

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Dissertations and Student Research:  
Architectural Engineering

Durham School of Architectural Engineering  
and Construction

---

11-2021

## Investigation of the Prevalence of Faults in the Heating, Ventilation, and Air-Conditioning Systems of Commercial Buildings

Amir Ebrahimifakhar  
*University of Nebraska-Lincoln*

Follow this and additional works at: <https://digitalcommons.unl.edu/archengdiss>



Part of the [Architectural Engineering Commons](#), and the [Mechanical Engineering Commons](#)

---

Ebrahimifakhar, Amir, "Investigation of the Prevalence of Faults in the Heating, Ventilation, and Air-Conditioning Systems of Commercial Buildings" (2021). *Dissertations and Student Research: Architectural Engineering*. 68.  
<https://digitalcommons.unl.edu/archengdiss/68>

This Article is brought to you for free and open access by the Durham School of Architectural Engineering and Construction at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Dissertations and Student Research: Architectural Engineering by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

INVESTIGATION OF THE PREVALENCE OF FAULTS IN THE  
HEATING, VENTILATION, AND AIR-CONDITIONING SYSTEMS OF  
COMMERCIAL BUILDINGS

by

Amir Ebrahimifakhar

A DISSERTATION

Presented to the Faculty of  
The Graduate College at the University of Nebraska  
In Partial Fulfillment of Requirements  
For the Degree of Doctor of Philosophy

Major: Architectural Engineering

Under the Supervision of Professor David Yuill

Lincoln, Nebraska

November, 2021

# INVESTIGATION OF THE PREVALENCE OF FAULTS IN THE HEATING, VENTILATION, AND AIR-CONDITIONING SYSTEMS OF COMMERCIAL BUILDINGS

Amir Ebrahimifakhar, Ph.D.

University of Nebraska, 2021

Advisor: David Yuill

This dissertation describes a large-scale investigation of heating, ventilation, and air-conditioning (HVAC) fault prevalence in commercial buildings in the United States. A multi-year dataset with 36,556 pieces of HVAC equipment including air handling units (AHUs), air terminal units (ATUs), and packaged rooftop units (RTUs) was analyzed to determine values for several HVAC fault prevalence metrics. The primary source of data for this study comes from three commercial fault detection and diagnostics (FDD) providers. Since each FDD provider uses different terms to refer to the same fault in an HVAC system, a mapping function was created for each FDD provider's dataset, to convert the fault reports to a single standardized fault identifier. The fault identifier is taken from a standard taxonomy that was created for this purpose.

Since the commercial FDD software outputs are inherently subject to some level of error, i.e., they could have false negatives and false positives, a field study was conducted to gain greater insight into the commercial FDD software results. Two buildings from among the buildings of one of the FDD providers were selected. The RTUs serving these two buildings were monitored for about two weeks using our installed data loggers. The actual faults in these buildings were identified using methods that we developed or selected from the literature. The results of the field study were compared with the FDD provider fault reports.

This study also proposes a data-driven FDD strategy for RTUs, using machine learning classification methods. The FDD task is formulated as a multi-class classification problem. Seven typical RTU faults are discriminated against one another as well as the normal condition. Nine classification methods were applied to a dataset of simulation data, which was split into a training set and a test set. The performance of the classifiers for individual faults was characterized using true positive rate and false positive rate statistical measures. The relative importance of input variables was analyzed, and is also discussed in the dissertation.

## **Dedication**

This thesis is dedicated to my mother.

## **Acknowledgments**

I am highly thankful to all people who helped me during my journey at the University of Nebraska-Lincoln. I am extremely grateful to my advisor, Dr. David Yuill, for his support during my Ph.D. program. His brilliant mind, willingness to help, and demand for excellence did make this project possible. Thank you, Dr. David Yuill, now and always. I could not ask for a better advisor. I am grateful to the members of my committee, Dr. Fadi Alsaleem, Dr. Josephine Lau, and Dr. Seunghee Kim for their time, and useful feedback throughout this project. Special thanks to Dr. Jessica Granderson, Eliot Crowe, and Dr. Yimin Chen from Lawrence Berkeley National Laboratory (LBNL), and Dr. Amanda Smith, and Dr. Hayden Reeve from Pacific Northwest National Laboratory (PNNL) for providing the data and collaborating with me in this project.

