



# Article Knowledge Discovery by Analyzing the State of the Art of Data-Driven Fault Detection and Diagnostics of Building HVAC

Arash Hosseini Gourabpasi and Mazdak Nik-Bakht\*

Compleccity Lab, Department of Building, Civil and Environmental Engineering (BCEE), Concordia University, Montreal, QC H3G 1M8, Canada; arash.hosseinigourabpasi@mail.concordia.ca

\* Correspondence: mazdak.nikbakht@concordia.ca

Abstract: The automated fault detection and diagnostics (AFDD) of heating, ventilation, and air conditioning (HVAC) using data mining and machine learning models have recently received substantial attention from researchers and practitioners. Various models have been developed over the years for AFDD of complete HVAC or its sub-systems. However, HVAC complexities, which partly have roots in its close coupling nature and interrelated dependencies, mean that understanding the relationship between faults and the suitability of the techniques remains an unanswered question. The literature analysis and interactive visualization of the data collected from the past implementation of AFDD models can provide useful insight to further explore this question by applying artificial intelligence (AI). Association rule mining (ARM) is deployed by this paper, using the frequent pattern (FP) growth algorithm to generate frequent fault sets for most common HVAC faults from the body of AFDD models developed in the literature to represent the status quo. A new model is developed for common HVAC faults and the techniques most frequently used to detect and diagnose them. A recommender system is developed using the ARM model to extract knowledge from the body of knowledge of HVAC data-driven AFDD in the form of rule-sets that reflect the associations. Findings of this review paper can significantly help civil and building engineers, as well as facility managers, in better management of building HVAC systems.

Keywords: data mining; AFDD; HVAC; machine learning; association rule mining; FP-Growth

# 1. Introduction

The heating, ventilation, air conditioning (HVAC), and refrigeration systems are arguably use up the most energy out of all a building's physical assets. HVAC/R systems regulate the temperature, humidity, quality, and air movement in buildings, making them critical for occupant comfort, health, and productivity. In Canadian commercial stores, HVAC and lighting combined contribute to 90% of energy consumption [1]. In 2011, heating systems, particularly furnaces (57%), followed by electric baseboards (27%) and boilers (5%), were the primary type of heating system used by Canadian households [2]. This energy consumption indicates the dependency of Canadian households and commercial buildings on the HVAC system, and hence emphasizes the importance of timely and accurate identification of its faults.

Performance in HVAC systems and sub-systems are negatively affected by system degradation, operational misuse, reduced maintenance, and sensor issues [3,4]. Many HVAC faults that require repair or immediate attention go unnoticed and cause progressive damages. The most common components where faults occur are the damper, fan, filter, and other parts such as sensors [5]. Furthermore, faults in HVAC systems affect the HVAC's energy consumption. For example, when refrigerant charge is less than 25% of the design value, it can reduce the energy efficiency by 15%. Moreover, 20% capacity loss is also reported in such situations [6]. The reasons mentioned above can lead to



Citation: Hosseini Gourabpasi, A.; Nik-Bakht, M. Knowledge Discovery by Analyzing the State of the Art of Data-Driven Fault Detection and Diagnostics of Building HVAC. CivilEng 2021, 2, 986-1008. https:// doi.org/10.3390/civileng2040053

Academic Editors: George D. Hatzigeorgiou, João Castro-Gomes, Cristina Fael, Miguel Nepomuceno and Akanshu Sharma

Received: 23 August 2021 Accepted: 4 November 2021 Published: 10 November 2021

Publisher's Note: MDPI stavs neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

increased energy usage in addition to user discomfort, shorter equipment life, and less reliability [4]. Malfunctioning sensors, components, and control systems and degrading systems in HVAC and lighting systems are the main reasons for energy wastage and an unsatisfactory indoor environment [7].

Fault detection and diagnostics of the HVAC system allow the asset manager to isolate and locate the faults in a timely manner. Current advancements in the Internet of Things (IoT) have led to the application of big data for creating automated fault detection and diagnostic (AFDD) models, which can be developed using machine learning (ML) techniques. The sensory data available in building automation systems (BAS) and building management systems (BMS) are used to detect the HVAC's faults and perform diagnostics. In asset management of buildings, energy management and maintenance models differ in scope and structure. While models for energy management describe continuous states (energy, temperature, etc.) and usually assume the HVAC to be in a healthy condition, the models used for maintenance do not consider human factors such as comfort and only describe discrete states, such as faulty/non-faulty states of equipment and fault typology [8].

HVAC faults can be categorized in at least three ways. The first is based on the 'cause of the fault'. A fault can have a natural cause or be human-made. The second classification is according to the 'fault's extent', which categorizes a fault into soft fault and sudden fault, also known as hard fault [9,10]. Hard faults can cause a system to stop working. A soft fault, however, causes performance degradation [11]. 'Sensor faults' are inevitable in HVAC [12] and have been subject to extensive research. Hence, they are considered the third category consisting of component faults such as actuators, sensors, and feedback controllers sensor fault [9,13]. Sensor faults can be further classified into bias, drifting, precision degradation, and complete failure [14].

Generally, the AFDD procedure (also referred to as data-driven FDD) for HVAC can be considered a multi-class classification problem. It uses/identifies relationships between data patterns (usually collected from BAS or BMS) and fault classes in the modeling process [9]. Fault detection can be defined as the process in which faulty operation is identified and classified from normal operations. Fault detection consists of two steps: (i) training and (ii) fault detection, also known as deployment. On the other hand, fault diagnosis is the process that looks for the causality of the identified fault. Fault detection can be performed independently, but generally, fault diagnosis is followed after a fault detection step [15]. FDD techniques can help detect and locate the HVAC system faults and mainly depend on having an accurate reference model and a sensitive fault detection and identification method [16].

AFDD can be performed in both offline and online manners. Offline AFDD needs a large amount of historical data and usually yields high accuracy rates due to the possibility of sufficient training. However, in online AFDD, the aim is to find new faults without necessarily being trained on the faulty sample [17]. Both offline and online AFDD widely take advantage of machine learning in supervised and unsupervised learning formats (usually respectively). In supervised learning-based techniques, online fault detection is made possible by offline training and online fault detection. Both methods can be used for fault detection purposes, as faults can be considered outliers or novelties. Unsupervised methods are particularly suitable for imbalanced datasets and when labeled data are unavailable [18].

While there has been a rich body of knowledge on the applications of ML in AFDD, the studies to date have not investigated the status quo of the relationships between fault classes and/or between the detection and diagnostics techniques used to identify them in the literature. The occurrence of one or more faults in an HVAC system can be associated with the appearance of other faults in the system. Revealing such interrelationships among faults can allow the asset managers to expect the event of the occurrence of one or more faults. Additionally, certain AFDD algorithms are used in the literature to detect specific fault types in particular HVAC systems. Providing a comprehensive overview of this relationship among fault types and

detection methods can allow the asset managers to better select proper algorithms(s) to identify faults of interest. Accordingly, this paper implements a comprehensive analysis of the AFDD models, reported to date for HVAC asset management and maintenance. It analyses associations between the most common HVAC fault types and then scrutinizes the affinity between AFDD techniques and the fault types they identify/diagnose, as per the best practices from the literature. The latter will facilitate a more in-depth understanding of relationships between the fault types and the suitability of the AFDD techniques.

In the upcoming sections of this paper, a holistic review of FDD and, in particular, datadriven FDD models is performed. Further, the applicable machine learning FDD algorithms utilized by the literature are reviewed, followed by the methodology implemented in this study to allow literature analysis and visualization of data to create a recommender system for the database under investigation and its validation process.

## 2. Literature Review

At a holistic level, HVAC can be studied at a system or local level [19]. Local-level classification can be divided into (i) sub-system level and (ii) equipment/component level [20], which we refer to as 'HVAC levels' in this paper. The full HVAC system consists of the sub-systems and/or components coupled together. In the past two decades, fault detection has been mainly applied to the HVAC at the sub-system level, and very few researchers have looked at detecting faults at the whole building level [7]. Hereafter in this paper, we use the term HVAC system in a general sense, by which we also refer to sub-systems and pieces of equipment in HVAC. System-level faults refer to the occurrence of a fault in one sub-system or equipment and its consequence at the system level [7].

