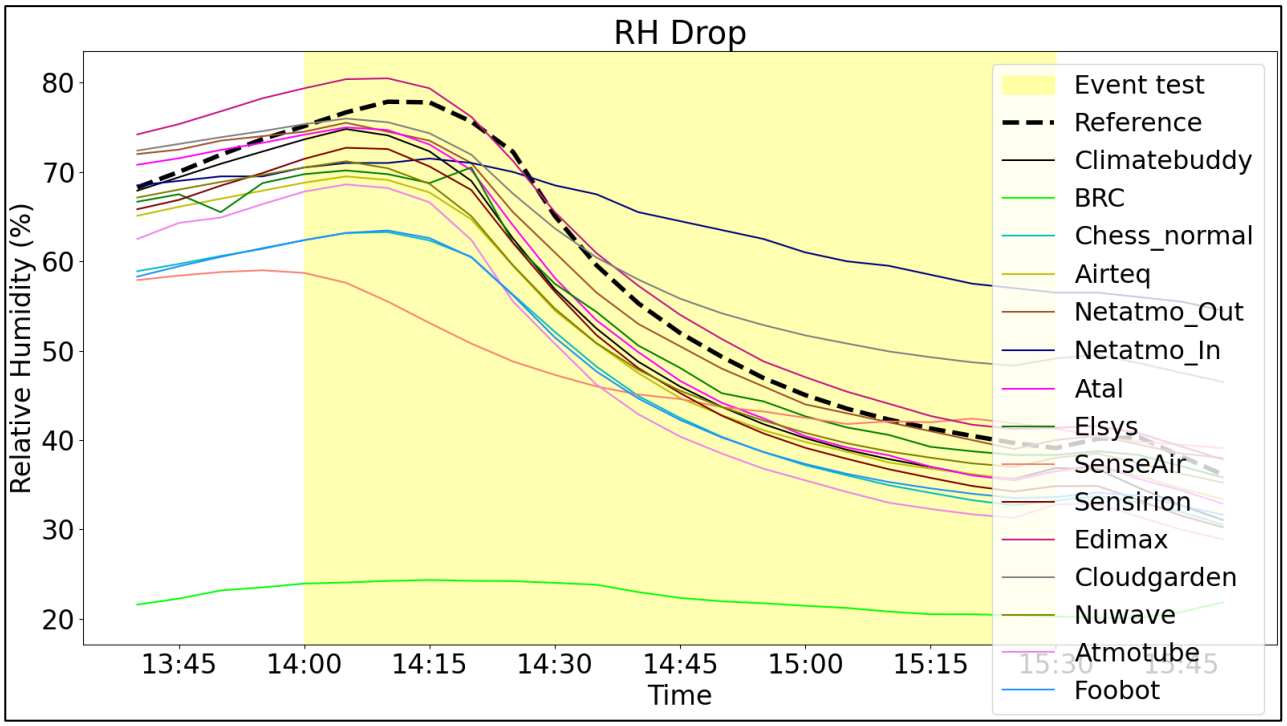

Figure 15: Time series plot: RH drop from 75% to 40% at 14 °C.

Subsequently, the five evaluation metrics were calculated for each event for each LCM separately. The values of each metric over multiple events were averaged out and are presented in Tables 12, 13, and 14 for each of T, CO_2, and RH.

Table 12: Calculated Average Metric Values for T.

LCM	CV (%)	R^2	NRMSE (%)	RMSE (°C)	MNB (%)
Airteq	1.07	1.00	4.82	0.97	3.83
Atal	2.37	0.99	6.67	1.31	1.11
Atmotube	1.46	0.99	7.65	1.58	-7.18
BRC	0.77	0.99	2.81	0.60	-1.64
Chess_infra	1.55	1.00	9.16	1.92	-8.95
Chess_normal	0.15	1.00	11.41	2.33	11.94
Cloudgarden	0.08	1.00	3.53	0.71	2.57
Edimax	0.65	1.00	6.88	1.43	-6.57
Elsys	3.85	0.99	6.63	1.39	-0.66
Foobot	0.38	1.00	4.00	0.83	4.03
Netatmo_In	0.55	1.00	4.22	0.88	-3.83
Netatmo_Out	0.44	0.99	5.86	1.21	-2.12
Nuwave	0.90	0.99	6.85	1.39	6.89
Sensirion	6.74	0.98	8.31	1.73	-3.58
Airteq	1.07	1.00	4.82	0.97	3.83

Table 13: Calculated Average Metric Values for CO_2.

LCM	CV (%)	R^2	NRMSE (%)	RMSE (ppm)	MNB (%)
BRC	68.43	0.96	77.07	955	2.63
Chess_normal	3.76	0.99	17.71	180	19.69
Cloudgarden	11.32	0.75	25.67	262	1.26
Edimax	2.35	0.88	19.95	267	4.02
Netatmo_In	7.26	0.82	32.91	415	23.34
Nuwave	1.47	1.00	2.84	33	-1.00
Sensirion	3.57	0.98	22.51	278	18.01

Table 14: Calculated Average Metric Values for RH.

LCM	CV (%)	R^2	NRMSE (%)	RMSE (%)	MNB (%)
Airteq	0.58	0.97	12.55	7.16	-4.96
Atal		0.98			-7.55
Atmotube	1.98	0.98	15.84	8.99	-15.57
BRC	0.72	0.91	61.90	35.76	-54.62
Chess_normal	0.23	0.99	19.96	11.43	-18.58
Cloudgarden	1.16	0.98	11.80	6.70	10.09
Edimax	0.70	0.99	3.59	2.05	1.79
Elsys	5.20	0.98	11.57	6.67	-6.93
Foobot	0.84	0.99	18.78	10.76	-16.93
Netatmo_In	0.60	0.95	19.98	11.37	11.40
Netatmo_Out	1.52	0.98	7.89	4.55	-3.92
Nuwave	1.34	0.97	10.69	6.11	-7.06
Sensirion	8.01	0.98	10.80	6.16	-4.78

All LCMs were rated on the metrics for each of T, CO_2, & RH separately as per the rating criteria in Table 11. Finally, a weighted average rating was obtained for all LCMs by assigning weights of 0.50, 0.33 & 0.17 to T, CO_2 & RH respectively. The weighted average rating results are shown in Figure 16. This has been calculated only for those LCMs which logged data for all three parameters during the experiments.

