# Parameter selection in data-driven fault detection and diagnosis of the air conditioning system

**Noor Asyikin Sulaiman[1], Md Pauzi Abdullah[2], Hayati Abdullah[3], Muhammad Noorazlan Shah Zainudin[4], Azdiana Md Yusop[5], Siti Fatimah Sulaiman[6]**

[1,4,5,6]Centre for Telecommunication Research & Innovation (CeTRI), Fakulti Kejuruteraan Elektronik & Kejuruteraan Komputer (FKEK), Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

[2,3]Centre of Electrical Energy Systems (CEES), School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), 81310 Skudai, Johor, Malaysia

**Article Info**

***Article history:***

Received Oct 1, 2018
Revised Dec 10, 2018
Accepted Jan 25, 2019

***Keywords:***

Air conditioning system
Data-driven FDD
Parameter selection

**ABSTRACT**

Data-driven fault detection and diagnosis system (FDD) has been proven as simple yet powerful enough to identify soft and abrupt faults in the air conditioning system, leading to energy saving. However, the challenge of data-driven FDD is to obtain reliable operation data from the actual building. Therefore, a lab-scaled centralised chilled water air conditioning system was successfully developed in this paper. All necessary sensors were installed to generate reliable operation data for the data-driven FDD. Nevertheless, if a practical system is considered, the number of sensors required would be extensive as it depends on the number of rooms in the building. Hence, parameters impact in the dataset were also investigated to identify critical parameters for fault classifications. The analysis results had identified four critical parameters for data-driven FDD: the rooms' temperature, $T_{TCx}$, supplied chilled water temperature, $T_{CHWS}$, supplied chilled water flow rate, $V_{CHWS}$, and supplied cooled water temperature, $T_{CWS}$. Results showed that the data-driven FDD successfully diagnosed all six conditions correctly with the proposed parameters for more than 92.3% accuracy; only 0.6% - 3.4% differed from the original dataset's accuracy. Therefore, the proposed parameters can reduce the number of sensors used for practical buildings, thus reducing installation costs without compromising the FDD accuracy.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*

***Corresponding Author:***

Noor Asyikin Sulaiman,
Centre for Telecommunication Research & Innovation (CeTRI),
Fakulti Kejuruteraan Elektronik & Kejuruteraan Komputer (FKEKK),
Universiti Teknikal Malaysia Melaka (UTeM),
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.
Email: noor_asyikin@utem.edu.my

## 1. INTRODUCTION

Faults in air conditioning systems, especially soft faults, are hard to detect. Even a regularly maintained building may suffer from soft faults without realising it [1]. Therefore, fault detection and diagnosis (FDD) plays an important role in building energy savings. Successful FDD can save up to 40% of air conditioning energy consumption [2]. One of the FDD methods is model-based FDD, which relies on mathematical modelling to represent the system. The detailed physical modelling derived using the first principle method is the most accurate way to describe the air conditioning system as proposed in [3]–[5]. However, since the system itself is a complex and dynamic system, the development of mathematical modelling is complex and requires detailed information regarding the system and is challenging to derive [6]. In contrast, simplified mathematical modelling using a lumped parameter approach developed in [7], [8] are simpler to derive. However, the number of available fault models of air conditioning systems is still limited [9]. One of the reasons was that most of the modellings are developed for a specific system. Thus, some adjustment needs to be made to use in other types of air conditioning system.

Recently, researchers are exploring more on data-driven FDD due to its simple yet reliable method. This method has gained much interest among researchers in many areas, such as in air conditioning systems [10]–[14], power generation systems [15]–[19] and motor drive systems [20], [21]. The method is simple to develop because it only requires historical data to train and validate its operational data. Thus, it is easy to develop, but it requires fault-free training data to classify other faults. Otherwise, the classifier model would recognise faults as the standard operating performance.

Current FDD trends for air conditioning systems only focus on individual component, such as the chiller as in [10], [14], [22], and air handling unit (AHU) as in [23]–[27]. However, no FDD research considers faults across the entire air-conditioning system even though all components are interconnected [9]. Thus, faults in one component may affect other components' parameters. Therefore, by combining faults across the entire system, the ability of the FDD system to diagnose with correct faults can be analysed. To fill up this gap, Chen [28] has proposed data-driven FDD using the Bayesian network (BN) for the whole building fault, including faults across chiller, AHU, and operation schedule. However, this research does not cover faults across the cooling tower, which is also one of the air conditioning system components. One of the limitations of his research is that some faults may not be identified under certain weather, operation, or internal load conditions. Indeed, it is one of the biggest challenges for data-driven FDD in the actual building.

There are many challenges to obtaining reliable fault-free and faults operation data in the actual building. Firstly, the initial building operation data might differ from those applied later in the building's lifetime. Furthermore, the external factors, such as environment and usage patterns, may vary the results as in [28]. It is also a challenge to simulate faults in actual buildings as it may disturb the thermal comfort of the occupants. Therefore, in our previous studies in [13], [29], we developed a lab-scaled chilled water air conditioning system. The data was used to develop three machine learning models as in [13]: deep learning, support vector machine (SVM) and multi-layer perceptron (MLP) for data-driven FDD of the entire system faults. It covers the entire system faults, which are faults across the chiller, AHU, and the cooling tower. Results showed that all models were successfully identified all faults for more than 95%.