Previous literature reviews in the domain of FDD have focused on overall FDD modeling methods [21–25] or data-driven methods [26]. Another group of review studies focuses on a specific step of the procedure, such as algorithms [27] or fault types [28]. However, the current paper is different in the way that it analyzes the models developed in the literature by looking at the features used, fault types identified, corresponding HVAC systems, and algorithms used for data-driven FDD models. Through an affinity analysis of these studies, we extract knowledge in the form of association rules and deploy them in the form of a recommender system. The scope of this study and the recommender system is mainly commercial buildings since the majority of AFDD models developed in the literature have been of this type.

#### 2.1. FDD Approaches and Techniques

Fault detection generally refers to the process that discovers any faulty operation and separates it from normal operations, and fault diagnosis refers to identifying the cause of the faulty operation [15]. For, e.g., a chiller, examples of system-level faults are refrigerant leak/undercharge, refrigerant overcharge, and excess oil. Sub-system-level faults can be condenser fouling, reduced condenser water flow, and non-condensable in refrigerant and reduced evaporator water flow [15]. While these are only some examples, a more complete list of the faults can be seen in Table 1. An accurate model and an appropriate threshold acre the critical factors in fault detection [3]. Common classifications of AFDD models found in the literature are shown in Figure 1 and introduced in the following paragraphs.

One classification for AFDD methods is the top-down versus bottom-up approach. The top-down approach detects faults that manifest themselves at the whole building level, whereas the bottom-up approach focuses on the component or sub-system level. In both approaches, models of ideal operation conditions are compared with actual measurements to detect faulty or abnormal behavior [29]. Whole-building fault detection usually makes use of a top-down fault detection strategy [7]; the top-down approach is comparatively more difficult than the bottom-up. In the top-down approach, further analysis is required to locate faults because of the system-level effect that causes the faults' symptoms to spread across the system [15,30].

Rank	Fault Category	Count
1	Limit issue	68
2	Stuck/Partially closed	67
3	Flow problems	54
4	Bias/Drift/Calibration	49
5	Leakage	41
6	Foul	38
7	Other faults	20
8	Non-functioning	20
9	Non-condensable	18
10	Control	18
11	Temperature issue	12
12	Speed	12
13	Set point	8
14	Performance	8
15	Capacity reduction	5
16	Blockage	4
17	Schedule	3
18	Sizing issue	3



Figure 1. Holistic FDD and data-driven modeling approaches.

Based on ASHRAE's recommendations [31] as found in [32], the two main modeling methods are the forward (classical) approach and the data-driven (inverse) approach. The forward approach is known as white box/engineering methods. Forward approaches usually require detailed knowledge of various system processes and interactions. Most simulation software tools use such approaches. The data-driven and model-based classifications have been found to be the most common FDD classification approaches [33–36],

which have also been referred to as model-free methods and model-based methods [37] in the literature.

The other common classification found in the literature categorizes the AFDD techniques into model-based methods, rule-based methods, and data-driven methods [5,38]. The data-driven methods are also called process history-based [30,39]. In some cases, knowledge-based is also included in data-driven models [34]. Further, model-based classification is also referred to as quantitative [39,40]. The other classifications found are analytical model-based, signal-based, and data-driven methods [11]. On the other hand, the data-driven models are classified by ASHRAE to calibrated simulation models, greybox models, and black-box models/empirical approaches. In simple terms, calibrated simulation is similar to forward approaches and requires detailed knowledge of the system and processes, but black-box models are data-driven and use statistical or artificial intelligence approaches to develop models. Grey-box models, on the other hand, are formulated using training data and physical principles [32].

Another classification of AFDD techniques considers the nature of the data analysis model used and provides two categories of multivariate statistical analysis and artificial intelligence (AI) methods. While principal component analysis (PCA) is the classical example of multivariate statistical analysis, artificial neural network (ANN) is a typical example of AI-based methods [41].

In summary, the above classifications show how terminology tends to vary among researchers and practitioners. Care should be taken as previous studies have used more than one classification scheme for the available methods. There exists no hard classification, and several AFDD models might lie in an intersection of various classes. Hence, the modeling approaches are not necessarily mutually exclusive and often can be a combination of one or more methods. The present study is dedicated to knowledge discovery from data-driven methods, i.e., bottom-up approaches and algorithms used by them. In the process of selecting relevant studies, we used ASHRAE's recommendations [31] to verify whether a model is data-driven. This was to overcome the literature discrepancies regarding the classification and categories of FDD methods.

# 2.2. Data-Driven FDD Algorithms Based on Machine Learning Approach

The AFDD techniques reviewed in the literature are broadly grouped and categorized into supervised and unsupervised learning. This study also covers more general algorithms, such as Bayesian network (BN) and ARM algorithms, which may not traditionally fit in any of these two broad categories. Most of the reviewed studies implementing AFDD are supervised methods and treat the FDD as essentially a classification problem. Unsupervised methods are mainly adopted in the pre-processing phase or are used for fault detection through clustering.

Figure 2 shows the machine learning algorithms for FDD based on learning type. SVM (support vector machine), decision tree, and regression methods are grouped into supervised, and dimensionality reduction techniques, instance-based classification and clustering belong to the unsupervised category. However, ANN/deep learning, ensemble learning, Bayesian networks, and ARM in the literature have used both supervised and unsupervised methods. Bayesian methods are used where event information is required to be included in the models. The events describe the states of discrete or continuous variables, such as a room being occupied or not by its occupants or considering the HVAC operation schedule, respectively. In hybrid methods, the machine learning approach for fault detection and diagnostics are different from one another. The algorithms are defined as they have appeared in the literature and are explained below.

AFDD, based on supervised learning, approaches the problem as a classification task. SVM is the most common FDD method of this type, used at the sub-system level. It is used for binary classification problems to separate faulty from normal data. This method finds a hyperplane in the high-dimensional feature space to separate the two types of data. The other variant, multi-class SVM, is used when more than two fault classes

exist [42]. Regression methods used in the AFDD literature for HVAC are commonly used in combination with other modeling techniques for fault detection purposes. Regression is often coupled with other methods and uses thresholds to determine faults in HVAC. Instance-based classification compares new problem instances with instances seen in the training set. K-nearest neighbor (k-NN) is a typical example of an instance-based classifier. Instance-based classification is the least-used method in the literature and has been used to identify a single type of fault or binary faults in the HVAC system.



Figure 2. Machine learning FDD algorithms based on learning type.

Unsupervised learning, including dimensionality reduction, instance-based classification, and clustering, are used when the classes of faults are not known. Dimensionality reduction methods such as PCA, Fisher discriminant analysis (FDA), and linear discriminant analysis (LDA) are used in this category. In PCA, which is the most commonly used unsupervised learning method, raw data are decomposed into two subspaces: the principal component subspace and residual subspace. The principal component data capture the main normal statistical correlations, and the residual data quantify the main variances to detect faults [36]. In clustering, two or more clusters are identified that separate normal from faulty data.

While most applications of the deep learning methods in the energy field are in load and power prediction, ANN/DL (deep learning) has also been extensively used for FDD of HVAC. On the other hand, ensemble methods utilize multiple learning algorithms to achieve better predictive performance than the same algorithms separately. Most of the algorithms in ensemble learning can be considered weak learners. Therefore, ensemble learning integrates multiple weak learners to create an improved FDD method.

Bayesian networks (BN) use graph theory and probability theory to perform data analysis and are preferred for their reasoning ability. In the reviewed literature, it was evident that BN is mostly used in system-level studies to detect and diagnose HVAC faults and support literature analysis results found for the type of HVAC used [43]. BN's distinct advantage is the ability to identify the causes and sort them from the most to the least probable, which can be used to prioritize inspection and maintenance [39]. Association rule mining (ARM) intends to identify a set of latent associations among attributes of a dataset and express them in the form of if-then rules. Both unsupervised algorithms and supervised learning algorithms are used in the literature for this purpose.

Last but not least, there are studies in the literature which utilize different machine learning algorithms rather for detection than the diagnosis of faults, where for fault detection, the algorithm can be either supervised or unsupervised, and for fault diagnosis, it can utilize the same or a different algorithm which can be either supervised or unsupervised. In the present paper, we classify such studies as 'hybrid methods'. It is worth mentioning that in the AFDD literature, the term hybrid is used for a variety of purposes, such as for the algorithms combining both supervised and unsupervised learning [26] and must not be confused with what are referred to as hybrid methods in this paper.