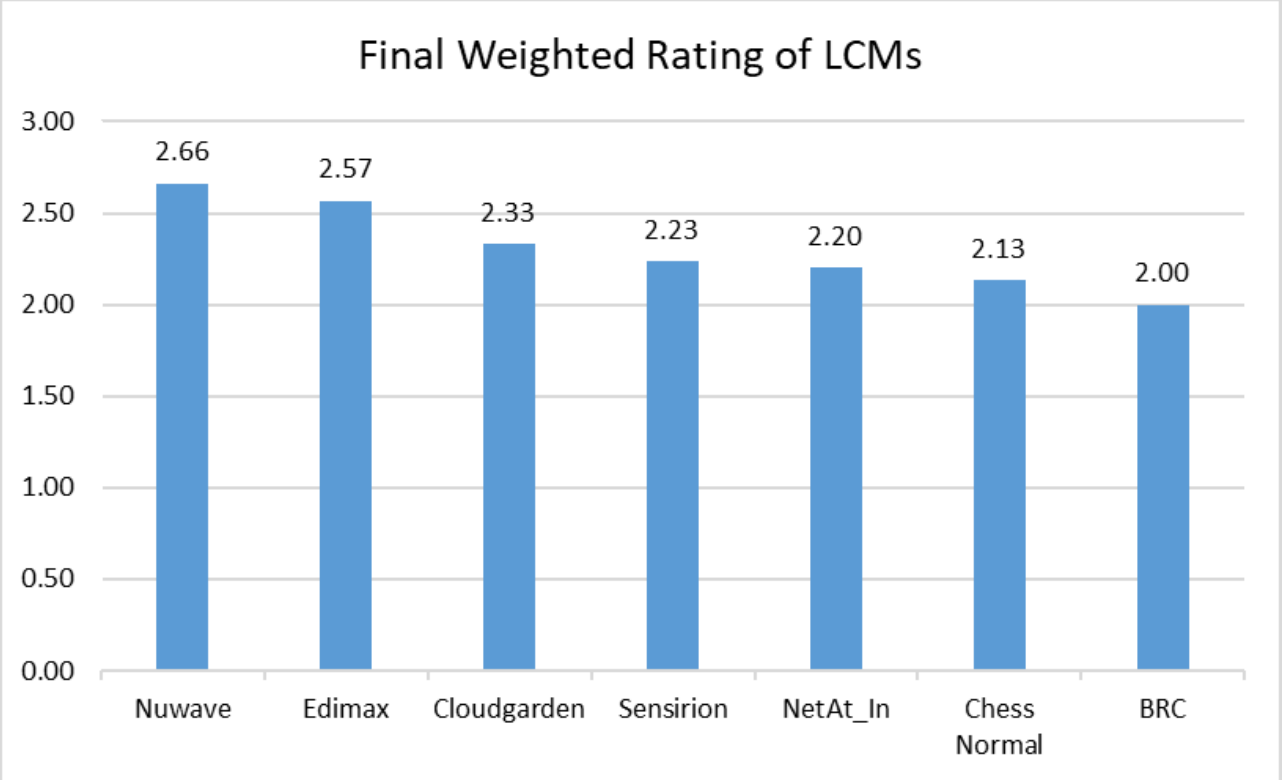

Figure 16: Weighted Average Rating of LCMs on a 0-3 scale.

3.4 Discussion

The guidelines provided in Duvall et al. [43] and Williams et al. [44] set a score of ±30% as acceptable for the metrics CV, NRMSE & MNB and a value of 0.7 or more as acceptable for R^2. Using these benchmark values for the parameter Temperature, the CV, R^2 and NRMSE values of all LCMs are within the acceptable limit. While the MNB values individually are well within the limit, the difference between the two most extreme values of MNB is close to 20.9%. The RMSE values for all LCMs are above 0.5 °C and therefore the ratings corresponding to RMSE are 2 or lower. The difference between the monitors with the two most extreme values is 1.73 °C which is quite significant. Only in the case of the monitor Foobot did the RMSE value meet the specified accuracy value. For the parameter, CO_2, the R^2 values of all monitors, the CV and NRMSE values of all monitors except BRC are within the acceptable limit. The RMSE values of all monitors except Nuwave are greater than

200 ppm, and show a high degree of disagreement amongst one another. The RMSE value of Nuwave meets the accuracy specification in certain concentrations of CO_2. For the parameter RH, the CV an R^2 values of all monitors fall within the acceptable limit. The NRMSE values of all monitors except BRC fall within the limit. The difference between the monitors with the two most extreme values of MNB is 66%.

Across all three parameters, the recurring trend is that the monitors do not satisfy the benchmark value of at least one metric. This reinforces the significance of using this rating method in identifying the best performing monitors. Such differences can be captured in the rating method and will be reflected in the final score which in turn will impact the choice of an appropriate LCM. The final, weighted average rating indicated that the LCMs Nuwave, Edimax and Cloudgarden performed the best across parameters and metrics. Their costs are €420, €387 and €400 respectively. This price range is significantly lower when compared against the cost of the research grade instruments (€2,500 for SBA sensor and €3,200 for the EE-08 transmitter with Squirrel 2040 data logger), and LCMs meet the acceptability criteria of the metrics they were evaluated on.

The advantage of employing Low Cost Monitors is that they can measure more than one parameter which provides the opportunity to use the reading of more than one parameter to detect and corroborate the presence of faults. The use of a rating criteria in line with the SRI framework will allow for integration of sensor availability in buildings and the evaluation of their performance into the SRI framework. A rating method similar to the one presented in this paper can be used for evaluating and selecting LCSs and LCMs for different applications by applying different weights for the parameters measured.

While the results have been encouraging and advocate for the use of some of the tested LCMs for FDD, it is important to note that the values measured by a sensor when deployed as a standalone sensor could differ from the values it measures when integrated into an LCM. A study that compares the differences between a single sensor and LCMs for each of the three parameters must be performed and in places where only one parameter needs to be measured, a single sensor can be used.

3.5 Conclusions

A controlled climate chamber test was performed to evaluate the performance of various LCMs by comparing them against RGIs. For the valuation, five metrics were calculated as per the method advocated by EPA: R^2, RMSE, NRMSE, CV & MNB. A weighted rating was obtained for each LCM depending on the value of the calculated metrics and the importance of different parameters for FDD. It was concluded that Nuwave, Edimax and Cloudgarden LCMs performed the best and can be chosen for use in FDD. The results of this study will be useful for building owners and operators who want to install several sensors to develop advanced diagnostic tools to eliminate energy wastage and ensure thermal comfort of occupants.

4 EVALUATION OF THE PERFORMANCE OF MACHINE LEARNING ALGORITHMS FOR TIME SERIES SYMPTOM DETECTION IN AIR HANDLING UNITS: A CASE STUDY

The results presented in this chapter were published as Gopalan, S., & Kramer, R. P. (2025). Evaluation of the Performance of Machine Learning Algorithms for Time Series Symptom Detection in Air Handling Units: A Case Study. In C. Zilio, F. Busato, L. Mazzarella, & M. Noro (editors), Proceedings of the 15th REHVA HVAC World Congress - CLIMA 2025 (Vol. 1, blz. 811-822). (Lecture Notes in Civil Engineering (LNCE); Vol. 762). Springer. https://doi.org/10.1007/978-3-032-06806-4_78

4.1 Introduction

The amount of research on the topic of Fault Detection and Diagnosis (FDD) has been growing over the last two decades [4] with a greater focus on data-driven methods. Data-driven methods can be classified into supervised learning-based methods (regression and classification) and unsupervised learning-based methods (which includes clustering and principal component analysis) [9]. Data-driven methods can potentially achieve high accuracy at low cost [45]. Supervised ML-based fault detection can be categorized into classification-based and regression-based methods. The review paper of Z. Chen et al. [10] lists examples of both categories. A classification-based fault detection reduces the process to a simple categorization of the datapoint under consideration into a '0' or '1'. For example, an offset in the sensor measuring indoor air temperature is also a fault ('1') and the fan being stuck is also a fault ('1'). It trivializes the very process of fault detection to just two categories without providing any information on the severity of the fault and its location in the HVAC system. In contrast, symptom detection is a process that tries to identify the effects of faults in sensor measurements. As explained in the Chapter 1, it is quite similar to how a doctor performs their role: they use the information about symptoms caused by the disease to diagnose the actual disease that caused it [8]. This process is quite close to how a HVAC engineer approaches the process of FDD. The training and application of a direct classification algorithm for symptom detection does not represent a proper way of symptom detection. This is because: (i) the relationship between faults and symptoms exist in terms of probabilities [13], (ii) a single fault can cause multiple symptoms whereas multiple faults can lead to the same symptom [13], and (iii) depending on factors such as weather, time of the day, building façade & zoning characteristics etc. a fault need not necessarily lead to a symptom. This necessitates the use of a threshold-based rule to determine whether a symptom has been observed or not. For signals/sensor measurements pertaining to CAV fans, static SAT/IAT setpoints, such a simple rule-based method would be preferrable. However, there are components such as HCV & CCV whose operation varies throughout the day depending on the outside air temperature as a result of which the measurements made by sensors such as HC SWT and CC RWT also vary accordingly. Additionally, setpoints are not static. Strategies such as dynamic setpoint control, MPC have made significant progress thereby necessitating the use of black-box modelling techniques using machine learning that can use learn intricate relationships among the various sensor measurements and the control strategies using data collected by the Building Management System (BMS). Such techniques also provide advantages such as not needing an in-depth knowledge of the total control strategy of the building being studied. A simple knowledge of how the HVAC system process and control works would suffice in developing these models by representing the target variable under study as a function of other system variables. Thus, a combination of a ML-based black-box model for predicting fault-free sensor/signal value and a threshold will give the best possible opportunity to both leverage the power of data-driven methods and also detect symptoms in a manner quite similar to how a HVAC engineer monitors buildings. Time series prediction is the technique that can be used to predict what should be the ideal, fault-free value for a target variable (sensor measurement) under question at a specific instant of time when the system variables (features) that serve as inputs to make this prediction are known.