Deep learning, SVM and MLP are among the most widely used for classification proses. For instance, deep learning was successfully proposed as FDD in Tennessee Eastman (TE) process as presented in [30]. Results show deep learning model outshines the other five classifier models. Likewise, Yan [31] successfully proposed SVM as FDD in the chiller system. SVM also shows the highest accuracy compared to other methods in detecting breast cancer [32], [33]. Meanwhile, MLP successfully diagnosed bladder cancer and predicted faults in yacht hydrodynamics, as portrayed in [34] and [35].

Even though the FDD in [13] successfully diagnosed the faults, it requires many sensors to be implemented in actual buildings. Nevertheless, most air conditioning systems in non-residential buildings have a limited number of sensors, and most of them were installed for control purposes only [6]. Hence, it needs a substantial additional cost to add more sensors to the building. Furthermore, the accuracy of the data-driven method depends on the parameter data collected from the system. The more parameters in the dataset, the better FDD accuracy will be produced, and the bigger the system is, the more parameters will be required. Therefore, it is essential to identify the impact of those parameters on their ability to detect faults. The unimportant parameters can be eliminated to reduce the installation cost without compromising FDD accuracy. Thus, the proposed parameters can still avoid unnecessary energy wastage with smaller installation costs.

In this paper, the impact of each parameter in FDD was investigated to identify the critical parameters. New dataset combinations were developed based on standard deviation and accuracy percentage values. Each combination was then evaluated using deep learning, SVM and MLP model developed in [13]. The performance of the proposed critical parameters was then compared with the performance of the original dataset in [13]. This paper was written in four sections, where some research backgrounds are presented in Section 1. Then, the research methodology is presented carefully in Section 2. It includes the development of the lab-scaled system, the fault simulation on the system and the investigation of each parameter's impact. Section 3 elaborates the outcome of this research in detail. Lastly, the conclusions are written up in Section 4.

## 2. RESEARCH METHOD

This section explains the research methodology of this research. It involves the development of the lab-scaled system and the selections of the parameters for the FDD.

### 2.1. Lab scaled of centralised chilled water air conditioning system

Figure 1 shows the lab-scaled system developed in this research as described in [7], [8], [13]. It consists of a chiller, cooling tower, AHU, and two rooms to replicate an actual centralised chilled water air conditioning system. The chiller used is a ready-made chiller system equipped with a chilled water tank, and the cooling tower is designed as a counter flow type. The AHU system has a cooling coil, a fan, supply and return ducts for each room, and dampers. The speed of the fan can be varied to achieve a specific supplied airflow rate. The rooms were constructed by insulated board and poly-carbonate, and each of them sizes 2.4m × 1.2m × 1.6m. Five bulbs rated 100watt each was installed in each room to simulate heat from equipment and occupants.