There are people in everyone's lives who make success possible. My wife, Paridokht Kamalimehr, steadfastly supported and encouraged me. Without her endless love and encouragement, I would never have been able to complete my Ph.D. program. My parents, brothers, and sister always were there for me and gave me lots of support. I love you all and I appreciate everything that you have done for me. My friends and colleagues, Dr. Yifeng Hu, and Dr. Adel Kabirikopaei, helped me whenever I needed it the most. I will never forget the long discussions we had about different building science topics in graduate school. Special thanks go to John and Jennifer Oliver who always supported me while I was studying abroad far away from my family. Thanks for being my friends.

# **Table of Contents**

CHAPTER 1. Introduction..... 1

    1.1. Study Scope..... 3

    1.2. HVAC Fault Prevalence Metrics..... 3

    1.3. Study Data ..... 4

    1.4. Verification..... 5

    1.5. Data-Driven FDD for RTUs..... 7

    1.6. Structure of the Thesis..... 9

CHAPTER 2. Literature Review ..... 10

    2.1. HVAC Fault Prevalence ..... 10

    2.2. RTU Fault Detection and Diagnostics ..... 21

CHAPTER 3. Methodology..... 27

    3.1. Analysis of FDD Records as an Indicator of HVAC Fault Prevalence ..... 27

        3.1.1. Data Overview ..... 27

        3.1.2. Data Preparation ..... 29

        3.1.3. Standardized Taxonomy for HVAC Faults ..... 31

        3.1.4. Metric Definitions..... 33

            3.1.4.1. Monthly Fault Presence (Metric 1) ..... 34

            3.1.4.2. Average Monthly Fault Presence (Metric 2) ..... 36

            3.1.4.3. Mean Number of Faults per Building per Month (Metric 3) ..... 36

    3.2. Field Study Verification ..... 37

        3.2.1. Sensor Faults ..... 41

        3.2.2. Economizer Damper Stuck..... 42

        3.2.3. Non-Condensable Gas ..... 44

        3.2.4. Abnormal Supply Fan Belt Tension ..... 45

        3.2.5. No Unoccupied Temperature Setback..... 45

    3.3. Machine Learning based FDD for RTUs ..... 46

CHAPTER 4. HVAC Fault Prevalence Results & Discussion..... 56

    4.1. Monthly Fault Presence (Metric 1) Results..... 56

4.1.1. AHU Results.....	56
4.1.2. ATU Results .....	59
4.1.3. RTU Results .....	62
4.2. Average Monthly Fault Presence (Metric 2) Results .....	64
4.2.1. AHU Results.....	64
4.2.2. ATU Results .....	69
4.2.3. RTU Results .....	71
4.3. Mean Number of Faults per Building per Month (Metric 3) Results.....	73
4.3.1. FDD provider A Results .....	74
4.3.2. FDD provider B Results .....	76
4.3.3. FDD provider C Results .....	78
CHAPTER 5. Field Study Results & Discussion .....	80
5.1. First Building Results.....	80
5.2. Second Building Results .....	85
5.3. Confidence Interval for Fault Prevalence.....	87
CHAPTER 6. Data-Driven FDD for RTUs Results & Discussion.....	89
CHAPTER 7. Conclusions and Recommendations for Future Research .....	100
7.1. HVAC Fault Prevalence Summary .....	100
7.2. Field Study Summary .....	102
7.3. Data-Driven FDD for RTUs Summary .....	103
References.....	106
APPENDIX A - Monthly Fault Presence (Metric 1) for AHU Faults.....	112
APPENDIX B - Monthly Fault Presence (Metric 1) for ATU Faults.....	115