This chapter presents the results of the study that compares multiple ML models in predicting the fault-free values of sensor readings (target variables) and consequently, their performance in detecting symptoms. The prediction of target variables by ML models uses various features, some of which could be influenced by the presence of a fault (hereafter referred to as symptomatic features). The use of such symptomatic features could result in the fault-free value of the target variable being predicted incorrectly and could result in the symptom not being detected. But, at the same time, dropping these features would also mean losing valuable

information needed to predict the target variable. To resolve this, a secondary study is proposed whose objective is to compare the performance of ML algorithms with and without these symptomatic features and also check if these symptomatic features affect the algorithms' prediction of fault-free values and consequently, symptom detection.

4.2 Methodology

The case study chosen for this research is the Kropman office building situated in the city of Breda in The Netherlands. Figures 1 and 2 in Chapter 2 shows the external appearance, zoning and the HVAC system layout of this building. Figure 17 provides the overview of the methodology adopted for this research.

Figure 17: An overview of the adopted methodology.

In the first step, the sensor values recorded by the Building Management System (BMS) were pre-processed by checking for missing values. The BMS records around 330 values from sensors (measuring temperature, pressure, mass flow rate and power) and control signals pertaining to the operation of the HVAC system at 1-minute intervals. Around 53 sensors and control signals pertaining to the AHU operation were chosen for this research. These included ambient air temperature, pressure difference across filters, fan speed control signal, air flow measurement, supply air pressure, valve opening control signals, supply and return water temperatures to and from the coils, boiler & chiller, supply air temperatures & zone air temperatures. Additional features such as hour of the day, day of the week, week & month of the year were added. Data collected from 16 September 2022 to 13 July 2023 was used. Around 28 timestamps were missing from the imported data. Interpolation was not done for these 28 minutes to prevent incorrect data from being used for training.

Subsequently, in step 2, the data corresponding to the heating season and cooling season operations were filtered out. After filtering the relevant dates, in step 3, the symptoms (target variables) to detect the faults were decided and features to predict the symptoms were chosen. For the heating season faults, the columns pertaining to cooling season operation were dropped; the Heating Coil Valve (HCV) opening control signal and the Heating Coil Supply Water Temperature (HC SWT) were chosen as symptoms to be studied. For the cooling season faults, the columns pertaining to heating season operation were dropped; the Indoor Air Temperature (IAT) North zone was chosen as the symptom to be studied. For each symptom, the remaining columns were chosen as input features to begin with. Then, the set of features were checked for multicollinearity. Only one of the features whose correlation was greater than 0.8 was retained. In this step, two separate sets of features were chosen: one without any feature that could be symptomatic under the influence of faults and one that includes such symptomatic features. The number of features with and without possible symptomatic features were 14 and 7 for HCV opening, 15 and 7 for HC SWT, and 16 and 10 for IAT North.

Following this, in step 4, a time series prediction model based on ML was developed for the chosen target variables. In this step, multiple ML algorithms were used to train time-series prediction models to check their fit of the training data by evaluating their performance on the test data. Seven ML algorithms- Support Vector Regression (SVR), Random Forest (RF), Decision Tree (DT), XGBoost, CatBoost, LightGBM and Multi-Layer Perceptron (MLP)- were studied for both sets of features. Following the model training, in step 5, the trained models were validated on a separate dataset kept aside (size being 25% of the training data size). In this phase, error metrics such as R^2, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) & Mean Absolute Percentage Error (MAPE) were calculated and the algorithm performing the best was chosen for use in symptom detection. In step 6, the chosen model is used to predict the fault-free value of the target variable using data from the fault experiment dates as presented in Table 15. Subsequently, the predicted target variable is compared with the actual value obtained from BMS and if the difference between the two is greater than a specific threshold value, then a symptom is deemed to have been detected. Finally, metrics such as True Positives (TPs), True Negatives (TNs), False Positives (FPs), False Negatives (FNs) and Accuracy were calculated to evaluate the performance of algorithms in detecting symptoms.

Table 15: Fault experiment schedule in the case study building.

Season	Fault Experiment	Date	Start Time	End Time	Duration (minutes)
Heating Season	Fault-free	02 & 03 Feb. 2023	07:01	16:45	585*
	Fault-free	06 & 07 Feb. 2023+	07:06	16:45	580*
	Fan stuck 40% speed	08 Feb. 2023	07:06	08:56	111
	HCV stuck 10% open	28 Feb. 2023	10:10	11:05	56
	HCV stuck 100% open	01 Mar. 2023	10:15	14:44	270
	HRW failure	06 Mar. 2023	07:06	16:45	580
Cooling Season	Fault-free	03, 05, 07 Jul. 2023	07:01	16:45	585*
	IAT sensor bias -3°C	04 Jul. 2023	07:01	16:45	585
	SAT sensor bias -3°C	10 Jul. 2023	08:33	16:45	493
	CCV stuck 20% open	11 Jul. 2023	08:25	12:30	246

* Duration for each of the stated dates; + For 07 Feb. 2023, 16 minutes were non-operational

4.3 Results

First, the results of training and testing ML algorithms to predict fault-free operating values of sensor readings are presented. Subsequently, the predicted sensor measurement trends by the chosen ML models applied to the application data, along with the actual data are presented. Finally, the symptoms detected by applying the symptom condition thresholds and fault detection metrics are presented.

4.3.1 Performance of ML algorithms

The R^2 and error metrics values obtained on the test data for the seven studied ML algorithms are presented in Tables 16-21. They are presented for the three studied target variables when considering both a higher number of features and a lower number of features. The algorithms have been arranged in the decreasing order of their R^2 values.

Table 16: Performance of HC SWT prediction models for heating season with higher number of features.

ML algorithm	R^2	MSE	RMSE	MAE	MAPE
MLP	0.9563	1.5618	1.2497	0.9042	2.7464
XGBoost	0.9514	1.7392	1.3188	0.8853	2.5711
LightGBM	0.9489	1.8289	1.3524	0.8911	2.5888
CatBoost	0.9368	2.2593	1.5031	0.9751	2.7343
RF	0.9188	2.9053	1.7045	1.1228	3.1834
DT	0.8996	3.5938	1.8957	1.2767	3.7214
SVR	0.7619	8.5184	2.9186	2.1052	6.4755

Table 17: Performance of HC SWT prediction models for heating season with lower number of features.

ML algorithm	R^2	MSE	RMSE	MAE	MAPE
XGBoost	0.5581	15.8167	3.9770	3.0783	10.0466
DT	0.5036	17.7666	4.2150	3.2496	10.4411
SVR	0.4533	19.5645	4.4232	3.4524	11.0422
LightGBM	0.4433	19.9232	4.4635	3.2931	10.5539
CatBoost	0.4285	20.4515	4.5223	3.4504	10.9940
MLP	0.4074	21.2100	4.6054	3.4499	10.9422
RF	0.3995	21.4902	4.6357	3.5192	11.0971

Table 18: Performance of HCV prediction models for heating season with higher number of features.

ML algorithm	R^2	MSE	RMSE	MAE
RF	0.5193	144.8356	12.0348	9.0758
XGBoost	0.5012	150.2998	12.2597	9.3859
LightGBM	0.3511	195.5314	13.9832	10.7670
MLP	0.3458	197.1204	14.0399	10.8979
CatBoost	0.3355	200.2199	14.1499	10.7835
SVR	0.2935	212.8631	14.5898	10.2863
DT	0.1528	255.2873	15.9777	12.1603

Table 19: Performance of HCV prediction models for heating season with lower number of features.

ML algorithm	R^2	MSE	RMSE	MAE
XGBoost	0.1486	256.5514	16.0172	12.9333
SVR	0.0309	292.0321	17.0889	13.2172
LightGBM	-0.1402	343.5945	18.5363	13.7486
RF	-0.1959	360.3820	18.9837	14.3760

ML algorithm	R^2	MSE	RMSE	MAE
CatBoost	-0.3349	402.2710	20.0567	16.1394
MLP	-0.5982	481.5961	21.9453	17.4033
DT	-1.3277	701.4149	26.4842	20.9772

Table 20: Performance of North IAT prediction models for cooling season with higher number of features.