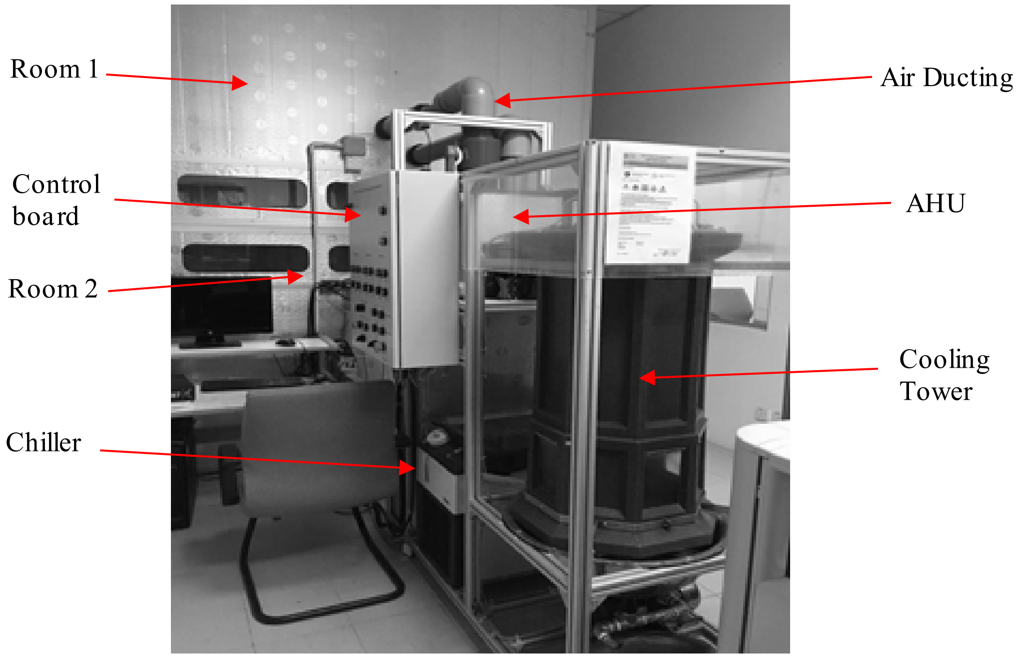


Figure 1. The lab-scaled of the chilled water system

The system is a set of standalone and self-contained equipment. It has a structured platform to accommodate the cooling tower, water-cooled chiller, and AHU system. Two rooms were installed next to the structured platform. Four lockable castor wheels were mounted at the bottom of the structure platform for easy mobilisation. The size of the platform is 64cm (W) x 150cm (L) x 170cm (H). A control board is used to control and operate the system. The system was equipped with fourteen sensors: thermocouple sensors, water flow rate sensors, airflow rate sensors, and current sensor, and the details of each sensor and the parameters measured were tabulated in Table 1. The system coefficient of performance (COP) was also analysed and presented in Sulaiman et al. [13]. The results show that the COPs reduce when the system has faults, which is consistent with the results energy audit of the actual system presented in Sulaiman et al. [1].

Table 1. List of the sensors in the lab-scaled system

| Sensor Type | Parameters measured |
|---|---|
| Temperature sensor | $T_{TC1}$ = Air temperature in Room 1 |
| | $T_{TC2}$ = Air temperature in Room 2 |
| | $T_{S1}$ = Air temperature at ducting Room 1 |
| | $T_{S2}$ = Air temperature at ducting Room 2 |
| | $T_{CHWS}$ = Supplied chilled water temperature |
| | $T_{CHWR}$ = Returned chilled water temperature |
| | $T_{CWS}$ = Supplied cooled water temperature |
| | $T_{CWR}$ = Returned cooled water temperature |
| Airflow rate sensor | $V_{S1}$ = Airflow rate at ducting Room 1 |
| | $V_{S2}$ = Airflow rate at ducting Room 2 |
| Water flow rate sensor | $V_{CHWS}$ = Supplied chilled water flow rate |
| | $V_{CHWR}$ = Returned chilled water flow rate |
| | $V_{CWS}$= Supplied cooled water flow rate |
| Current sensor | $C_{CH}$ = Compressor current |

All parameters in Table 1 was logged during various conditions simulations in the lab-scaled system. The conditions simulated as described in Table 2. The location and type of faults were also portrayed in the table. It includes five faults throughout the entire system and one normal without fault condition. Three machine learning models were used to classify all conditions as described in Table 3. The parameter setting for each model is displayed in the table. All classifier models have successfully identified all conditions as presented in Sulaiman et al. [13].

Table 2. List of conditions simulated in the lab-scaled system [13]

| Condition | Location of fault | Type of fault |
|---|---|---|
| Normal (no-fault) | -- | -- |
| Evaporator Clogging | Chiller | Soft |
| Compressor Failure | Chiller | Abrupt |
| Cooling Tower Fan Faulty | Cooling Tower | Soft |
| Damper Stuck | AHU | Soft |
| Air Ducting Leakage | AHU | Soft |

Table 3. Simulation parameters [13]

| Models | Parameter setting |
|---|---|
| Deep learning | Activation function for hidden layer: sigmoid<br>Activation function for the output layer: softmax<br>Optimization: Stochastic Gradient Descent (SGD) |
| SVM | Kernel function: polykernel |
| MLP | Activation function: sigmoid |

### 2.2. Parameter selection

Out of fourteen sensors, six were installed in the two rooms, three sensors for each room. If a practical system is considered, the number of sensors required would be extensive as it depends on the number of rooms in the actual building. In other words, more cost is needed as three sensors are required for each room. Therefore, it is essential to investigate the impact of these parameters in classifying the faults. Insignificant sensors can be eliminated to reduce installation costs. However, the elimination must not affect accuracy. Table 4 represents the list of sensors and their location throughout the entire system. The data were categorised into two, Group A and Group B. Group A is a set of parameters related to the rooms, and Group B is a set of parameters associated with the central unit.

Table 4. List of conditions simulated in the lab-scaled system

| Group | Location of the sensors | Parameters measured |
|---|---|---|
| Group A (Sensors located at rooms) | Room 1 | $T_{TC1}$<br>$T_{S1}$<br>$V_{S1}$ |
| | Room 2 | $T_{TC2}$<br>$T_{S2}$<br>$V_{S2}$ |
| Group B (Sensors located at the central unit) | The central unit of the system | $T_{CHWS}$<br>$T_{CHWR}$<br>$T_{CWS}$<br>$T_{CWR}$<br>$V_{CHWS}$<br>$V_{CHWR}$<br>$V_{CWS}$<br>$C_{CH}$ |

In general, the number of parameters can be presented as,

$$n_P = n_A N_{ROOM} + n_B = 3N_{ROOM} + 8, \qquad , \tag{1}$$

where $n_P$ represents the number of total parameters, $n_A$ is the number of parameters from Group A, $N_{ROOM}$ is the total number of rooms, and $n_B$ is the number of parameters from Group B. Equation (1) indicates that the more rooms used in the system, the more parameters will increase. Therefore, it is essential to identify the critical parameters to detect all six conditions in FDD. Hence it can minimise the number of sensors used in a practical system and eventually reduce the cost. The values of standard deviation and accuracy were used to investigate the impact of these parameters to detect faults without compromising the performance of the classifiers.