#### 3. Methodology of the Study

The high-level methodology of this study is as illustrated in Figure 3 and consists of a systematic approach for collection, analysis and synthesis of academic studies related to AFDD. The scoped literature included 'data-driven' studies (as defined earlier) between the years 2015 and 2021. Major knowledge repositories, including Scopus, Web of Science, and Google scholar were targeted for data collection. The criteria for initial screening of studies to be included in the analysis were to have indicated (i) a complete list of features being used, (ii) the faults being considered, and (iii) data-driven fault detection or both fault detection and diagnostics techniques being developed. Focusing on both features and analysis methods is instrumental for this research, since the aim is to provide the big picture of data requirements and analysis for AFDD of the HVAC systems.



Figure 3. High-level methodology used for data collection, analysis, and knowledge discovery.

A total of 109 papers were initially collected. Further, these studies were reviewed, and the papers offering a complete FD (fault detection) model or FDD model were targeted. Accordingly, a total of 82 studies were selected for analysis, which is listed in Table A1 in the

Appendix A section of this paper. The machine learning algorithms used in pre-processing and post-processing stages were excluded, and only those data-driven techniques that have application during the fault detection and diagnosis of HVAC were considered. Other supplementary information collected from each study includes the HVAC type, source of data for the AFDD process, and the data collection frequency. The sources of data include both synthetic and real data.

Initially, the features and faults are gathered for the respective HVAC systems investigated, and further, the analysis outcomes are visualized in the form of an interactive Sankey diagram. Further, two separate models are trained to extract knowledge in the form of association rules. The first model is developed specifically for common HVAC faults reported in the literature. The second model includes the common HVAC faults and the data-driven techniques used to detect them. The unsupervised machine learning technique, association rule mining, is adopted through the Frequent Pattern (FP)-Growth algorithm to find the frequent itemsets and associations among them. Finally, the rules derived are validated using experts' opinions through an online survey questionnaire.

# 4. Data Analysis and Knowledge Discovery for HVAC System AFDD

In the data analysis stage of this study, firstly, the inputs used for AFDD of building HVAC were investigated and analyzed to understand the importance of various features in fault detection. Then, the common faults associated with these features were investigated and classified in accordance with the different HVAC levels (i.e., system, sub-system, and component/equipment) that they correspond to. Furthermore, different algorithms used for identifying and diagnosing the faults were studied and categorized within the big picture of the current state of practice in AFDD. Finally, knowledge was extracted by implementing and developing machine learning-based models that could indicate associations among HVAC faults and detection/diagnosis methods reported in the literature.

#### 4.1. Feature Analysis

Figure 4 summarizes the features used in the analyzed literature for AFDD. They are ranked based on their frequency of use in the HVAC FDD models reported in the literature. It is evident that 'temperature' is the single most crucial feature used for AFDD, as its application also extends to the second most frequently used feature, i.e., the 'calculated measure'. This feature commonly uses arithmetic operations such as subtraction and often uses features such as 'temperature' or 'pressure' as the calculation component; for example, the calculated measure is used to show the temperature difference between the supply and return air or pressure difference between the entrance and the exit (inlet and outlet) to indicate pressure drop or increase. Other frequently used parameters include the 'pressure' and 'flow rate'. State-representative information and energy-related parameters such as 'Opening/position', which represents physical characteristics such as position or percentage of a valve being open or closed, and 'Load' and 'Energy' categories are among other attributes frequently used by AFDD models.

## 4.2. Fault Analysis

Before initiating the AFDD analysis, it is essential to identify the fault types. Several fault classification systems exist in the literature, such as [10,44–46]; however, they cannot be used in this study. Some classifications are specific to a particular sub-system, such as chillers [10], or, if they cover the whole HVAC system [44–46], they are too detailed and elaborate and cannot support the abstraction required for rule mining. Accordingly, in this study, eighteen (18) fault categories were created and introduced to solely organize, categorize and analyze more than 400 faults reported in the literature investigated for HVAC's most common faults detected using data-driven methods in this paper. It must be noticed that the categories shown in Table 1 are not meant to provide a comprehensive classification of all fault types. The faults considered apply to the HVAC system, sub-system, and/or components. The faults are categorized based on the following procedure.

The categories are created using a hypernym keyword. The faults are hyponym and belong to only one of the eighteen hypernyms created. Then, logical reasoning is performed to assign each fault to the category that it best represents. In cases where a hyponym consists of more than one word in its description, the first word will be selected, and the assignment is carried out based on that word. For example, for the fault type referred to as 'control unstable', the term 'control' is considered the primary word, and 'unstable' is a condition associated with controlling. Hence, the fault is assigned to the 'control' category. The only exception applies to faults that include bias/drift. In particular, for sensor faults, we skip the sensor type, even if it is the first word of the fault description, and look at the following term in the description.

1 Temperature (395)	2 Calculated	5 Energy/ power	9 Frequency (18)	14 Load (7)
	(52)	(28) 6 Other	10 Humidity (14)	15 Voltage (6)
	3 Pressure (46)	(24) 7 Opening/	11 Control (11)	16 Solar/radiation (5)
	4	position (22) 8	12 Speed/velocity (10)	17 Set point/ set value (3)
	Flow rate (33)	Electric currents (19)	13 Efficiency (10)	18 Rate(cooling) (3)

**Figure 4.** Common features selected for AFDD of HVAC (18 categories formed for 706 features used in the analyzed literature and the numbers in brackets represent the frequency of occurrence in the literature).

The categories of the faults are sorted in Table 1 in descending order of occurrence frequency in our database. The 'Limit issue', which is the dominant category, comprises faults related to over/undercharge, excess oil, or reduced evaporator. The second category, 'stuck/partially closed', includes faults such as exhausted air, damper stuck (fully open), or cooling coil valve partially closed (15% open). The other categories' names such as 'temperature issue', 'blockage', 'speed', and 'non-functioning' are self-explanatory. 'Flow problems' and sensor-related faults, which are categorized as 'bias/drift/calibration' alongside 'leakage' and 'foul'-related faults, comprise the top six frequent categories of HVAC faults. The 'other faults' comprises of different types of faults that did not form a category due to limited appearance in the database. The fault categories such as 'set point' and 'non-condensable' belong to a particular type of fault, and on the other hand, fault categories such as 'control' and 'performance' belong to a more diverse pool of faults pertaining to their respective categories. It is evident that most studies have relied purely on sensory data and very few categories with a small occurrence in our database represent faults that may be detected given static information such as 'schedule' and 'sizing issue', which are categories with the lowest counts in the table.

#### 4.3. Analysis of Data-Driven FDD Algorithms

The AFDD algorithms utilized by the studies are as shown in Figure 5, along with their frequency of occurrence in our database. It has to be noted that this distribution is only reflective of ML-based models reported by the selected papers, and a large majority of

other types of algorithms (e.g., rule-based expert systems that are common in commercial settings) are outside the scope, and hence are absent from the picture. The most commonly adopted data-driven algorithms used for FDD are SVM, neural networks/deep learning and dimensionality reduction techniques. On the other hand, algorithms such as regression, clustering, instance-based classification, and ensemble learning methods are the least-utilized algorithms for FD/FDD of HVAC systems. One possible explanation of the lesser extent of adoption of other FDD algorithms can be that presently most researchers consider FDD of HVAC a classification problem, and hence lesser research is invested in exploring other methods.



Figure 5. The extent of data-driven techniques used for FDD of HVAC, based on the reviewed papers.

#### 4.4. Analysis of HVAC's Most Common Faults Detected through AFDD

Data collected for each category (features, faults, and FDD Techniques) were initially analyzed separately to determine the current status of the AFDD for different HVAC levels. The Sankey diagram shown in Figure 6 depicts the relationship between HVAC levels, FDD techniques, and the faults associated with them. An online version of the diagram is available in [47] and can be used interactively to precisely show the weight of each node. Weights are representative of the count for the specific node (i.e., the number of times they have appeared in our database). As seen from the Sankey diagram, while the majority of the present works are focused on the sub-system level, the HVAC system, the device, component, or part levels are least investigated. Hence, more studies are needed to understand the effectiveness of AFDD at a component level. The recent trends of studies indicate algorithms such as SVM and ANN have been used more than other methods, including dimensionality reduction techniques such as PCA [27]. SVM is almost exclusively adopted for sub-system HVAC fault detection, whereas ANN and dimensionality reduction methods are split equally for HVAC system and sub-system FDD. On the other hand, Bayesian networks and decision trees have mostly been utilized at the whole system level.