ML algorithm	R^2	MSE	RMSE	MAE	MAPE
XGBoost	0.8643	0.0139	0.1182	0.0882	0.3791
RF	0.8399	0.0164	0.1284	0.1018	0.4396
LightGBM	0.8320	0.0173	0.1315	0.1024	0.4420
CatBoost	0.8079	0.0198	0.1406	0.1093	0.4725
DT	0.6815	0.0328	0.1811	0.1519	0.6527
SVR	0.6803	0.0329	0.1814	0.1322	0.5718
MLP	-2.0742	0.3164	0.5625	0.5286	2.2753

Table 21: Performance of North IAT prediction models for cooling season with lower number of features.

ML algorithm	R^2	MSE	RMSE	MAE	MAPE
SVR	0.7547	0.0252	0.1589	0.1317	0.5684
CatBoost	0.6181	0.0393	0.1983	0.1476	0.6388
XGBoost	0.5669	0.0446	0.2111	0.1789	0.7701
LightGBM	0.4739	0.0541	0.2327	0.1762	0.7606
RF	0.4604	0.0555	0.2357	0.2053	0.8859
DT	0.3187	0.0701	0.2648	0.2081	0.9012
MLP	-1.1726	0.2236	0.4729	0.4267	1.8419

The range of values obtained for R^2 for HC SWT prediction drops from the range of 0.76 - 0.96 to the range of 0.39 – 0.56 when symptomatic features were removed. These features contribute notably to the model prediction. For predicting the HCV Opening value, the best model with higher number of features could only produce an R^2 value of around 0.52. This indicates that one or more features that would help in predicting the HCV opening value are missing. In the scenario with lower number of features, the best performing model's R^2 drops further to around 0.15. For the IAT North zone temperature prediction for the cooling season, except MLP, the rest of the algorithms produce R^2 values in the range of 0.68 – 0.87 with higher number of features, which goes down to 0.31 – 0.76 when the features that are considered symptomatic are removed. Further, the results show that XGBoost regression performs the best in three out of the six studied scenarios, second best in two scenarios and third best in one scenario. SVR has a better relative ranking amongst the algorithms when using lower number of features to train the model. In contrast, Random Forest performs badly when there are limited number of features. LightGBM and CatBoost (except in one scenario) have average relative rankings consistently compared to the other algorithms. Except in one scenario, DT regression consistently ranks in the bottom positions (last in three scenarios, second-last and third-last in one scenario each). MLP performs badly with lower number of features. Only in the scenario of IAT North zone temperature prediction did MLP finish last even with higher number of features. Based on the obtained results, the best performing algorithms were chosen to be used for symptom detection for the respective scenarios.

4.3.2 Application of ML algorithms for predicting target variables

The best performing ML algorithms from section 5.3.1 were used to predict the target symptom variable for the application data. Figures 20 to 22 show the comparison of the trend predicted by the ML models and the actual recorded values by the sensors. The objective of plotting such figures is to check if the predicted value agrees with the actual value and what is the range of variation of the predicted value from the actual value. The range of variation provides input on what should be the threshold to determine the presence of a symptom. While 4 fault-free days and 3 fault-free days were studied for the heating season and cooling season respectively in this research, figures for only one fault-free day for each scenario are presented.

Figure 18 presents the predicted and actual trends for both HCV opening and HC SWT for the scenario with higher number of features. For the HCV opening, it can be seen that the predicted trend does not follow the actual trend over the day. A high proportion of the deviation varies between -10% to 10% and some values are also beyond this range. In contrast, the trend of predicted HC SWT follows the trend of actual HC SWT and the deviations lie in the range of -2 °C to +2 °C. The difference between the performance of the predicted HCV and predicted HC SWT can be traced back to R^2 values obtained for the best performing values which is just 0.5193 for the HCV opening whereas it is 0.9563 for the HC SWT.

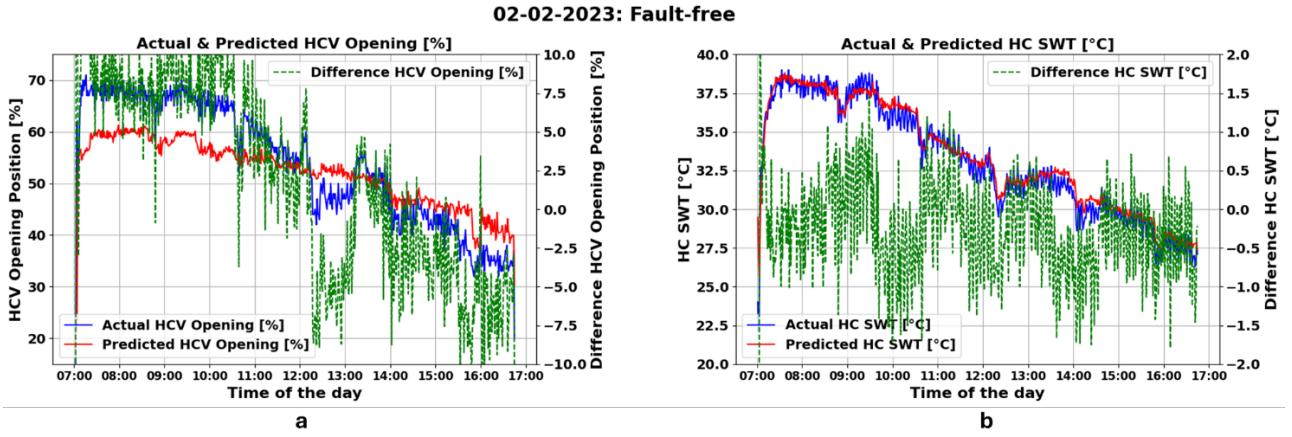

Figure 18: Comparison of predicted and actual time series trend using the set of higher number of features for a. HCV Opening [%] using Random Forest and b. HC SWT [°C] using MLP.

Figure 19 shows the same plots for HCV opening and HC SWT but for the scenario with lower *number of* features. In line with the results of the R^2 score, the trends of the predicted and actual values deviate more than that for the scenario with higher *number of* features. The differences between the predicted and actual values are not the same as in the scenario with higher *number of* features.

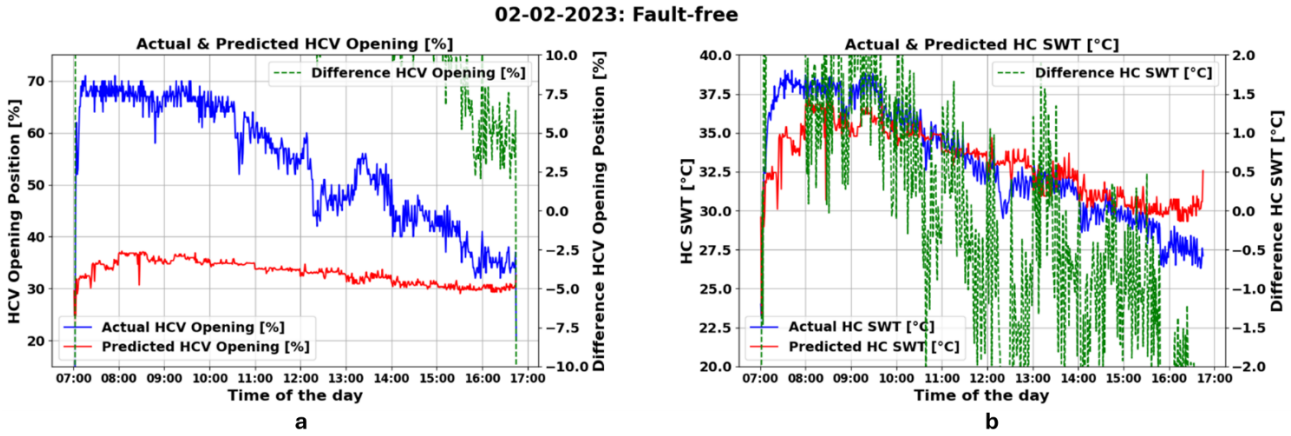

Figure 19: Comparison of predicted and actual time series trend using the set of lower number of features for a. HCV Opening [%] using Random Forest and b. HC SWT [°C] using MLP.