### 2.3. Standard deviation

In statistics, a standard deviation is used as a measure of variation in the dataset. A low value of standard deviation represents the data is close to the mean value. In contrast, a high value indicates that the data has a broader range and is farther than its mean value. In this paper, the standard deviation can be used to identify which parameters have notably changed throughout the simulation. Thus, it can be used to analyse the impact of parameter selection in identifying the faults. Table 5 shows the standard deviation value of each parameter in the dataset. The subscript $x$ in parameter Group A denotes the room number, where $x = 1,2,\ldots, N_{ROOM}$.

Table 5. Standard deviation value for all parameters in the dataset

| Group | Parameters | Parameters measured |
|---|---|---|
| Group A | $V_{SX}$ | 10.03 |
| | $T_{SX}$ | 5.2 |
| | $T_{TCX}$ | 2.57 |
| Group B | $T_{CHWS}$ | 5.67 |
| | $T_{CWR}$ | 5.31 |
| | $T_{CWS}$ | 4.45 |
| | $T_{CHWR}$ | 4.03 |
| | $V_{CHWS}$ | 2.38 |
| | $V_{CHWR}$ | 1.83 |
| | $C_{CH}$ | 1.53 |
| | $V_{CWS}$ | 1.22 |

Table 5 shows that $V_S$ and $T_{CHWS}$ have the highest standard deviation value for each group. In contrast, $T_{TC}$ and $V_{CWS}$ have the lowest standard deviation values. It shows that parameters of $V_S$ and $T_{CHWS}$ significantly changed during simulations compared to $T_{TC}$ and $V_{CWS}$ data. Therefore, the higher value of standard deviation may represent a more significant impact on the fault simulations. There is also a possibility that the low value of standard deviation was less critical in fault classification and can be removed from the dataset.

The parameters selection for new datasets of Group A and Group B are described in Table 6 and Table 7. One parameter was eliminated for every dataset formed in both Table 6 and Table 7. The datasets were formed based on the standard deviation values shown in Table 5. Datasets in Group A and Group B were then combined one by one as,

$$Combination\ dataset = \{\ A1B1; A1B2; \ldots; A2B1; A2B2; \ldots; AnBm\}, \quad (2)$$

where $n$ is the number of datasets in Group A, and $m$ is the number of datasets in Group B. Each combination was tested and compared with all three machine learning classifiers in Table 3.

Table 6. The selection of parameters in Group A

| Dataset | List of parameters |
|---|---|
| Original | $V_{Sx}$, $T_{Sx}$, $T_{TCx}$ |
| A1 | $V_{Sx}$, $T_{Sx}$ |
| A2 | $T_{TCx}$, $V_{Sx}$ |
| A3 | $T_{TCx}$, $T_{Sx}$ |
| A4 | $V_{Sx}$ |
| A5 | $T_{Sx}$ |
| A6 | $T_{TCx}$ |

Table 7. The selection of parameters in Group B

| Dataset | List of parameters |
|---|---|
| Original | $T_{CHWS}$, $T_{CWR}$, $T_{CWS}$, $T_{CHWR}$, $V_{CHWS}$, $V_{CHWR}$, $C_{CH}$, $V_{CWS}$ |
| B1 | $T_{CHWS}$, $T_{CWR}$, $T_{CWS}$, $T_{CHWR}$, $V_{CHWS}$, $V_{CHWR}$, $C_{CH}$ |
| B2 | $T_{CHWS}$, $T_{CWR}$, $T_{CWS}$, $T_{CHWR}$, $V_{CHWS}$, $V_{CHWR}$ |
| B3 | $T_{CHWS}$, $T_{CWR}$, $T_{CWS}$, $T_{CHWR}$, $V_{CHWS}$ |
| B4 | $T_{CHWS}$, $T_{CWR}$, $T_{CWS}$, $T_{CHWR}$ |
| B5 | $T_{CHWS}$, $T_{CWR}$, $T_{CWS}$ |
| B6 | $T_{CHWS}$, $T_{CWR}$ |
| B7 | $T_{CHWS}$ |

### 2.4. Accuracy

The accuracy of the deep learning classifier was analysed when one of the parameters was removed from the dataset. The results represent the ability of the classifier to identify and classify the faults. Therefore, the higher accuracy obtained when a parameter was taken out from the dataset represents that the parameter does not impact the fault classification. However, should the accuracy decrease much when the parameter was eliminated from the dataset, the parameter significantly impacts the fault classification. The results were presented in Table 8, while Table 9 shows the parameters selection for new datasets of Group B. The datasets of Group A remain unchanged, as in Table 4. Similarly, each dataset's combination was tested and compared with three machine learning classifiers