Fault Type

Figure 6. Relationship between system level, fault types, and FDD techniques used in the literature.

Certain methods are utilized more often for specific fault types. For example, from a total of 68 'limit issue' faults found in the literature, SVM (18 times), ANN (13), Bayesian networks (11), and ARM (11 times) have been used frequently to detect this type of fault. Nevertheless, for faults such as 'stuck/partially closed' and 'flow problems', neural networks and deep learning appear to be the most utilized methods. On the other hand, the clustering method has been mostly used for sensor-related faults such as 'bias/drift' and less frequently for 'non-functioning' and never been used for 'limit issue' faults. The other approach for analyzing FDD techniques is to assess each method for the faults detected individually. It can be seen that Bayesian networks, hybrid methods, and ARM methods appear to be able to detect all of the six most common faults in the HVAC even though these methods are comparatively used to a lesser extent.

#### 4.5. Knowledge Discovery through Machine Learning

In order to understand the latent relationships and associations between the common HVAC faults and/or AFDD techniques, association rule mining (ARM) has been used. ARM is an unsupervised machine learning procedure in which the aim is to observe the frequently occurring patterns, correlations, and associations in a dataset. Association mining is performed in two steps. The first step is to generate 'frequent itemsets'. The second is generating rules, where rules are generated and filtered based on set constraints. Two models were trained in this study: one for detecting affinity between various fault types and a second model to investigate the association between the FDD techniques and the HVAC's most common faults.

The FP-growth algorithm is an improved affinity analysis algorithm, in which the number of scans of the database is reduced to find the frequent itemsets [48]. In this study, FP-growth was implemented in the model to generate frequent itemsets of fault types and then extract relationships of a high level of *support* and *confidence* as rules. The rules take the form of a 'premise', followed by a 'conclusion'. The metrics considered in the model development are *support* and *confidence*, where *confidence* is used as a measure of the strength of the rule and *support* correlates to statistical significance. The equation for *support* of a rule and *confidence* of a rule are as shown below:

$$\operatorname{Rule}: (X \to Y) \tag{1}$$

(37.30)

$$Support(X \to Y) = Pr(X, Y) = \frac{n(X, Y)}{N}$$
(2)

Confidence
$$(X \to Y) = Pr(Y|X) = \frac{Pr(X,Y)}{Pr(X)}$$
 (3)

where *X* and *Y* are independent items or itemsets, *n* is the relative frequency of occurrence and *N* is the total transaction numbers.

Minimum *support* and minimum *confidence* are needed to eliminate the unimportant association rules [48,49]. Syntactic constraints were enforced for the second model to add restrictions on items that can be included in the rule. The developed model using Python and the dataset used for the analysis can be accessed online through Github (please see the 'data availability statement' at the end of the paper).

#### 4.5.1. Model 1: Common HVAC Faults

The frequent itemsets are created using the FP-growth algorithm, which has been assigned minimum support of 20% for the frequent itemsets and minimum confidence of 70% for detecting the association rules. By applying these criteria, five frequent fault types and 13 rules are identified through the FP-growth algorithm as shown in Figure 7. On the left-hand side of each rule are the premises and, on the right, after, the arrow is the conclusion. For example, rule #5 indicates that "if 'limit issue' fault and 'foul' fault are found simultaneously in the HVAC system for the designed FDD algorithm found in the literature, then it is likely that the system is also designed to detect 'flow problem'-related issues", with 77.3% confidence. The rules mined have either one or two fault categories in their premises. In seven of the rules that have one fault category in their premises, rule #9 and rule #10 have equal support of 22%, and confidence of 100%, which indicates "if 'non-condensables' fault occurs then there are equal chances that 'foul' and 'flow problems' related issues can be existing *separately*" or as per rule #11 the 'foul' and 'flow problems' can appear simultaneously. The first rule mined indicates that when 'flow problems' are found using the particular FDD algorithm, then it is likely that the FDD algorithm can detect 'leakage' in the HVAC system considered, which shows the correlation among these two fault categories in the database recorded. The other six rules mined show how the faults in the HVAC can be interrelated as they have two premises. Rules #12 and #13 have the highest confidence and represent how different combinations of fault in their premises and respective conclusion can be indicative of the correlation between 'foul', 'non-condensable', and 'flow problems' fault

categories in the given database. Model 1's mined rules only indicate correlations found in the database for the select FDD algorithms specifically designed to detect the faults investigated and cannot be used to investigate causality or indicate that FDD algorithms were designed to detect faults simultaneously.



```
2- [Foul] --> [Limit issue] (confidence: 0.710)
3- [Leakage] --> [Foul] (confidence: 0.714)
4- [Flow problems] --> [Foul] (confidence: 0.741)
5- [Limit issue, Foul] --> [Flow problems] (confidence: 0.773)
6- [Foul, Flow problems] --> [Limit issue] (confidence: 0.850)
7- [Foul, Flow problems] --> [Non-Condensables] (confidence: 0.900)
8- [Limit issue, Flow problems] --> [Foul] (confidence: 0.944)
9- [Non-Condensables] --> [Foul] (confidence: 1.000)
10- [Non-Condensables] --> [Flow problems] (confidence: 1.000)
11- [Non-Condensables] --> [Foul, Flow problems] (confidence: 1.000)
12- [Foul, Non-Condensables] --> [Flow problems] (confidence: 1.000)
13- [Flow problems, Non-Condensables] --> [Foul] (confidence: 1.000)
```

Figure 7. Association rules for the co-occurrence of common HVAC faults for the literature analyzed.

4.5.2. Model 2: Common Data-Driven Techniques for Detecting Each HVAC Fault Type

A second ARM model is developed for faults and FDD techniques found in the literature to determine the association between the faults and the methods used to detect and diagnose HVAC faults. The accuracy and performance of the FDD methods are not considered, and only their quantitative adoption in the literature is considered as a measure for the effectiveness of an FDD algorithm for detecting certain fault types. The *support* for FDD techniques was determined and selected to understand how frequently the items for the methods under investigation appear in the dataset.

The followings are the frequency of occurrence, i.e., the *support*, of different analysis methods in our database: 20% for SVM, 19% for ANN and 17% for dimensionality reduction techniques, and 11% for Bayesian networks. A minimum support of 2% and minimum confidence of 50% were selected for model 2, which is appropriate when compared to the highest support (20%) found, which is indicative of a limited number of algorithms in our database. Setting lower thresholds for the second model leads to the generation of a large number of rules that need syntactic constraints to prune and only show the associated faults and methods. The rules found for 100% confidence are removed at the 2% support, as this was considered an indicator of the availability of a few examples, and hence may not represent useful rules. We further limited the rules to those with a single item in their conclusion, which should belong to one of the FDD techniques.

A total of 16,703 rules were mined before being pruned (an excerpt of which is shown in Figure 8). Four methods, namely SVM, ANN, dimensionality reduction, and decision tree (with a *confidence* of 50%, 67%, 50%, and 67%, respectively) resulted in forming 12 rules where eight rules belong to SVM; two rules were found for ANN, and dimensionality reduction and decision tree have one rule each. Other than rule #9, which belongs to ANN, all rules have more than one item in their premises.



SVM Confidence = 50%

```
1- [Leakage, Control] --> [SVM]
2- [Stuck/partially closed, Temperature issue] --> [SVM]
3- [Stuck/partially closed, Control] --> [SVM]
4- [Control, Speed] --> [SVM]
5- [Leakage, Stuck/partially closed, Control] --> [SVM]
6- [Leakage, Control, Speed] --> [SVM]
7- [Stuck/partially closed, Control, Speed] --> [SVM]
8-
   [Leakage, Stuck/partially closed, Control, Speed] ->[SVM]
Neural Networks/Deep Learning Confidence = 67%
9- [Set point] --> [Neural Networks/Deep Learning]
10- [bias/drift/calibration, Control] --> [Neural Networks/Deep
Learning])
Dimensionality reduction Confidence = 50%
11- [Foul, Other faults] --> [Dimensionality reduction]
Decision tree Confidence = 67%
12- [Non-functioning, Speed] --> [Decision tree]
```

**Figure 8.** Excerpt of a deductive self-organizing graph (ISOM) for rules generated for techniques used for different HVAC fault types (minimum *support* = 2%).