## List of Tables

Table 1. Essential and preferred data required for study .....	6
Table 2. Strengths and weaknesses of the field verification .....	7
Table 3. Data sampling methods.....	12
Table 4. Fault occurrence metric definitions .....	18
Table 5. Summary of HVAC fault prevalence studies.....	20
Table 6. FDD data sources.....	28
Table 7. Number of buildings by building type .....	28
Table 8. Number of buildings by Building America climate zone .....	29
Table 9. Standard binary daily fault (BDF) data.....	30
Table 10. Example list of the AHU faults in the developed taxonomy .....	31
Table 11. Example list of the ATU faults in the developed taxonomy .....	32
Table 12. Example list of the RTU faults in the developed taxonomy .....	32
Table 13. Field study site and equipment descriptions .....	37
Table 14. BAS measurements.....	39
Table 15. RTU site visit data .....	40
Table 16. Specifications of the RTU system.....	49
Table 17. The statistics of the input variables of the total data set .....	52
Table 18. Fault class scenarios of the datasets.....	53
Table 19. An illustrative confusion matrix with three classes .....	54
Table 20. FDD provider C fault report for the first building .....	81
Table 21. Actual faults identified for the first building .....	82
Table 22. RTUs with and without non-condensable gas.....	84
Table 23. RTUs with normal and abnormal supply fan belt tension .....	84
Table 24. FDD provider C fault report for the second building.....	86
Table 25. Field study results for the second building .....	87
Table 26. 95% confidence interval for the prevalence of RTU faults .....	88
Table 27. A full list of R packages and functions used.....	89
Table 28. The list of tuning parameters for each of the classification methods.....	92

## List of Figures

Figure 1. Graphical depiction of monthly fault presence (Metric 1) .....	36
Figure 2. Our installed sensor beside the zone temperature and relative humidity sensor .....	42
Figure 3. Economizer section of the one of the RTUs in the field.....	44
Figure 4. Schematic of a typical vapor compression refrigeration cycle .....	50
Figure 5. Monthly fault presence (metric 1) for AHU simultaneous heating and cooling for FDD provider A .....	57
Figure 6. Monthly fault presence (metric 1) for AHU mixed air temperature abnormal for FDD provider A .....	57
Figure 7. Monthly fault presence (metric 1) for AHU heating coil valve leakage for FDD provider B.....	58
Figure 8. Monthly fault presence (metric 1) for AHU heating coil valve hunting for FDD provider B.....	59
Figure 9. Monthly fault presence (metric 1) for ATU reheat coil valve stuck for FDD provider A .....	60
Figure 10. Monthly fault presence (metric 1) for ATU discharge airflow abnormal for FDD provider A .....	60
Figure 11. Monthly fault presence (metric 1) for ATU discharge air damper stuck for FDD provider B .....	61
Figure 12. Monthly fault presence (metric 1) for ATU reheat coil valve hunting for FDD provider B.....	62
Figure 13. Monthly fault presence (metric 1) for RTU cooling failure for FDD provider C.....	63
Figure 14. Monthly fault presence (metric 1) for RTU heating failure for FDD provider C.....	63
Figure 15. Average monthly fault presence (metric 2) for AHU faults from FDD provider A .....	66
Figure 16. Average monthly fault presence (metric 2) for AHU faults from FDD provider B .....	67
Figure 17. Comparison of average monthly fault presence for AHU faults between FDD providers A and B .....	68
Figure 18. Average monthly fault presence (metric 2) for ATU faults from FDD provider A .....	69
Figure 19. Average monthly fault presence (metric 2) for ATU faults from FDD provider B.....	70
Figure 20. Comparison of average monthly fault presence for ATU faults between FDD providers A and B .....	71
Figure 21. Average monthly fault presence (metric 2) for RTU faults from FDD provider C.....	72
Figure 22. Average monthly fault presence (metric 2) for RTU faults from FDD provider B.....	73
Figure 23. Mean number of faults per building per month distribution for FDD provider A.....	75
Figure 24. Mean number of faults per building per equipment per month distribution for FDD provider A .....	75
Figure 25. Mean number of faults per building per month distribution for FDD provider B.....	77
Figure 26. Mean number of faults per building per equipment per month distribution for FDD provider B .....	77
Figure 27. Mean number of faults per building per month distribution for FDD provider C.....	79
Figure 28. Mean number of faults per building per equipment per month distribution for FDD provider C .....	79
Figure 29. Normal and faulted zone air temperature sensors.....	83
Figure 30. Unoccupied temperature setback.....	85