Table 8. The accuracy of the classifier when each of these parameters was deleted from the original dataset

| Group | Parameters deleted | Parameters measured |
|---|---|---|
| Group A | $V_{SX}$ | 91.4% |
| | $T_{TCX}$ | 93.1% |
| | $T_{SX}$ | 94.0% |
| Group B | $T_{CWS}$ | 91.5% |
| | $T_{CHWS}$ | 93.7% |
| | $V_{CHWS}$ | 93.8% |
| | $T_{CWR}$ | 94.0% |
| | $C_{CH}$ | 94.0% |
| | $T_{CHWR}$ | 94.1% |
| | $V_{CWS}$ | 94.3% |
| | $V_{CHWR}$ | 94.5% |

Table 9. The selection of parameters in Group B

| Dataset | List of parameters |
|---|---|
| Original | $T_{CHWS}$, $T_{CWR}$, $T_{CWS}$, $T_{CHWR}$, $V_{CHWS}$, $V_{CHWR}$, $C_{CH}$, $V_{CWS}$ |
| B11 | $T_{CWS}$, $T_{CHWS}$, $V_{CHWS}$, $T_{CWR}$, $C_{CH}$, $T_{CHWR}$, $V_{CWS}$ |
| B12 | $T_{CWS}$, $T_{CHWS}$, $V_{CHWS}$, $T_{CWR}$, $C_{CH}$, $T_{CHWR}$ |
| B13 | $T_{CWS}$, $T_{CHWS}$, $V_{CHWS}$, $T_{CWR}$, $C_{CH}$ |
| B14 | $T_{CWS}$, $T_{CHWS}$, $V_{CHWS}$, $T_{CWR}$ |
| B15 | $T_{CWS}$, $T_{CHWS}$, $V_{CHWS}$, $C_{CH}$ |
| B16 | $T_{CWS}$, $T_{CHWS}$, $V_{CHWS}$ |
| B17 | $T_{CWS}$, $T_{CHWS}$ |
| B18 | $T_{CWS}$ |

## 3. RESULTS AND DISCUSSION

Table 10 shows the results of the best combination datasets formed using both methods discussed in the previous section. The combination was selected for the least parameters with a minimum of 90% accuracy. For instance, dataset A1B5 combined dataset A1 from Group A and dataset B5 from Group B. It was the best combination in dataset A1 with a minimum number of parameters that reached 90% accuracy. Dataset A5 was not listed because all combinations with any datasets Group B produced below 90% accuracy. The number of parameters required for each dataset was developed as in Equation (1). The first part of the equation represents the parameters from Group A, while the second part represents the parameters in Group B. Based on the equations, the number of sensors depends on the number of rooms in the system. The results show that datasets A4B3, A6B3, A4B16, and A6B16 required the least number of sensors when the number of rooms increased, as compared to others.

Table 10. Results for the best combination datasets formed

| Method | Dataset | Number of parameters required, $n_P$ | Number of sensors for | | | |
|---|---|---|---|---|---|---|
| | | | $N_{room}=1$ | $N_{room}=2$ | $N_{room}=3$ | $N_{room}=4$ |
| Standard deviation | A1B5 | $2N_{ROOM}+3$ | 5 | 7 | 9 | 11 |
| | A2B5 | $2N_{ROOM}+3$ | 5 | 7 | 9 | 11 |
| | A3B4 | $2N_{ROOM}+4$ | 6 | 8 | 10 | 12 |
| | A4B3 | $N_{ROOM}+5$ | 6 | 7 | 8 | 9 |
| | A6B3 | $N_{ROOM}+5$ | 6 | 7 | 8 | 9 |
| Accuracy | A1B16 | $2N_{ROOM}+3$ | 5 | 7 | 9 | 11 |
| | A2B17 | $2N_{ROOM}+2$ | 4 | 6 | 8 | 10 |
| | A3B16 | $2N_{ROOM}+3$ | 5 | 7 | 9 | 11 |
| | A4B16 | $N_{ROOM}+3$ | 4 | 5 | 6 | 7 |
| | A6B16 | $N_{ROOM}+3$ | 4 | 5 | 6 | 7 |

Based on the investigation results in Table 10, the datasets combination of Dataset A4 and A6 for standard deviation and accuracy selection methods were identified as the minimum number of required sensors. Table 11 compares the classification results from our previous study in [13] with the highlighted datasets in Table 10: A4B3, A6B3, A4B16, and A6B16. Three machine learning classifiers: deep learning, support vector machine (SVM), and multi-layer perceptron (MLP), were used to measure the accuracy of all five datasets. The accuracy of these newly combined datasets was a bit lesser than the original dataset in [13], around 0.6% - 3.4%. Nonetheless, the differences were not much and are still reliable.