Figure 20a shows the predicted trend and actual trend for IAT North for the scenario with higher number of features. Here, the deviation between the two ranges from -0.5 °C to +0.5 °C. Figure 20b shows the trends for the scenario with lower number of features. Here, the predicted value has changed very slightly and the difference with the actual value is between -0.75 °C to +0.25 °C.

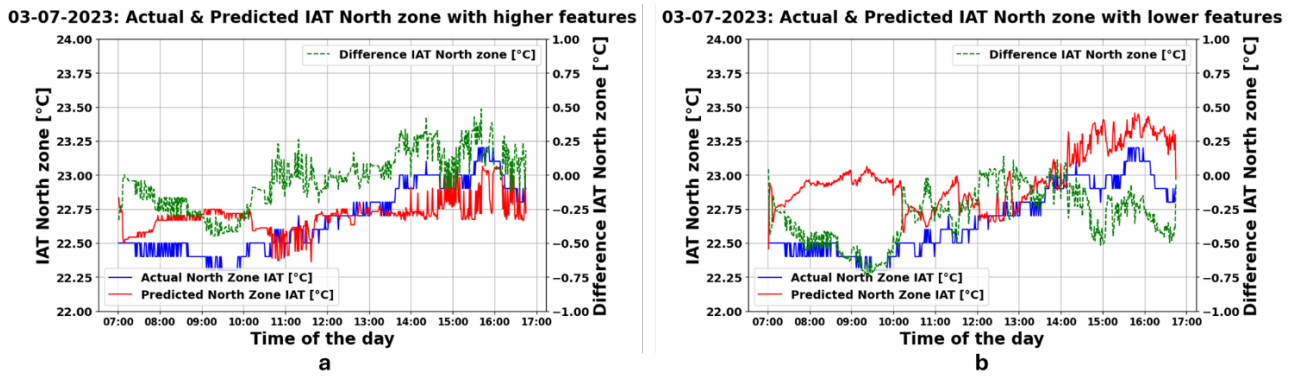

Figure 20: Comparison of predicted and actual time series trend using the set of higher number of features for a. HCV Opening [%] using Random Forest and b. HC SWT [°C] using MLP.

4.3.3 Symptom Detection

From the plots of time series trends presented in the previous section, a threshold of ±10% was chosen for the HCV opening symptom, a threshold of ±2°C was chosen for the HC SWT symptom and a threshold of ±0.5°C for the IAT North symptom. Table 22 presents the TPs, TNs, FPs, FNs and accuracy for the six separate scenarios studied in this research.

Table 22: Performance of North IAT prediction models for cooling season with lower number of features.

Season	Set of features	ML algorithm	Symptom	TPs (%)	TNs (%)	FPs (%)	FNs (%)	Accuracy (%)
Heating	higher	RF	HCV Opening	27.68	52.48	16.99	2.85	80.16
Heating	lower	XGBoost	HCV Opening	27.71	15.31	54.16	2.82	43.02
Heating	higher	MLP	HC SWT	29.15	66.50	2.97	1.38	95.65
Heating	lower	XGBoost	HC SWT	30.23	36.81	32.66	0.30	67.04
Heating	higher	RF + MLP	Any symptom	30.50	51.97	17.50	0.03	82.47
Heating	lower	XGBoost	Any symptom	30.53	2.64	66.83	0	32.87
Cooling	higher	XGBoost	IAT North	36.80	56.97	0.03	6.20	93.77
Cooling	lower	SVR	IAT North	35.53	51.45	5.55	7.47	86.98

It can be observed for the heating season that with reduction in features, the accuracy of symptom detection decreases: close to half for HCV opening symptom, more than 28 %pt decrease for the HC SWT symptom and close to 50 %pt decrease when considering any symptom. For the IAT North symptom, in the cooling season, the decrease in accuracy is comparatively smaller. The FPs increase irrespective of symptom and season with a decrease in input features. This comes at the expense of TNs. The TPs, across seasons and symptoms, hold steady with a decrease in the number of features. The HC SWT symptom detection and IAT North symptom detection have higher accuracies when compared to the other symptoms although the drop in accuracy is steeper for the HC SWT symptom detection when there are lower number of features as compared to the drop in accuracy for the IAT North symptom.

4.4 Conclusions

This research compared the performances of multiple ML algorithms in predicting fault-free trends of target variables used for symptom detection. The performances of algorithms were compared for three symptoms (two for heating season and one for cooling season faults) and for two scenarios each (one with a higher number of features including those which can be symptomatic under the presence of a fault and one with a lower number of features excluding all potential symptomatic features). Across the six scenarios studied, XGBoost regression gave the best prediction for three scenarios; MLP, RF and SVR gave the best prediction for one scenario each. Across all the symptoms, the ML models predicted better and detected symptoms with a higher accuracy when using a greater number of features. The FNs results did not show any notable indication of wrong prediction of fault-free values of target variables when using features that could be symptomatic under the presence of faults. The dropping of “symptomatic features” led to a poorer fit of the developed ML models which in turn led to higher FPs and a reduced accuracy. Limitations include the use of limited symptoms and the testing of only one NN (MLP). Future work includes the detection of all possible symptoms for the studied faults, the study of a larger number of NNs, and comparison of their performances with GBDTs.

5 CONCLUSIONS

Within the B4B project, multiple studies were performed to identify the sensors which can be used to detect faults, the LCSs and LCMs which are suitable candidates for use in the HVAC system for making buildings FDD-ready and the ML algorithm that is best suited for fault (symptom) detection. The main conclusions from these studies are:

1. Sensors for fault detection:

i. The combination of μ and 3σ should be used to detect symptoms when using a WPM-based symptom detection approach using descriptive statistics. This is because this combination gives the least amount of symptoms (outliers) in the fault-free case, which is the expectation in practice.

ii. Excluding the HCV leak 0% fault, for the heating season, out of 260 patterns of symptom occurrences caused by 13 faults across 20 sensors, 134 were deemed as statistically significant when comparing the difference of symptom occurrences between the faulty and fault-free cases for the heating season. A further 15 patterns of symptom occurrences, across three different faults, were also deemed statistically significant even though they were either contrary to the knowledge of how HVAC systems work or lack generalisability. Excluding the CCV leak 0% fault, for the cooling season, out of 264 patterns of symptom occurrences caused by 12 faults across 22 sensors, 139 were deemed as statistically significant. A further 3 patterns of symptom occurrences were also deemed statistically significant even though they were contrary to the knowledge of how HVAC systems work.

iii. The results of the statistical significance tests were used to determine fault-symptom relationships which in turn could be used to decide on the necessary sensors required to detect AHU faults. The IAT North sensor can detect 12 heating season faults. Chiller power meter can detect 10 cooling season faults. Fan failure fault leads to detectable symptoms at the highest number of sensors during both heating and cooling season.

2. Choice of LCMs for sensor installation:

It was concluded that Nuwave, Edimax and Cloudgarden LCMs performed the best and can be chosen for use in FDD.

3. Choice of ML for symptom detection:

i. Across six scenarios studied, XGBoost regression gave the best prediction for three scenarios; MLP, RF and SVR gave the best prediction for one scenario each. XGBoost should be preferred when training ML models.

ii. Across all the symptoms, the ML models predicted better and detected symptoms with a higher accuracy when using a greater number of features.

iii. The FNs results did not show any notable indication of wrong prediction of fault-free values of target variables when using features that could be symptomatic under the presence of faults.

iv. The dropping of “symptomatic features” led to a poorer fit of the developed ML models which in turn led to higher FPs and a reduced accuracy.

REFERENCES

[1] United Nations Environment Programme, '2021 Global status report for buildings and construction Towards a zero-emissions, efficient and resilient buildings and construction sector', 2021. [Online]. Available: www.globalabc.org.

[2] J. Kim, J. Cai, and J. E. Braun, 'Common Faults and Their Prioritization in Small Commercial Buildings; February 2017 - December 2017', 2017. [Online]. Available: www.nrel.gov/publications.

[3] S. Katipamula and M. R. Brambley, 'Review article: Methods for fault detection, diagnostics, and prognostics for building systems—A review, part I', *HVAC and R Research*, vol. 11, no. 1, pp. 3–25, 2005, doi: 10.1080/10789669.2005.10391123.