Figure 31. Estimated and true test accuracy for different classification methods.....	90
Figure 32. Confusion matrices for four classification methods: (a) SVM, (b) LR, (c) KNN, (d) LDA .....	91
Figure 33. SVM 10-fold CV accuracy rates as a function of gamma values while $\text{cost}=10^7$ .....	92
Figure 34. TPR values for each class for all nine classification methods .....	94
Figure 35. FPR values for each class for all nine classification methods .....	95
Figure 36. Variable importance plots for different classification methods: (a) BA, (b) RF, (c) AD, (d) XGB .....	96
Figure 37. Confusion matrices for two classification methods after applying oversampling: (a) SVM, (b) LR .....	97
Figure 38. TPR values for each fault class before and after applying the oversampling: (a) SVM, (b) LR.....	98
Figure 39. FPR values for each fault class before and after applying the oversampling: (a) SVM, (b) LR.....	99

## **CHAPTER 1. Introduction**

This dissertation presents the study for a large-scale investigation of heating, ventilation, and air-conditioning (HVAC) fault prevalence in commercial buildings in the United States. HVAC faults are studied for several years, but there is no large-scale effort to quantify the magnitude of prevalence of individual HVAC faults. A comprehensive literature review was performed to understand the status of knowledge, key gaps, and potential value in doing research on quantifying the prevalence of HVAC faults in commercial buildings. Our literature review showed that there is only small number of studies that investigated the prevalence of faults in HVAC systems. Most of these studies had small sample sizes (small number of buildings and HVAC equipment), and mainly focused on specific climate zones or building types. Also, fault prevalence values change significantly between different studies. These limitations show there is a need for empirical data on the prevalence of HVAC faults in commercial buildings at the desired level of granularity.

Air handling units (AHUs), air terminal units (ATUs), and packaged rooftop units (RTUs) are very common in the United States commercial buildings. Unfortunately, the performance of these HVAC systems is often far from optimal. Fault detection and diagnostics (FDD) is a powerful tool that can monitor the operation of HVAC equipment and detect their problems. Although there is a significant growth in using of the FDD tools, there is still lack of reliable data about the prevalence of HVAC faults within the commercial buildings.

In this study, a multi-year dataset, provided by three commercial fault detection and diagnostics (FDD) company, including thousands of HVAC equipment from multiple climate zones and building types is analyzed to determine a range of HVAC fault prevalence metrics. Since each FDD company had different data formats, fault names, and fault reporting, data from each company are converted to a standard format, which is called binary daily fault (BDF) data. To quantitatively characterize the HVAC fault prevalence, several metrics are defined: monthly fault presence, average monthly fault presence, and mean number of faults per building per month. The most common AHU, ATU, and RTU faults in commercial buildings are found, and fault prevalence values calculated for different FDD providers are compared.

FDD software results inherently have a certain level of error, i.e., they might have false negatives and false positives. In order to better understand the correlation between fault reporting by FDD tools, and the true presence of faults, field verification has been carried out on a small subset of those buildings for which FDD data were collected. This allows verification of the presence of flagged faults, and checking for faults that were not flagged by the FDD tools. These site visits were conducted on two retail buildings that are each served by multiple packaged rooftop units (RTU).

A significant challenge of commercial FDD providers is to make sense of the building automation system (BAS) points, which provide the inputs for FDD algorithms. For some fault types, there are not sufficient sensors to be able to detect them. For example, non-condensable gas in the refrigerant and loose fan belt faults are not targeted by FDD providers, but can easily be found with field measurements. As efforts to standardize the

naming of BAS data and metadata (ontologies) progress, the opportunities for low-cost FDD are expanding. The issue of the accuracy and utility of diagnostic outputs is expected to become increasingly important in this environment. Lessons learned by comparisons, such as those produced in this study, may be valuable for potential adopters, users, and developers of FDD.

The basic elements of this study are addressed in the following sections.

### **1.1. Study Scope**

Since a comprehensive study on HVAC fault prevalence in commercial buildings in the United States is an ambitious task, this study focuses on selected mechanical systems that are commonly used in commercial buildings in the United States, rather than on all possible system types. Specifically, this study includes air handling units (AHUs), air terminal units (ATUs), and packaged rooftop units (RTUs). AHUs are a key element of the HVAC systems that are common in large commercial buildings. ATUs are one of the major building HVAC systems and directly affect the building zone comfort. RTUs serve the conditioning requirements for nearly half of the United States commercial building floor space (DOE, 2011).