Table 11. Comparison results between the original dataset, Dataset A4B3, A6B3, A4B16, and A6B16

| | | Original dataset [13] | Dataset A4B3 | Dataset A6B3 | Dataset A4B16 | Dataset A6B16 |
|---|---|---|---|---|---|---|
| Classification accuracy | Deep learning | 94% | 93.2% | 91.8% | 93.4% | 92.3% |
| | SVM | 97% | 94.6% | 94.3% | 94.3% | 93.6% |
| | MLP | 99.4% | 97.5% | 97.3% | 97.4% | 96.6% |
| Parameters | Group A | $V_{Sx}$<br>$T_{Sx}$<br>$T_{TCx}$ | $V_{Sx}$ | $T_{TCx}$ | $V_{Sx}$ | $T_{TCx}$ |
| | Group B | $T_{CHWS}$<br>$T_{CWR}$<br>$T_{CWS}$<br>$T_{CHWR}$<br>$V_{CHWS}$<br>$V_{CHWR}$<br>$C_{CH}$<br>$V_{CWS}$ | $T_{CHWS}$<br>$V_{CHWS}$<br>$T_{CWS}$<br>$T_{CHWR}$<br>$T_{CWR}$ | $T_{CHWS}$<br>$V_{CHWS}$<br>$T_{CWS}$<br>$T_{CHWR}$<br>$T_{CWR}$ | $T_{CHWS}$<br>$V_{CHWS}$<br>$T_{CWS}$ | $T_{CHWS}$<br>$V_{CHWS}$<br>$T_{CWS}$ |

The original dataset has three parameters Group A and eight parameters of Group B. In comparison, datasets A4B3 and A6B3 have one parameter of Group A and five parameters of Group B. Although Group A's parameter is different, the parameters of Group B are the same for both datasets. Similarly, it is the same case for datasets A4B16 and A6B16, where Group B has the same parameters for both datasets. For information, datasets A4B16 and A6B16 have one parameter from Group A and three parameters of Group B. Moreover, the parameters of Dataset B16 were part of the parameters of Dataset B3. It can be concluded that $T_{CHWS}$, $V_{CHWS}$, and $T_{CWS}$ were among the critical parameters in Group B to classify all six conditions. As for Group A, either $T_{TCx}$ or $V_{Sx}$ can be regarded as equally crucial for data-driven FDD because both datasets had almost similar accuracy. Mathematically, the list of proposed parameters for data-driven FDD for centralised chilled water air conditioning system can be written as follow,

$$Parameter, P = \{ P_{ROOM}, P_{CENTRAL}\} \quad (3)$$
where $P_{ROOM} = \{T_{TC,x} | x = 1,2, \dots, N_{ROOM}\} \; OR \; \{V_{S,x} | x = 1,2, \dots, N_{ROOM}\}$
and $P_{CENTRAL} = \{T_{CHWS}, V_{CHWS}, T_{CWS}\}$.

The minimum number of parameters required to identify six conditions, as described in Table 10 successfully, can be expressed as,

$$n_P = N_{ROOM} + 3, \quad (4)$$

where $N_{ROOM}$ represents the total number of rooms. The constant 3 indicates the three critical Group B parameters, which are $T_{CHWS}$, $V_{CHWS}$, $T_{CWS}$. The other parameter associated with the number of rooms is either $V_S$ or $T_{TC}$. In this research, two thermocouples were used to measure $T_{TC}$, the temperature of each room, while for the airflow sensor, model SD2001 from ifm electronic was used to measure $V_S$. The price of an airflow sensor is very much higher than the price of thermocouples. Therefore, in terms of cost, all four parameters in Dataset A6B16 can be considered the critical parameters to identify six classes of faults for this research at a lower cost than dataset A4B16. Although the accuracy of the Dataset A4B16 was slightly higher than the Dataset A6B16, the difference was not much and was still above 90%.

## 4. CONCLUSION

This paper has presented the developed lab-scaled of a centralised chilled water air-conditioning system to represent the actual system. It is a complete system with a cooling tower, chiller, AHU and two rooms. Six conditions had successfully simulated in the lab-scaled system and presented in our previous study. However, if a practical system is considered, the number of sensors required would be extensive as it depends on the number of rooms in the building. In other words, more cost is needed as the number of sensors is increased with the number of rooms. Therefore, this paper has proposed critical parameters for data-driven FDD of a centralised chilled water system. The impact of each parameter was identified and carefully analysed to maintain a good FDD accuracy. Four critical parameters were proposed in this paper: the rooms' temperature, $T_{TCx}$, supplied chilled water temperature, $T_{CHWS}$, supplied chilled water flow rate, $V_{CHWS}$, and supplied cooled water temperature, $T_{CWS}$. Results showed that the data-driven FDD successfully diagnosed all six conditions with the proposed parameters for more than 92.3% accuracy. Furthermore, the results were only differed by 0.6% - 3.4%, which was almost similar to our previous study. With the proposed parameters, only critical parameters to be installed in the actual building thus can reduce the sensors installation cost.