At 50% *confidence*, the rules consist of the following fault categories, namely 'leakage', 'control', 'stuck/partially closed', and 'speed', which when combined form rule #9. The initial four rules have two premises made up from the combination of these fault categories, and rules #6 and #7 have three fault categories in their premises which are detected using the SVM algorithm. The ANN algorithms are found to be utilized for diagnosing the 'set point' faults or a combination of sensor-related issues and the 'control' category of faults. Rule #11 indicates the applicability of dimensionality reduction techniques when 'Foul'

or 'other faults' are found together in our database. The decision tree technique, which has the highest joint *confidence* of 67%, is used for detecting 'non-functioning' and 'speed' categories of faults. The findings of this study merely indicate how specific types of faults are often addressed in the sampled research literature, using specific types of algorithms, and they do not provide information on the actual co-occurrence of the faults in building mechanical systems, nor on the performance of the data-driven algorithms with respect to faults.

# 5. Validation and Discussion

The rules discovered through the first model, i.e., the association among HVAC's most common faults, were validated by taking experts' opinions through structured surveys and are validated in the context of the data gathered from the academic papers reviewed to reflect experts' opinions. For the second model, no survey was conducted since the results represent the current state of application for AFDD literature considered in this study. The rules are meant to help better understand the status quo and not necessarily represent best practices. Further, the accuracy of the algorithms was set aside from the comparison due to several reasons, such as comparison among the works based on algorithm performance may not be a fair judgment due to the drastic changes among the contexts of problems reported in the literature (e.g., the system level, fault type being detected, quantity and quality of available data, etc.).Additionally, all studies have not reported their accuracy or have not used the same performance measures.

The survey contained thirteen questions corresponding to the rules detected by the first model. The correspondents were given the Likert scale anchors for the frequency of use, i.e., 'never', 'almost never', 'occasionally/sometimes', 'almost every time', and 'every time'. In addition to these, an 'I do not know' option was added to reduce the uncertainty resulting from enforcing the respondents to answer all the questions in the survey. The survey was made available to respondents with HVAC and FDD experience in the industry or those with relevant research background expertise and was made available for two months.

A total of 13 responses were recorded out of 117 circulated questionnaires (i.e., a response rate of 11%). While the small number of participants may not allow for the validation of results through a statistical analysis approach, here we provide a general overview of the participating experts on the latent patterns detected from the analysis of the literature. The survey results are shown in Figure 9; as can be seen, the scale 'occasionally/sometimes' is, in general, dominating. The criteria indicative of the accuracy of rules is to have recorded a response to either of the following responses: 'Almost every time' or 'Every time'. The responses that receive 'Never', 'Almost never', or 'I don't know' are considered not to be associated with actual rules based on expert opinion. However, the experts' 'Occasionally/Sometimes' response is neither an indicator for strong support or rejection of survey questions, and can only suggest that the ARM supports the correlations found in the investigated database.

Rules #5, #6, and #12 have the highest cumulative responses with above 30% of votes in "Every time" and "Almost every time" categories of responses. For rule #5, when 'limit issue' and 'foul' issues are detected simultaneously by the FDD algorithm, it is likely that the system has a 'limit issue'-related problem. In rule #12, when 'foul' and 'Non-condensables' issues are found simultaneously, it is likely that the system has 'flow problems'-related problems. These three rules can be further investigated to understand if they can be applied beyond the database under investigation. However, rule #7, i.e., "If the 'foul' and 'f/low problems' are detected in the HVAC system for the recorded faults associated with the FDD algorithm being investigated, then the Non-Condensable fault in the system is also likely to occur" is found to need further investigation as it has received the least support from the experts.



☑ Never ■Almost never ■Occasionally/Sometimes ■Almost everytime ■Every time ■I don't know

Figure 9. Diverging bar chart for the frequency of association among HVAC common faults. \* Check Figure 7 for the rules.

A total of 25 rules are mined using the ARM models, which facilitates understanding the relationship between the faults and the suitability of the techniques for FDD purposes. The mined rules alongside the infographic presented in the form of the Sankey diagram allow the asset managers to better comprehend the FDD procedure for different levels of HVAC. In particular, the rules found in model 1, such as *"When 'foul' and 'non-condensables' faults are found together, there is a high chance for 'flow problem'* or *When 'flow problems' and 'non-condensable' issues are found in an HVAC, then 'foul' fault will also occur"* can be interpreted as important rules, as the high *confidence* of 100% shows the significance of the mined rule, and it is worthwhile to be further investigated when there are more data available.

Algorithms and faults at all three levels of HVAC were investigated. Our results show that previous research endeavors have mostly been focused on the sub-system-level FDD (as shown by the 56% share of this level from the whole studies reviewed) and has often used the SVM, ANN, and dimensionality reduction techniques. The second research area focuses on system-level FDD (taking a 40% share) where ANN, dimensionality reduction techniques, and Bayesian networks are the dominant algorithms. At the component/device/part level, which covers only 4% of the studies, the SVM and ARM methods are used; further data is needed to investigate the part and component-level FDD.

Additionally, model 2 provides four FDD algorithms and their associated faults in the form of a set of rules that allows the asset managers to decide on the type of algorithm that can be selected for AFDD of the HVAC system faults. For example, the SVM algorithm is found to be effective in FDD when fault types belong to 'leakage', 'stuck/partially

closed', and 'control' issues. It was found that some algorithms are used more often for detecting particular faults. The algorithms that can be utilized for each category of the top six common HVAC faults separately are shown in below Table 2 and are organized in descending order.

Table 2. Recommer	d algorithms for an	individual category	of HVAC syste	ems most common faults
	()	<b>()</b>		

Fault Category	Recommended Algorithms
Limit issue	SVM-ANN-BN
Stuck/Partially closed	ANN-SVM-DT
Flow problems	ANN-SVM-BN
Bias/Drift/Calibration	ANN-Dimensionality reduction methods-SVM
Leakage	SVM-ANN-Dimensionality reduction methods
Foul	SVM-ANN-Dimensionality reduction methods

#### 6. Conclusions

In addition to data availability, understanding the association among faults (manifested in the rules offered by this paper) and the suitability of data-driven algorithms to detect and diagnose them is essential for AFDD. The knowledge extracted from the wealth of reviewed machine learning techniques can aid in better comprehending the complexities that exist in HVAC systems at various levels. The first model developed in this study assists the asset managers/facility managers to better understand the associations among faults and anticipation of other fault types that can be expected when certain faults are identified in the HVAC system by the literature for the database considered. Moreover, the association rules of the second model can be used to understand the status quo of FDD algorithm adoption to assist asset managers to better train data-driven FDD systems using the most suitable algorithms based on the fault(s) of importance.

This study contributes to the body of knowledge by exploring and analyzing the data found in the literature to develop these two sets of rules based on the present status quo of black-box models. The investigation of relationships between fault type co-occurrence, and also associations among fault types and machine learning models used for data-driven FDD can significantly help the implementation of AFDD in practice. Recommender systems can be developed on the basis of the rules extracted and validated in this study to recommend fault check and diagnosis techniques. Such a recommender system can facilitate realistic anticipation of associated faults and provide support regarding the suitability of FDD algorithms for a single fault or a combination of faults. In addition, the implementation of such recommender systems can benefit both the building maintenance program and energy consumption aspect of the facility management by identifying the likelihood of occurrence of one or more fault(s) by the observation of other faults in the HVAC system.

Despite their capabilities, the currently available AFDD techniques less often make use of any physical knowledge of the built facility and are mainly fed with features from BMS/BAS. For example, due to the lack of knowledge of sensor locations, weather data, and occupancy information, in many cases, it is difficult to effectively detect and diagnose the cause of faults in HVAC systems. AFDD models can benefit from contextual and spatial information from sources such as BIM to enhance the process of fault detection and diagnostics, fault propagation, and analysis in building HVAC systems at different levels. This must be taken into closer consideration by future studies. Accordingly, the future work can be divided into two streams: ones that can work on FDD models and improve the existing challenges of FDD methods and create more accurate models with fewer false alarms. The second direction is the analysis of the models developed using the same/different data sources. Methods such as feature selection can be analyzed to understand the relationship that may exist between them. All stated directions can be enhanced by adding contextual and spatial information, by improving the user's understanding of the system and creating more robust AFDD models.