[4] W. Kim and S. Katipamula, 'A review of fault detection and diagnostics methods for building systems', *Sci Technol Built Environ*, vol. 24, no. 1, pp. 3–21, Jan. 2018, doi: 10.1080/23744731.2017.1318008.

[5] N. Fernandez, Y. Xie, S. Katipamula, M. Zhao, W. Wang, and C. Corbin, 'Impacts of Commercial Building Controls on Energy Savings and Peak Load Reduction', 2017.

[6] J. Granderson, G. Lin, and R. Singla, 'Lawrence Berkeley National Laboratory Recent Work Title Commercial Fault Detection and Diagnostics Tools: What They Offer, How They Differ, and What's Still Needed Permalink https://escholarship.org/uc/item/4j72k57p Publication Date', Aug. 2018. doi: 10.20357/B7V88H.

[7] N. Djuric and V. Novakovic, 'Review of possibilities and necessities for building lifetime commissioning', Feb. 2009. doi: 10.1016/j.rser.2007.11.007.

[8] A. Taal, L. Itard, and W. Zeiler, 'A reference architecture for the integration of automated energy performance fault diagnosis into HVAC systems', *Energy Build*, vol. 179, pp. 144–155, Nov. 2018, doi: 10.1016/j.enbuild.2018.08.031.

[9] Y. Zhao, T. Li, X. Zhang, and C. Zhang, 'Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future', Jul. 01, 2019, *Elsevier Ltd*. doi: 10.1016/j.rser.2019.04.021.

[10] Z. Chen *et al.*, 'A review of data-driven fault detection and diagnostics for building HVAC systems', *Appl Energy*, vol. 339, Jun. 2023, doi: 10.1016/j.apenergy.2023.121030.

[11] J. Schein, 'Results from Field Testing of Embedded Air Handling Unit and Variable Air Volume Box Fault Detection Tools', Oct. 2006. [Online]. Available: https://www.researchgate.net/publication/237457515

[12] K. Bruton, P. Raftery, B. Kennedy, M. M. Keane, and D. T. J. O'Sullivan, 'Review of automated fault detection and diagnostic tools in air handling units', 2014, *Springer Science and Business Media Netherlands*. doi: 10.1007/s12053-013-9238-2.

[13] Y. Zhao, J. Wen, F. Xiao, X. Yang, and S. Wang, 'Diagnostic Bayesian networks for diagnosing air handling units faults – part I: Faults in dampers, fans, filters and sensors', *Appl Therm Eng*, vol. 111, pp. 1272–1286, Jan. 2017, doi: 10.1016/j.applthermaleng.2015.09.121.

[14] M. S. Breuker and J. E. Braun, 'Common faults and their impacts for rooftop air conditioners', *HVAC and R Research*, vol. 4, no. 3, pp. 303–318, 1998, doi: 10.1080/10789669.1998.10391406.

[15] M. C. Comstock, J. E. Braun, and E. A. Groll, 'The sensitivity of chiller performance to common faults', *HVAC and R Research*, vol. 7, no. 3, pp. 263–279, 2001, doi: 10.1080/10789669.2001.10391274.

[16] S. H. Cho, H. C. Yang, M. Zaheer-Uddin, and B. C. Ahn, 'Transient pattern analysis for fault detection and diagnosis of HVAC systems', *Energy Convers Manag*, vol. 46, no. 18–19, pp. 3103–3116, Nov. 2005, doi: 10.1016/j.enconman.2005.02.012.

[17] L. Morawska *et al.*, 'Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?', Jul. 01, 2018, *Elsevier Ltd*. doi: 10.1016/j.envint.2018.04.018.

[18] J. P. Sá, M. C. M. Alvim-Ferraz, F. G. Martins, and S. I. V. Sousa, 'Application of the low-cost sensing technology for indoor air quality monitoring: A review', Nov. 01, 2022, *Elsevier B.V.* doi: 10.1016/j.eti.2022.102551.

[19] I. Demanega, I. Mujan, B. C. Singer, A. S. Anđelković, F. Babich, and D. Licina, 'Performance assessment of low-cost environmental monitors and single sensors under variable indoor air quality and thermal conditions', *Build Environ*, vol. 187, Jan. 2021, doi: 10.1016/j.buildenv.2020.107415.

[20] H. Zheng, V. Krishnan, S. Walker, M. Loomans, and W. Zeiler, 'Laboratory evaluation of low-cost air quality monitors and single sensors for monitoring typical indoor emission events in Dutch daycare centers', *Environ Int*, vol. 166, Aug. 2022, doi: 10.1016/j.envint.2022.107372.

[21] Y. Chen, G. Lin, Z. Chen, J. Wen, and J. Granderson, 'A simulation-based evaluation of fan coil unit fault effects', *Energy Build*, vol. 263, May 2022, doi: 10.1016/j.enbuild.2022.112041.

[22] Y. Chen, Z. Chen, G. Lin, Y. Zhang, and S. Ye, 'A novel evaluation method of measurement sensitivities on common faults in VAV HVAC systems', *Build Environ*, vol. 261, Aug. 2024, doi: 10.1016/j.buildenv.2024.111683.
[23] J. Hwan Cha *et al.*, 'A new integrated framework to fault detection and diagnosis of air handling unit: Emphasizing the impact of symptoms', *Energy Build*, vol. 319, Sep. 2024, doi: 10.1016/j.enbuild.2024.114474.
[24] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, 'Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median', *J Exp Soc Psychol*, vol. 49, no. 4, pp. 764–766, Jul. 2013, doi: 10.1016/j.jesp.2013.03.013.
[25] Y. Yu, D. Woradechjumroen, and D. Yu, 'A review of fault detection and diagnosis methodologies on air-handling units', 2014, *Elsevier Ltd*. doi: 10.1016/j.enbuild.2014.06.042.
[26] A. Thamban, 'Fault detection and diagnosis of the low ΔT syndrome in cooling coils of chilled water systems', Eindhoven, Aug. 2022. Accessed: Jul. 07, 2024. [Online]. Available: https://research.tue.nl/en/publications/fault-detection-and-diagnosis-of-the-low-t-syndrome-in-cooling-co
[27] S. Gopalan, A. Rijs, S. Chitkara, A. Thamban, and R. Kramer, 'Fault prioritisation for Air Handling units using fault modelling and actual fault occurrence data', *Energy Build*, vol. 319, Sep. 2024, doi: 10.1016/j.enbuild.2024.114476.
[28] M. J. Gestwick and J. A. Love, 'Trial application of ASHRAE 1051-RP: Calibration method for building energy simulation', *J Build Perform Simul*, vol. 7, no. 5, pp. 346–359, Sep. 2014, doi: 10.1080/19401493.2013.838698.
[29] ASHRAE, 'ASHRAE Guideline 14-2014, measurement of Energy, Demand, and Water Savings', Atlanta, Dec. 2014. [Online]. Available: www.ashrae.org
[30] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, 'A Review on Outlier/Anomaly Detection in Time Series Data', *ACM Comput Surv*, vol. 54, no. 3, pp. 1–33, Apr. 2022, doi: 10.1145/3444690.
[31] R. Wu and E. J. Keogh, 'Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress', *IEEE Trans Knowl Data Eng*, vol. 35, no. 3, pp. 2421–2429, Mar. 2023, doi: 10.1109/TKDE.2021.3112126.
[32] A. P. Field, *Discovering Statistics Using IBM SPSS Statistics*, 5th ed. Newbury Park: Sage, 2018.
[33] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences Second Edition*, Second. Mahwah: Lawrence Erlbaum Associates, Publishers, 1988.
[34] S. Nakagawa and I. C. Cuthill, 'Effect size, confidence interval and statistical significance: a practical guide for biologists', *Biological Reviews*, vol. 82, no. 4, pp. 591–605, Nov. 2007, doi: 10.1111/j.1469-185X.2007.00027.x.
[35] D. Cousineau and S. Chartier, 'Outliers detection and treatment: a review', *Int J Psychol Res (Medellin)*, vol. 3, no. 1, pp. 58–67, 2010.
[36] J. Cohen, 'Statistical Power Analysis', *Curr Dir Psychol Sci*, vol. 1, no. 3, pp. 98–101, Jun. 1992, doi: 10.1111/1467-8721.ep10768783.
[37] U.S. Department of Energy, 'EnergyPlusTM Version 22.1.0 Documentation Engineering Reference', 2022.
[38] Y. Zhao, J. Wen, and S. Wang, 'Diagnostic Bayesian networks for diagnosing air handling units faults - Part II: Faults in coils and sensors', *Appl Therm Eng*, vol. 90, pp. 145–157, Jul. 2015, doi: 10.1016/j.applthermaleng.2015.07.001.
[39] Y. Chen, J. Wen, O. Pradhan, L. J. Lo, and T. Wu, 'Using discrete Bayesian networks for diagnosing and isolating cross-level faults in HVAC systems', *Appl Energy*, vol. 327, Dec. 2022, doi: 10.1016/j.apenergy.2022.120050.
[40] T. Li, Y. Zhao, C. Zhang, J. Luo, and X. Zhang, 'A knowledge-guided and data-driven method for building HVAC systems fault diagnosis', *Build Environ*, vol. 198, Jul. 2021, doi: 10.1016/j.buildenv.2021.107850.
[41] D. Dey and B. Dong, 'A probabilistic approach to diagnose faults of air handling units in buildings', *Energy Build*, vol. 130, pp. 177–187, Oct. 2016, doi: 10.1016/j.enbuild.2016.08.017.
[42] L. Schellen, M. Loomans, W. Van, M. Lichtenbelt, A. Frijns, and M. De Wit, 'Assessment of thermal comfort in relation to applied low exergy systems The design of a climate chamber and the use of a thermophysiological model', in *Proceedings of Conference: Adapting to Change: New Thinking on Comfort*, 2010. [Online]. Available: http://nceub.org.uk
[43] R. M. Duvall *et al.*, 'Performance Testing Protocols, Metrics, and Target Values for Fine Particulate Matter Air Sensors Use in Ambient, Outdoor, Fixed Site, Non-Regulatory Supplemental and Informational Monitoring Applications', Feb. 2021.
[44] R. Williams *et al.*, 'Air Sensor Guidebook', Jun. 2014. [Online]. Available: www.epa.gov/ord
[45] G. Hu, T. Zhou, and Q. Liu, 'Data-Driven Machine Learning for Fault Detection and Diagnosis in Nuclear Power Plants: A Review', *Front Energy Res*, vol. 9, May 2021, doi: 10.3389/fenrg.2021.663296.