## ACKNOWLEDGEMENTS

The authors are grateful to Centre for Research and Innovation Management (CRIM), Universiti Teknikal Malaysia Melaka (UTeM), Universiti Teknologi Malaysia (UTM), and the Ministry of Higher Education (MOHE) for funding this research.

## REFERENCES

[1] M. F. Othman, H. Abdullah, N. A. Sulaiman, and M. Y. Hassan, "Performance evaluation of an actual building air-conditioning system," in *IOP Conference Series: Materials Science and Engineering*, 2013, vol. 50, no. 1, doi: 10.1088/1757-899X/50/1/012051.
[2] S. Wu and J. Q. Sun, "A top-down strategy with temporal and spatial partition for fault detection and diagnosis of building HVAC systems," *Energy Build.*, vol. 43, no. 9, pp. 2134–2139, 2011, doi: 10.1016/j.enbuild.2011.04.020.
[3] Z. Li *et al.*, "An Effective Fault Detection and Diagnosis Approach for Chiller System," *IFAC-PapersOnLine*, vol. 52, no. 10, pp. 55–60, 2019, doi: 10.1016/j.ifacol.2019.10.026.
[4] M. E. S. Trothe, H. R. Shaker, M. Jradi, and K. Arendt, "Fault isolability analysis and optimal sensor placement for fault diagnosis in smart buildings," *Energies*, vol. 12, no. 9, 2019, doi: 10.3390/en12091601.
[5] E. K. Alexandersen, M. R. Skydt, S. S. Engelsgaard, M. Bang, M. Jradi, and H. R. Shaker, "A stair-step probabilistic approach for automatic anomaly detection in building ventilation system operation," *Build. Environ.*, vol. 157, no. February, pp. 165–171, Jun. 2019, doi: 10.1016/j.buildenv.2019.04.036.
[6] C. G. Mattera, J. Quevedo, T. Escobet, H. R. Shaker, and M. Jradi, "A Method for Fault Detection and Diagnostics in Ventilation Units Using Virtual Sensors," *Sensors (Basel).*, vol. 18, no. 11, pp. 1–21, 2018, doi: 10.3390/s18113931.
[7] N. A. Sulaiman, M. F. Othman, and H. Abdullah, "Fuzzy logic control and fault detection in centralized chilled water system," *Proc. - 2015 IEEE Symp. Ser. Comput. Intell. SSCI 2015*, pp. 8–13, 2015, doi: 10.1109/SSCI.2015.265.
[8] N. A. Sulaiman, M. F. Othman, and H. Abdullah, "Fuzzy logic control of centralized chilled water system," *J. Teknol.*, vol. 72, no. 2, pp. 57–62, 2015, doi: 10.11113/jt.v72.3884.
[9] Y. Li and Z. O'Neill, "A critical review of fault modeling of HVAC systems in buildings," *Build. Simul.*, vol. 11, no. 5, pp. 953–975, 2018, doi: 10.1007/s12273-018-0458-4.
[10] B. Li, F. Cheng, X. Zhang, C. Cui, and W. Cai, "A Novel Semi-supervised Data-driven Method for Chiller Fault Diagnosis with Unlabeled Data," *Appl. Energy*, vol. 285, pp. 1–13, 2021, doi: 10.1016/j.apenergy.2021.116459.
[11] W. S. Yun, W. H. Hong, and H. Seo, "A data-driven fault detection and diagnosis scheme for air handling units in building HVAC

systems considering undefined states," *J. Build. Eng.*, vol. 35, 2021, doi: 10.1016/j.jobe.2020.102111.

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

[13] N. A. Sulaiman, P. Abdullah, H. Abdullah, M. N. S. Zainuddin, and A. Md Yusop, "Fault detection for air conditioning system using machine learning," *IAES Int. J. Artif. Intell.*, vol. 9, no. 1, pp. 109–116, 2020, doi: 10.11591/ijai.v9.i1.pp109-116.

[14] Y. Fan, X. Cui, H. Han, and H. Lu, "Chiller fault diagnosis with field sensors using the technology of imbalanced data," *Appl. Therm. Eng.*, vol. 159, no. June, 2019, doi: 10.1016/j.applthermaleng.2019.113933.

[15] A. K. Abbas, S. Hamad, and N. A. Hamad, "Single line to ground fault detection and location in medium voltage distribution system network based on neural network," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 2, pp. 621–632, 2021, doi: 10.11591/ijeecs.v23.i2.pp621-632.

[16] A. T. Radhi, W. H. Zayer, and A. M. Dakhil, "Classification and direction discrimination of faults in transmission lines using 1D convolutional neural networks," *Int. J. Power Electron. Drive Syst.*, vol. 12, no. 3, 2021, doi: 10.11591/ijpeds.v12.i3.pp1928-1939.

[17] N. F. Fadzail, S. M. Zali, M. A. Khairudin, and N. H. Hanafi, "Stator winding fault detection of induction generator based wind turbine using ANN," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 1, pp. 126–133, 2020, doi: 10.11591/ijeecs.v19.i1.pp126-133.