**Author Contributions:** Conceptualization, M.N.-B. and A.H.G. methodology, M.N.-B.; software, A.H.G.; validation, A.H.G. and M.N.-B.; formal analysis, A.H.G.; investigation, A.H.G.; resources, A.H.G.; data curation, A.H.G.; writing—original draft preparation, A.H.G.; writing—review and editing, M.N.-B.; visualization, A.H.G.; supervision, M.N.-B.; project administration, A.H.G.; funding acquisition, M.N.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in https://github. com/arashhosseiniarash/association-rule-mining.

**Conflicts of Interest:** The authors declare no conflict of interest.

# Appendix A

Table A1. List of the 82 studies considered for analysis.

No.	Author(s)	Title	Year
1.	K. Yan, J. Huang, W. Shen, and Z. Ji	Unsupervised learning for fault detection and diagnosis of air handling units	2020
2.	K. Yan, A. Chong, and Y. Mo	Generative adversarial network for fault detection diagnosis of chillers	2020
3.	A. Ranade, G. Provan, A. El-Din Mady, and D. O'Sullivan	A computationally efficient method for fault diagnosis of fan-coil unit terminals in building Heating Ventilation and Air Conditioning systems	2020
4.	S. Miyata, J. Lim, Y. Akashi, Y. Kuwahara, and K. Tanaka	Fault detection and diagnosis for heat source system using convolutional neural network with imaged faulty behavior data	2020
5.	Z. Zhang, H. Han, X. Cui, and Y. Fan,	Novel application of multi-model ensemble learning for fault diagnosis in refrigeration systems	2020
6.	Y. Fan, X. Cui, H. Han, and H. Lu	Chiller fault detection and diagnosis by knowledge transfer based on adaptive imbalanced processing	2020
7.	A. Montazeri and S.M. Kargar,	Fault detection and diagnosis in air handling using data-driven methods	2020
8.	J. Liu et al.	Data-driven and association rule mining-based fault diagnosis and action mechanism analysis for building chillers	2020
9.	M. Elnour, N. Meskin, and M. Al-Naemi	Sensor data validation and fault diagnosis using Auto-Associative Neural Network for HVAC systems	2020
10.	Z. Li et al.	Machine learning based diagnosis strategy for refrigerant charge amount malfunction of variable refrigerant flow system	2020
11.	Y. Fan, X. Cui, H. Han, and H. Lu	Feasibility and improvement of fault detection and diagnosis based on factory-installed sensors for chillers	2020
12.	K. Yan, Z. Ji, H. Lu, J. Huang, W. Shen, and Y. Xue	Fast and Accurate Classification of Time Series Data Using Extended ELM: Application in Fault Diagnosis of Air Handling Units	2019
13.	A. Motomura et al.	Fault evaluation process in HVAC system for decision making of how to respond to system faults	2019
14.	Z. Li et al.	An efficient online wkNN diagnostic strategy for variable refrigerant flow system based on coupled feature selection method	2019
15.	G. Li and Y. Hu	An enhanced PCA-based chiller sensor fault detection method using ensemble empirical mode decomposition based denoising	2019
16.	D. Li, D. Li, C. Li, L. Li, and L. Gao	A novel data-temporal attention network based strategy for fault diagnosis of chiller sensors	2019

# Table A1. Cont.

No.	Author(s)	Title	Year
17.	D. Li, Y. Zhou, G. Hu, and C. J. Spanos	Handling Incomplete Sensor Measurements in Fault Detection and Diagnosis for Building HVAC Systems	2019
18.	H. Han, X. Cui, Y. Fan, and H. Qing	Least squares support vector machine (LS-SVM)-based chiller fault diagnosis using fault indicative features	2019
19.	D. Bigaud, A. Charki, A. Caucheteux, F. Titikpina, and T. Tiplica	Detection of Faults and Drifts in the Energy Performance of a Building Using Bayesian Networks	2019
20.	A. Beghi, R. Brignoli, L. Cecchinato, G. Menegazzo, and M. Rampazzo	A data-driven approach for fault diagnosis in HVAC chiller systems	2019
21.	J. Liu, M. Zhang, H. Wang, W. Zhao, and Y. Liu	Sensor Fault Detection and Diagnosis Method for AHU Using 1-D CNN and Clustering Analysis	2019
22.	C. Zhong, K. Yan, Y. Dai, N. Jin, and B. Lou	Diagnosis of Air Handling Units Using Generative Adversarial Networks	2019
23.	C. Yang, W. Shen, B. Gunay, and Z. Shi	Toward Machine Learning-based Prognostics for Heating Ventilation and Air-Conditioning Systems	2019
24.	L. Gao, D. Li, D. Li, L. Yao, L. Liang, and Y. Gao	A Novel Chiller Sensors Fault Diagnosis Method Based on Virtual Sensors	2019
25.	M. Tahmasebi, K. Eaton, N. Nassif, and R. Talib	Integrated Machine Learning Modeling and Fault Detection Approach for Chilled Water Systems	2019
26.	J. Liu, G. Li, B. Liu, K. Li, and H. Chen	Knowledge discovery of data-driven-based fault diagnostics for building energy systems: A case study of the building variable refrigerant flow system	2019
27.	A. Behravan, M. Abboush, and R. Obermaisser	Deep Learning Application in Mechatronics Systems' Fault Diagnosis, a Case Study of the Demand-Controlled Ventilation and Heating System	2019
28.	H. Zhang, H. Chen, Y. Guo, J. Wang, G. Li, and L. Shen	Sensor fault detection and diagnosis for a water source heat pump air-conditioning system based on PCA and preprocessed by combined clustering	2019
29.	M. Elnour, N. Meskin, and M. Al-Naemi	Sensor Fault Diagnosis of Multi-Zone HVAC Systems Using Auto-Associative Neural Network	2019
30.	Y. Fan, X. Cui, H. Han, and H. Lu	Chiller fault diagnosis with field sensors using the technology of imbalanced data	2019
31.	B. Jin, D. Li, S. Srinivasan, SK. Ng, K. Poolla, and A. Sangiovanni-Vincentelli	Detecting and Diagnosing Incipient Building Faults Using Uncertainty Information from Deep Neural Networks	2019
32.	K. Yan and J. Hua	Deep Learning Technology for Chiller Faults Diagnosis	2019
33.	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	2019
34.	Y.H. Eom, J.W. Yoo, S.B. Hong, and M.S. Kim	Refrigerant charge fault detection method of air source heat pump system using convolutional neural network for energy saving	2019
35.	K. Yan, C. Zhong, Z. Ji, and J. Huang	Semi-supervised learning for early detection and diagnosis of various air handling unit faults	2018
36.	K. Yan, L. Ma, Y. Dai, W. Shen, Z. Ji, and D. Xie	Cost-sensitive and sequential feature selection for chiller fault detection and diagnosis	2018
37.	Z. Wang, Z. Wang, X. Gu, S. He, and Z. Yan	Feature selection based on Bayesian network for chiller fault diagnosis from the perspective of field applications	2018
38.	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	2018
39.	M. Hu et al.	A machine learning Bayesian network for refrigerant charge faults of variable refrigerant flow air conditioning system	2018
40.	Y. Guo et al.	Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving	2018
41.	M. Dey, S.P. Rana, and S. Dudley	Smart building creation in large scale HVAC environments through automated fault detection and diagnosis	2018
42.	M. Dey, S.P. Rana, and S. Dudley	Semi-Supervised Learning Techniques for Automated Fault Detection and Diagnosis of HVAC Systems	2018
43.	F. Simmini, M. Rampazzo, A. Beghi, and F. Peterle	Local Principal Component Analysis for Fault Detection in Air-Condensed Water Chillers	2018
44.	Y. Chen and J. Wen	Development and Field Evaluation of Data-driven Whole Building Fault Detection and Diagnosis Strategy	2018