APPENDICES

Appendix A

To demonstrate the simplified formula of SOP, the South Zone T symptom for the CCV stuck 0% flow is used. There are five operating windows as shown in Table 3. First, the SOP is calculated based on the formula prescribed by Y. Chen et al. [14]. It is the weighted sum of SOP within each window where the weight for each window's SOP is its operating ratio.

$$SOP = \frac{225}{2886} \times 0.0822 + \frac{10051}{11339} \times 0.3230 + \frac{9168}{9601} \times 0.2735 + \frac{6911}{7637} \times 0.2176 + \frac{3085}{3637} \times 0.1024$$

$$SOP = 0.8387 = 83.87\%$$

Now, the SOP equation above is re-arranged in the following manner:

$$SOP = \frac{225}{2886} \times 0.0822 + \frac{10051}{11339} \times 0.3230 + \frac{9168}{9601} \times 0.2735 + \frac{6911}{7637} \times 0.2176 + \frac{3085}{3637} \times 0.1024$$

$$SOP = \frac{225}{2886} \times \frac{2886}{35100} + \frac{10051}{11339} \times \frac{11339}{35100} + \frac{9168}{9601} \times \frac{9601}{35100} + \frac{6911}{7637} \times \frac{7637}{35100} + \frac{3085}{3637} \times \frac{3637}{35100}$$

$$SOP = \frac{225}{35100} + \frac{10051}{35100} + \frac{9168}{35100} + \frac{6911}{35100} + \frac{3085}{35100}$$

$$SOP = \frac{225 + 10051 + 9168 + 6911 + 3085}{35100}$$

$$SOP = 0.8387 = 83.87\%$$

Appendix B

Figure 12 presents the complete fault-symptom relationships for the heating season. Here the explanation for only those results are provided which are either contrary to the expected results based on the knowledge of how HVAC system works or not applicable to all HVAC systems.

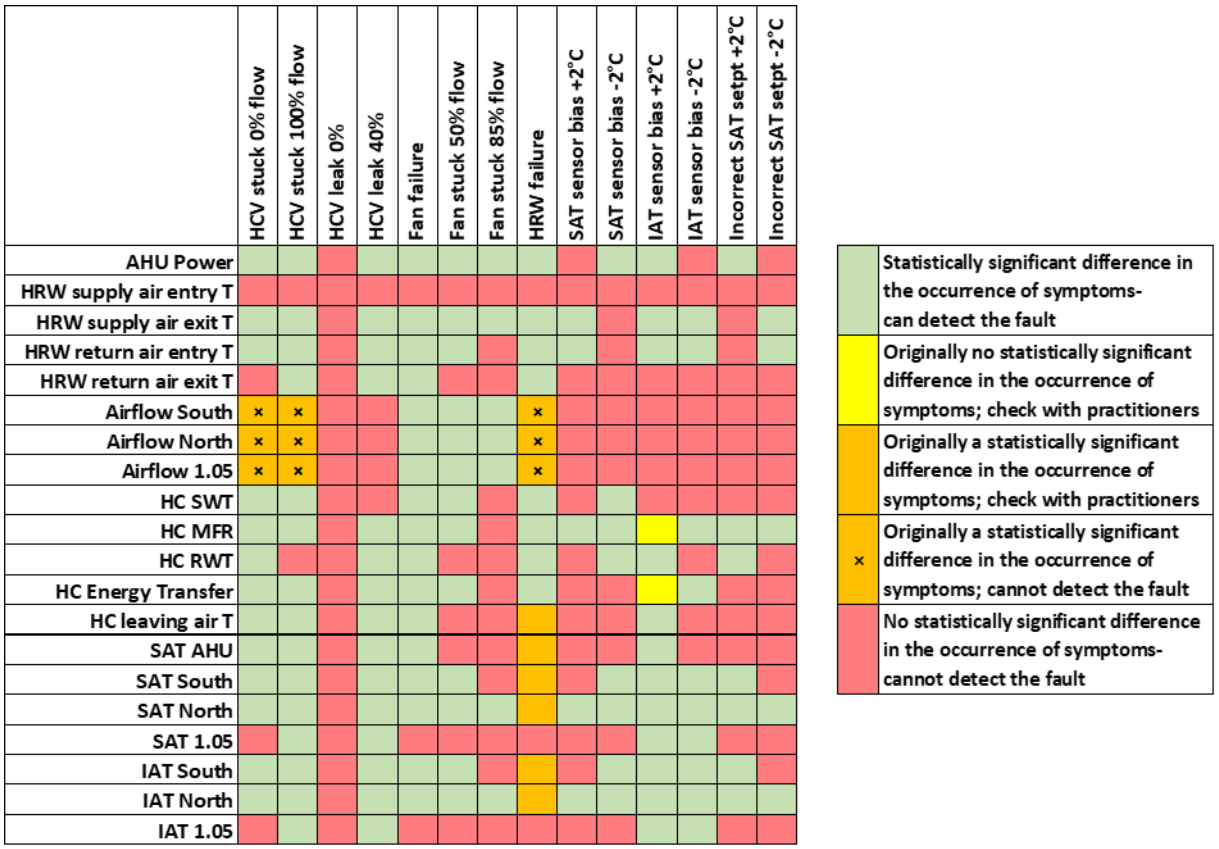

Figure 21: Complete fault-symptom relationships for the heating season.