[18] N. F. Fadzail and S. M. Zali, "Fault detection and classification in wind turbine by using artificial neural network," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 3, pp. 1687–1693, 2019, doi: 10.11591/ijpeds.v10.i3.pp1687-1693.

[19] A. Kaur, Y. S. Brar, G. Leena, E. Engineering, and A. Info, "Fault detection in power transformers using random neural networks," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 1, pp. 78–84, 2019, doi: 10.11591/ijece.v9i1.pp78-84.

[20] R. Jyothi, T. Holla, U. R. K, and R. Jayapal, "Machine learning based multi class fault diagnosis tool for voltage source inverter driven induction motor," *Int. J. Power Electron. Drive Syst.*, vol. 12, no. 2, pp. 1205–1215, 2021, doi: 10.11591/ijpeds.v12.i2.pp1205-1215.

[21] S. L. Souad, B. Azzedine, and S. Meradi, "Fault diagnosis of rolling element bearings using artificial neural network," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 5, pp. 5288–5295, 2020, doi: 10.11591/ijece.v10i5.pp5288-5295.

[22] X. J. Luo, K. F. Fong, Y. J. Sun, and M. K. H. Leung, "Development of clustering-based sensor fault detection and diagnosis strategy for chilled water system," *Energy Build.*, vol. 186, pp. 17–36, 2019, doi: 10.1016/j.enbuild.2019.01.006.

[23] K. Yan, J. Huang, W. Shen, and Z. Ji, "Unsupervised learning for fault detection and diagnosis of air handling units," *Energy Build.*, vol. 210, p. 109689, 2020, doi: 10.1016/j.enbuild.2019.109689.

[24] C. G. Mattera, M. Jradi, M. R. Skydt, S. S. Engelsgaard, and H. R. Shaker, "Fault detection in ventilation units using dynamic energy performance models," *J. Build. Eng.*, vol. 32, 2020, doi: 10.1016/j.jobe.2020.101635.

[25] C. G. Mattera, J. Quevedo, T. Escobet, H. R. Shaker, and M. Jradi, "Fault Detection and Diagnostics in Ventilation Units Using Linear Regression Virtual Sensors," in *International Symposium on Advanced Electrical and Communication Technologies, ISAECT 2018 - Proceedings*, 2019, pp. 1–6, doi: 10.1109/ISAECT.2018.8618755.

[26] S. Deshmukh, S. Samouhos, L. Glicksman, and L. Norford, "Fault detection in commercial building VAV AHU: A case study of an academic building," *Energy Build.*, vol. 201, pp. 163–173, 2019, doi: 10.1016/j.enbuild.2019.06.051.

[27] J. Li, Y. Guo, J. Wall, and S. West, "Support vector machine based fault detection and diagnosis for HVAC systems," *Int. J. Intell. Syst. Technol. Appl.*, vol. 18, no. 1–2, pp. 204–222, 2019, doi: 10.1504/IJISTA.2019.097752.

[28] Y. Chen, "Data-driven Whole Building Fault Detection and Diagnosis," Drexel University, Philadelphia, Pennsylvania, 2019.

[29] N. A. Sulaiman, K. W. Chuink, M. N. S. Zainudin, A. M. Yusop, S. F. Sulaiman, and M. P. Abdullah, "Data-driven fault detection and diagnosis for centralised chilled water air conditioning system," *Prz. Elektrotechniczny*, vol. 2022, no. 1, 2022.

[30] S. Heo and J. H. Lee, "Fault detection and classification using artificial neural networks," in *IFAC PapersOnLine*, 2018, pp. 470–475, doi: 10.1016/j.ifacol.2018.09.380.

[31] K. Yan, Z. Ji, and W. Shen, "Online fault detection methods for chillers combining extended kalman filter and recursive one-class SVM," *Neurocomputing*, vol. 228, pp. 205–212, 2017, doi: 10.1016/j.neucom.2016.09.076.

[32] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," *Procedia Comput. Sci.*, vol. 83, no. Fams, pp. 1064–1069, 2016, doi: 10.1016/j.procs.2016.04.224.

[33] A. Bharat, N. Pooja, and R. A. Reddy, "Using Machine Learning algorithms for breast cancer risk prediction and diagnosis," *2018 IEEE 3rd Int. Conf. Circuits, Control. Commun. Comput. I4C 2018*, no. x, pp. 1–4, 2018, doi: 10.1109/CIMCA.2018.8739696.

[34] I. Lorencin, N. Anđelić, J. Španjol, and Z. Car, "Using multi-layer perceptron with Laplacian edge detector for bladder cancer diagnosis," *Artif. Intell. Med.*, vol. 102, no. May 2019, 2020, doi: 10.1016/j.artmed.2019.101746.

[35] S. B. Segota, N. Anđelić, J. Kudláček, and R. Čep, "Artificial neural network for predicting values of residuary resistance per unit weight of displacement," *Pomor. Zb. - J. Marit. Transp. Sci.*, vol. 57, pp. 9–22, 2019, doi: 10.18048/2019.57.01.