# Table A1. Cont.

No.	Author(s)	Title	Year
45.	K. Yan, C. Zhong, Z. Ji, and J. Huang	Evaluating Semi-supervised Learning for Automated Fault	2018
46.	Y. Chen, J. Wen, T. Chen, and O. Pradhan	Detection and Diagnosis of Air Handling Units Bayesian Networks for Whole Building Level Fault Diagnosis and Isolation	2018
47.	G. Li et al.	An improved decision tree-based fault diagnosis method for practical variable refrigerant flow system using virtual sensor-based fault indicators	2018
48.	X. Liu, Y. Li, X. Liu, and J. Shen	Fault diagnosis of chillers using very deep convolutional network	2018
49.	R. Huang et al.	An effective fault diagnosis method for centrifugal chillers using	2018
		associative classification	-010
50.	Z. Wang, L. Wang, K. Liang, and Y. Tan,	principal component analysis	2018
51.	J. Liu, G. Li, H. Chen, J. Wang, Y. Guo, and J. Li	A robust online refrigerant charge fault diagnosis strategy for VRF systems based on virtual sensor technique and PCA-EWMA method	2017
52.	K. Yan, Z. Ji, and W. Shen	Online fault detection methods for chillers combining extended	2017
		kalman filter and recursive one-class SVM Combining knowledge and historical data for system-level fault	
53.	K. Verbert, R. Babuška, and B. De Schutter	diagnosis of HVAC systems	2017
54	P.M. Van Every, M. Rodriguez, C.B. Jones,	Advanced detection of HVAC faults using unsupervised SVM	2017
01.	A.A. Mammoli, and M. Martínez-Ramón	novelty detection and Gaussian process models	2017
55.	W.J.N. Turner, A. Staino, and B. Basu	approach	2017
56.	S. Sun, G. Li, H. Chen, Q. Huang, S. Shi, and W. Hu	A hybrid ICA-BPNN-based FDD strategy for refrigerant charge	2017
57	C Chi at al	Refrigerant charge fault diagnosis in the VRF system using	2017
57.		Bayesian artificial neural network combined with Relief Filter	2017
58.	S.C. Mukhopadhyay, O.A. Postolache, K.P. Jayasundera, and A.K. Swain, Eds.	Sensors for everyday life: environmental and food engineering	2017
59.	K. Mittal, J.P. Wilson, B.P. Baillie, S. Gupta, G.M. Bollas, and P.B. Luh	Supervisory Control for Resilient Chiller Plants Under Condenser Fouling	2017
60.	Y. Guo et al.	Modularized PCA method combined with expert-based multivariate decoupling for FDD in VRF systems including indoor unit faults	2017
61.	Y. Guo et al.	An enhanced PCA method with Savitzky-Golay method for VRF system sensor fault detection and diagnosis	2017
62.	Y. Chen and J. Wen	A whole building fault detection using weather based pattern matching and feature based PCA method	2017
63.	L. Chang, H. Wang, and L. Wang	Cloud-Based parallel implementation of an intelligent classification algorithm for fault detection and diagnosis of HVAC systems	2017
64.	Z. Wang, Z. Wang, S. He, X. Gu, and Z.F. Yan	Fault detection and diagnosis of chillers using Bayesian network	2017
011		merged distance rejection and multi-source non-sensor information	-017
65.	Y. Chen and J. Wen	matching and PCA method	2017
66.	J. Wang et al.	under heating mode	2017
67.	S. Shi et al.	An efficient VRF system fault diagnosis strategy for refrigerant charge amount based on PCA and dual neural network model	2017
68.	R. Yan, Z. Ma, Y. Zhao, and G. Kokogiannakis	A decision tree based data-driven diagnostic strategy for air handling units	2016
69.	K. Sun, G. Li, H. Chen, J. Liu, J. Li, and W. Hu	A novel efficient SVM-based fault diagnosis method for multi-split air conditioning system's refrigerant charge fault amount	2016
70.	J. Liu, Y. Hu, H. Chen, J. Wang, G. Li, and W. Hu	A refrigerant charge fault detection method for variable refrigerant flow (VRF) air-conditioning systems	2016
71.	J. Liu, H. Chen, J. Wang, G. Li, H. Li, and W. Hu	Fault diagnosis of refrigerant charge based on PCA and decision tree for variable refrigerant flow systems	2016
72.	G. Li et al.	An improved fault detection method for incipient centrifugal chiller faults using the PCA-R-SVDD algorithm	2016

Table	A1.	Cont.
-------	-----	-------

No.	Author(s)	Title	Year
73.	G. Li et al.	A sensor fault detection and diagnosis strategy for screw chiller system using support vector data description-based D-statistic and DV-contribution plots	2016
74.	D. Li, G. Hu, and C. J. Spanos	A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis	2016
75.	Y. Hu, G. Li, H. Chen, H. Li, and J. Liu	Sensitivity analysis for PCA-based chiller sensor fault detection	2016
76.	S. He, Z. Wang, Z. Wang, X. Gu, and Z. Yan	Fault detection and diagnosis of chiller using Bayesian network classifier with probabilistic boundary	2016
77.	Y. Gao, S. Liu, F. Li, and Z. Liu	Fault detection and diagnosis method for cooling dehumidifier based on LS-SVM NARX model,	2016
78.	A. Beghi, R. Brignoli, L. Cecchinato, G. Menegazzo, M. Rampazzo, and F. Simmini	Data-driven Fault Detection and Diagnosis for HVAC water chillers	2016
79.	R. Yan, Z. Ma, G. Kokogiannakis, and Y. Zhao	A sensor fault detection strategy for air handling units using cluster analysis	2016
80.	D.A.T. Tran, Y. Chen, H.L. Ao, and H.N.T. Cam	An enhanced chiller FDD strategy based on the combination of the LSSVR-DE model and EWMA control charts	2016
81.	D.A.T. Tran, Y. Chen, and C. Jiang	Comparative investigations on reference models for fault detection and diagnosis in centrifugal chiller systems	2016
82.	C. Audivet Durán and M.E. Sanjuán	On-Line Early Fault Detection of a Centrifugal Chiller Based on Data Driven Approach	2016

# References

- 1. Major Energy Retrofit Guidelines for Commercial and Institutional Buildings–Non-Food Retail. Available online: https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/oee/buildings/pdf/RetrofitGuidelines-e.pdf (accessed on 16 October 2021).
- 2. Government of Canada. Households and the Environment: Energy Use: Analysis. Available online: https://www150.statcan.gc. ca/n1/pub/11-526-s/2013002/part-partie1-eng.htm (accessed on 5 August 2020).
- 3. Chakraborty, D.; Elzarka, H. Early detection of faults in HVAC systems using an XGBoost model with a dynamic threshold. *Energy Build.* **2019**, *185*, 326–344. [CrossRef]
- Beghi, A.; Brignoli, R.; Cecchinato, L.; Menegazzo, G.; Rampazzo, M. A data-driven approach for fault diagnosis in HVAC chiller systems. In Proceedings of the 2015 IEEE Conference on Control Applications (CCA), Sydney, Australia, 21–23 September 2015; pp. 966–971.
- 5. Yan, R.; Ma, Z.; Zhao, Y.; Kokogiannakis, G. A decision tree based data-driven diagnostic strategy for air handling units. *Energy Build.* **2016**, *133*, 37–45. [CrossRef]
- 6. Sun, S.; Li, G.; Chen, H.; Huang, Q.; Shi, S.; Hu, W. A hybrid ICA-BPNN-based FDD strategy for refrigerant charge faults in variable refrigerant flow system. *Appl. Therm. Eng.* **2017**, *127*, 718–728. [CrossRef]
- Chen, Y.; Wen, J. A whole building fault detection using weather based pattern matching and feature based PCA method. In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; pp. 4050–4057.
- 8. Baldi, S.; Zhang, F.; Le Quang, T.; Endel, P.; Holub, O. Passive versus active learning in operation and adaptive maintenance of Heating, Ventilation, and Air Conditioning. *Appl. Energy* **2019**, 252, 113478. [CrossRef]
- 9. Sun, L.; Wu, J.; Jia, H.; Liu, X. Research on fault detection method for heat pump air conditioning system under cold weather. *Chin. J. Chem. Eng.* **2017**, *25*, 1812–1819. [CrossRef]
- 10. Comstock, M.C.; Braun, J.E.; Groll, E.A. A survey of common faults for chillers/Discussion. Ashrae Trans. 2002, 108, 819.
- 11. Li, D.; Hu, G.; Spanos, C.J. A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis. *Energy Build.* 2016, 128, 519–529. [CrossRef]
- 12. Hu, Y.; Li, G.; Chen, H.; Li, H.; Liu, J. Sensitivity analysis for PCA-based chiller sensor fault detection. *Int. J. Refrig.* 2016, 63, 133–143. [CrossRef]
- 13. Li, D.; Zhou, Y.; Hu, G.; Spanos, C.J. Handling Incomplete Sensor Measurements in Fault Detection and Diagnosis for Building HVAC Systems. *IEEE Trans. Autom. Sci. Eng.* **2020**, *17*, 833–846. [CrossRef]
- 14. Padilla, M.; Choinière, D. A combined passive-active sensor fault detection and isolation approach for air handling units. *Energy Build.* **2015**, *99*, 214–219. [CrossRef]
- 15. Han, H.; Cui, X.; Fan, Y.; Qing, H. Least squares support vector machine (LS-SVM)-based chiller fault diagnosis using fault indicative features. *Appl. Therm. Eng.* **2019**, *154*, 540–547. [CrossRef]
- 16. Tran, D.A.T.; Chen, Y.; Jiang, C. Comparative investigations on reference models for fault detection and diagnosis in centrifugal chiller systems. *Energy Build.* **2016**, 133, 246–256. [CrossRef]