i. In addition to the fan faults, the statistical significance testing resulted in three other faults- HCV stuck 0% flow, HCV stuck 40% flow and HRW failure- leading to a statistically significant difference in the occurrence of symptoms at the air flow rate meters for the three zones. The SOP values for all the three air flow meters for these three faults are 3.7%, 8.7% and 2.1% respectively. However, in reality, a fault of the HCV or the HRW does not affect the fan operation. Additionally, a HVAC system modelled in EnergyPlus with constant volume fan can control the ventilation only by using the On/Off strategy [37]. The obtained symptom occurrences at the three air flow meters for these three faults require further investigation which is outside the scope of this research. Comparing the effect sizes of the changes in symptom occurrence obtained at the airflow meters for these three faults against the effect sizes obtained for a mild fan fault such as the fan stuck at 85% air flow fault, it is clear that the order of magnitude of the effect sizes also differ by a very high factor. Therefore, the airflow meters need not be considered as symptoms for the HCV stuck faults and the HRW failure faults.

ii. When considering the HC MFR sensor, the IAT sensor bias -2 °C fault leads to a statistically significant change in the occurrence of symptoms with an SOP value of 1.2%. This fault has an effect size of 0.25. In contrast, the IAT sensor bias +2 °C fault does not lead to a statistically significant change in the occurrence of symptoms despite leading to an SOP value of 1.8% and an effect size of 0.32. A comparison of the SDCD values indicated that the IAT sensor bias -2 °C fault leads to an SDCD of 6 minutes and the IAT sensor bias +2 °C fault leads to an SDCD of 4 minutes when compared to the fault-free case where the SDCD value is only 1 minute. In Gopalan et al. [27], an explanation on how a HVAC system using a PI controller would react to an IAT positive bias fault was provided. It was concluded that there would be a net decrease in the heat transfer at the heating coil. As a result, the sensors at the HC MFR and HC Energy Transfer sensors can indeed be used to detect symptoms caused as a result of the IAT sensor bias +2 °C. Note: In the results of the statistical significance testing, there has been no fault case where there has been a statistically significant difference in the occurrence of symptoms at the HC Energy Transfer sensor but no statistically significant difference in the occurrence of symptoms at the HC MFR sensor. Therefore, the HC MFR sensor is also chosen for symptom detection.

iii. The HRW failure fault leads to a statistically significant difference in the occurrence of symptoms at the following sensors: HC Air leaving T sensor, SAT AHU sensor, SAT and IAT sensors at both North and South zones. In practice, the heat that could not be recovered by the HRW will be compensated by the heating coil to meet the SAT requirements. In case the heating coil cannot provide the required energy, then both the SAT & IAT setpoints are not met, leading to detectable symptoms. It is recommended to consult HVAC practioners/building engineers on a case-by-case basis before deciding if there is a need to use these sensors to detect HRW failure.

Figure 13 presents the complete fault-symptom relationships for the cooling season. Here the explanation for only those results are provided which are contrary to the expected results based on the knowledge of how HVAC system works.

	CCV stuck 0% flow	CCV stuck 50% flow	CCV leak 0% flow	CCV leak 40% flow	Fan failure	Fan stuck 50% airflow	Fan stuck 85% airflow	SAT sensor bias +2°C	SAT sensor bias -2°C	IAT sensor bias +2°C	IAT sensor bias -2°C	Incorrect SAT setpt +2°C	Incorrect SAT setpoint -2°C
Chiller Power													
Airflow South	×												
CC South SWT													
CC South MFR													
CC South RWT													
CC South Energy Transfer													
SAT South													
IAT South													
Airflow North	×												
CC North SWT													
CC North MFR													
CC North RWT													
CC North Energy Transfer													
SAT North													
IAT North													
Airflow 1.05	×												
CC 1.05 SWT													
CC 1.05 MFR													
CC 1.05 RWT													
CC 1.05 Energy Transfer													
SAT 1.05													
IAT 1.05													

	Statistically significant difference in the occurrence of symptoms- can detect the fault
	Originally no statistically significant difference in the occurrence of symptoms; check with practitioners
×	Originally a statistically significant difference in the occurrence of symptoms; cannot detect the fault
	No statistically significant difference in the occurrence of symptoms- cannot detect the fault

Figure 22: Complete fault-symptom relationships for the cooling season.

i. In addition to the fan faults, the figure shows that the CCV stuck 0% flow fault leads to a statistically significant difference in the occurrence of symptoms at the air flow rate meters for the three zones. The SOP values for the three air flow meters for this fault are 9.1%, 6.4% and 7.4% respectively. However, quite similar to the results obtained for the heating season, the order of magnitude of the effect sizes obtained at the airflow meters for this fault differ from the order of magnitude of the effect sizes obtained at the airflow meters for the fan stuck 85% air flow fault by a very high factor. The airflow meters need not be considered as symptoms for the HCV stuck faults.

ii. The IAT sensor biases do not lead to any statistically significant difference in the occurrence of symptoms across all the sensors. It appears that the thermostat bias fault object in EnergyPlus did not cause any significant differences in the occurrence of symptoms at all the studied sensors. However, a working knowledge of the AHU operation indicates that the IAT negative sensor bias fault leads to a reduction in the energy transferred at the cooling coil and an increase in SAT value, as presented in Gopalan et al. [27]. It is recommended to use IAT sensors to detect both positive and negative IAT sensor bias faults. Additionally, it is also recommended to use CC energy transfer and SAT sensor to

detect IAT sensor negative bias fault. It must be noted that the IAT sensor bias faults could also possibly be detected at the chiller power meter. However, the knowledge of how a change in cooling energy demand at the cooling coils affects the chiller operation and its energy consumption is unknown.

Appendix C

Table 23: Features used in the ML prediction models.

Scenario	Set of features	Features used
HC SWT	higher	'Ambient air temperature [°C]', 'Pressure difference supply air filter [Pa]', 'Heating Coil Valve Opening control signal [%]', 'Supply fan speed control signal [%]', 'Air Handling Unit Supply Air Temperature [°C]', 'IAT South zone [°C]', 'Calculated air flow North zone [m3/h]', 'IAT North zone [°C]', 'Calculated air flow zone 1.05 [m3/h]', 'IAT zone 1.05 [°C]', 'Pressure difference return air filter [Pa]', 'Heat recovery wheel operation status [%]', 'Day of the week', 'hour of the day', 'week of the year'
	lower	'Ambient air temperature [°C]', 'IAT North zone [°C]', 'Calculated air flow zone 1.05 [m3/h]', 'IAT zone 1.05 [°C]', 'Day of the week', 'hour of the day', 'week of the year'
HCV Opening	higher	'Ambient air temperature [°C]', 'Pressure difference supply air filter [Pa]', 'Supply fan speed control signal [%]', 'Air Handling Unit Supply Air Temperature [°C]', 'IAT South zone [°C]', 'Calculated air flow North zone [m3/h]', 'IAT North zone [°C]', 'Calculated air flow zone 1.05 [m3/h]', 'IAT zone 1.05 [°C]', 'Pressure difference return air filter [Pa]', 'Heat recovery wheel operation status [%]', 'Day of the week', 'hour of the day', 'week of the year'
	lower	'Ambient air temperature [°C]', 'IAT North zone [°C]', 'Calculated air flow zone 1.05 [m3/h]', 'IAT zone 1.05 [°C]', 'Day of the week', 'hour of the day', 'week of the year'
IAT North	higher	'Ambient air temperature [°C]', 'Supply fan speed control signal [%]', 'Supply Air Temperature North zone [°C]', 'Return air temperature North zone [°C]', 'Calculated air flow zone 1.05 [m3/h]', 'CO_2 concentration zone 1.05 [ppm]', 'Chiller power [kW]', 'Heat recovery wheel operation status [%]', 'Chilled water supply temperature from chiller [°C]', 'South zone cooling coil chilled water mass flow rate [L/h]', 'North zone cooling coil chilled water mass flow rate [L/h]', 'Return water temperature North zone cooling coil [°C]', 'Pressure difference chilled water inlet & outlet North zone cooling coil [kPa]', 'Day of the week', 'hour of the day', 'week of the year'
	lower	'Ambient air temperature [°C]',

Scenario	Set of features	Features used
		'Supply fan speed control signal [%]', 'Calculated air flow zone 1.05 [m3/h]', 'CO_2 concentration zone 1.05 [ppm]', 'Chiller power [kW]', 'Chilled water supply temperature from chiller [°C]', 'Pressure difference chilled water inlet & outlet North zone cooling coil [kPa]', 'Day of the week', 'hour of the day', 'week of the year'