- 17. Yan, K.; Ji, Z.; Shen, W. Online fault detection methods for chillers combining extended kalman filter and recursive one-class SVM. *Neurocomputing* **2017**, *228*, 205–212. [CrossRef]
- Van Every, P.M.; Rodriguez, M.; Jones, C.B.; Mammoli, A.A.; Martínez-Ramón, M. Advanced detection of HVAC faults using unsupervised SVM novelty detection and Gaussian process models. *Energy Build.* 2017, 149, 216–224. [CrossRef]
- 19. Zhang, Z.; Han, H.; Cui, X.; Fan, Y. Novel application of multi-model ensemble learning for fault diagnosis in refrigeration systems. *Appl. Therm. Eng.* 2020, *164*, 114516. [CrossRef]
- Deshmukh, S.; Glicksman, L.; Norford, L. Case study results: Fault detection in air-handling units in buildings. *Adv. Build. Energy Res.* 2018, 14, 305–321. [CrossRef]
- 21. Shi, Z.; O'Brien, W. Development and implementation of automated fault detection and diagnostics for building systems: A review. *Autom. Constr.* **2019**, 104, 215–229. [CrossRef]
- 22. Afram, A.; Janabi-Sharifi, F. Review of modeling methods for HVAC systems. Appl. Therm. Eng. 2014, 67, 507–519. [CrossRef]
- Kim, W.; Katipamula, S. A review of fault detection and diagnostics methods for building systems. *Sci. Technol. Built Environ.* 2017, 24, 3–21. [CrossRef]
- Zhao, Y.; Li, T.; Zhang, X.; Zhang, C. Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. *Renew. Sustain. Energy Rev.* 2019, 109, 85–101. [CrossRef]
- 25. Rogers, A.; Guo, F.; Rasmussen, B. A review of fault detection and diagnosis methods for residential air conditioning systems. *Build. Environ.* **2019**, *161*, 106236. [CrossRef]
- 26. Mirnaghi, M.S.; Haghighat, F. Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: A comprehensive review. *Energy Build.* **2020**, *229*, 110492. [CrossRef]
- 27. Zhao, Y.; Zhang, C.; Zhang, Y.; Wang, Z.; Li, J. A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis. *Energy Built Environ.* **2020**, *1*, 149–164. [CrossRef]
- 28. Li, Y.; O'Neill, Z. A critical review of fault modeling of HVAC systems in buildings. Build. Simul. 2018, 11, 953–975. [CrossRef]
- Bang, M.; Engelsgaard, S.S.; Alexandersen, E.K.; Skydt, M.R.; Shaker, H.R.; Jradi, M. Novel real-time model-based fault detection method for automatic identification of abnormal energy performance in building ventilation units. *Energy Build.* 2018, 183, 238–251. [CrossRef]
- 30. Ranade, A.; Provan, G.; Mady, A.E.-D.; O'Sullivan, D. A computationally efficient method for fault diagnosis of fan-coil unit terminals in building Heating Ventilation and Air Conditioning systems. *J. Build. Eng.* **2020**, *27*, 100955. [CrossRef]
- 2013 ASHRAE Handbook-Fundamentals (SI Edition). Available online: https://app-knovel-com.lib-ezproxy.concordia.ca/ hotlink/toc/id:kpASHRAEC1/ashrae-handbook-fundamentals/ashrae-handbook-fundamentals (accessed on 16 October 2021).
- 32. Serale, G.; Fiorentini, M.; Capozzoli, A.; Bernardini, D.; Bemporad, A. Model Predictive Control (MPC) for Enhancing Building and HVAC System Energy Efficiency: Problem Formulation, Applications and Opportunities. *Energies* **2018**, *11*, 631. [CrossRef]
- Beghi, A.; Cecchinato, L.; Peterle, F.; Rampazzo, M.; Simmini, F. Model-based fault detection and diagnosis for centrifugal chillers. In Proceedings of the 2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol), Barcelona, Spain, 7–9 September 2016; pp. 158–163.
- Beghi, A.; Brignoli, R.; Cecchinato, L.; Menegazzo, G.; Rampazzo, M.; Simmini, F. Data-driven Fault Detection and Diagnosis for HVAC water chillers. *Control. Eng. Pract.* 2016, 53, 79–91. [CrossRef]
- Yan, K.; Huang, J.; Shen, W.; Ji, Z. Unsupervised learning for fault detection and diagnosis of air handling units. *Energy Build.* 2020, 210, 109689. [CrossRef]
- 36. Li, G.; Hu, Y. An enhanced PCA-based chiller sensor fault detection method using ensemble empirical mode decomposition based denoising. *Energy Build.* **2019**, *183*, 311–324. [CrossRef]
- Tran, D.A.T.; Chen, Y.; Chau, M.Q.; Ning, B. A robust online fault detection and diagnosis strategy of centrifugal chiller systems for building energy efficiency. *Energy Build.* 2015, 108, 441–453. [CrossRef]
- 38. Li, Z.; Tan, J.; Li, S.; Liu, J.; Chen, H.; Shen, J.; Huang, R.; Liu, J. An efficient online wkNN diagnostic strategy for variable refrigerant flow system based on coupled feature selection method. *Energy Build.* **2018**, *183*, 222–237. [CrossRef]
- 39. Bigaud, D.; Charki, A.; Caucheteux, A.; Titikpina, F.; Tiplica, T. Detection of Faults and Drifts in the Energy Performance of a Building Using Bayesian Networks. *J. Dyn. Syst. Meas. Control.* **2019**, *141*, 101011. [CrossRef]
- 40. Katipamula, S.; Brambley, M.R. Review Article: Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems— A Review, Part I. *HVAC&R Res.* 2005, *11*, 3–25. [CrossRef]
- 41. Li, D.; Li, D.; Li, C.; Li, L.; Gao, L. A novel data-temporal attention network based strategy for fault diagnosis of chiller sensors. *Energy Build*. **2019**, *198*, 377–394. [CrossRef]
- Yan, K.; Zhong, C.; Ji, Z.; Huang, J. Semi-supervised learning for early detection and diagnosis of various air handling unit faults. Energy Build. 2018, 181, 75–83. [CrossRef]
- 43. He, S.; Wang, Z.; Wang, Z.; Gu, X.; Yan, Z. Fault detection and diagnosis of chiller using Bayesian network classifier with probabilistic boundary. *Appl. Therm. Eng.* **2016**, *107*, 37–47. [CrossRef]
- 44. Roth, K.W.; Westphalen, D.; Llana, P.; Feng, M. *The Energy Impact of Faults in U.S. Commercial Buildings*; Purdue University: West Lafayette, Indiana, 2004; p. 9.
- 45. Frank, S.M.; Kim, J.; Cai, J.; Braun, J.E. Common Faults and Their Prioritization in Small Commercial Buildings: February 2017–December 2017; National Renewable Energy Laboratory: Golden, CO, USA, 2017.

- Roth, K.W.; Westphalen, D.; Feng, M.Y.; Llana, P.; Quartararo, L. Energy Impact of Commercial Building Controls and Performance Diagnostics: Market Characterization, Energy Impact of Building Faults and Energy Savings Potential; US Department of Energy: Cambridge, MA, USA, 2005; p. 413.
- 47. Available online: https://github.com/arashhosseiniarash/sankey-diagram-journal (accessed on 1 July 2021).
- 48. Zhang, W.; Liao, H.; Zhao, N. Research on the FP Growth Algorithm about Association Rule Mining. In 2008 International Seminar on Business and Information Management; IEEE: Piscataway, NJ, USA, 2008; Volume 1, pp. 315–318.
- 49. Agrawal, R.; Imieliński, T.; Swami, A. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, 25–28 May 1993; pp. 207–216.