### **1.2. HVAC Fault Prevalence Metrics**

There are many different approaches for expressing HVAC fault prevalence. For example, the HVAC fault occurrence rates could be specified on a monthly, seasonal or

annual time basis. In addition to time basis, there are several options for the physical boundaries that are drawn to determine HVAC fault prevalence. HVAC faults can be determined at different levels of component physical granularity. For example, temperature sensor frozen, return air temperature sensor frozen, or economizer sensor frozen. Moreover, HVAC fault prevalence can be determined at the building level, or the equipment level. For example, average number of faults per building per month, or average number of faults per unit per month.

In order to determine the HVAC fault prevalence metrics to be calculated in this study, several questions are established:

- For each month of the year, how often is HVAC fault type ‘X’ observed to be present?
- Which HVAC faults are most often observed to be present?
- How many HVAC faults are observed to be present each month for a given building?

### **1.3. Study Data**

The primary source of data for this study comes from commercial fault detection and diagnostics (FDD) software outputs. The reason is that commercial FDD software outputs can be obtained at relatively low cost, for a large number of buildings and HVAC systems. Since commercial FDD software outputs are subject to some level of error, i.e.,

they could have false negatives and false positives, a field study will be conducted to verify the FDD software results.

Although the study data includes the largest and most diverse dataset that can practically be obtained from commercial FDD providers, it is still not an ideal random sample representative of the population of United States commercial buildings as a whole.

However, it is sufficient to calculate HVAC fault prevalence metrics with an acceptable degree of precision and confidence. Ideally, field studies can use random selection to control different sources of selection bias, such as sampling bias or volunteer bias.

However, in this study, random selection is not a practical option.

Data requested from commercial FDD providers includes essential and preferred requirements shown in Table 1. Essential data requirements directly relate to the reporting of HVAC faults in a format that can be translated into a standardized format, along with descriptions of building type, location, and HVAC systems. Since flexibility is allowed on received FDD data format, each FDD provider's data will be translated into a standardized format for HVAC fault prevalence analysis.

#### **1.4. Verification**

Since the commercial FDD software results inherently contain a certain level of error, these results are complemented with a verification method based on direct field inspection of commercial buildings.

Table 1. Essential and preferred data required for study

<b>Essential data requirements</b>
Labeled HVAC faults with timestamps
HVAC system description labels (e.g., AHU, ATU, or RTU)
Total number of HVAC systems within the sample
Geographical description of building (e.g., zip code or climate zone)
<b>Preferred data requirements</b>
HVAC System details (e.g., age, configuration, or manufacturer)
Information about fault intensity and fault impact
Raw building automation system (BAS) data
Building characteristics (e.g., type or size)

Commercial FDD tools are designed to work with a practical set of constraints. For example, they might be designed to avoid false positives, even at the cost of imposing false negatives, and to focus on HVAC faults that are most cost-effective to detect.

One verification method employed in this study is to do manual analysis of building automation system (BAS) data to determine HVAC faults that were not detected by an FDD software. A subset of study data from a small number of commercial buildings is selected to check whether manual analysis of BAS data will provide additional insights.

Another verification method is manual inspection of commercial buildings, i.e., field testing. This method provides the highest fidelity verification. However, it has the highest cost. Therefore, it is applied to a small subset of commercial buildings. For this purpose,

a set of buildings are selected from among the buildings that are customers of the FDD providers.

HVAC field testing could be invasive and needs building owners to provide access to mechanical rooms, roof, and plenums above occupied zones, and to determine liability concerns. It would be more efficient to combine the site visit with an analysis of BAS data, and with short-term monitoring (e.g., two weeks) using dataloggers selected for this goal.

One important advantage of direct verification is that this method will make the overall findings far more credible for many of the potential users of the findings, i.e., FDD researchers, building managers, HVAC standards officials, etc.

Table 2 summarizes the strengths and weaknesses of the field verification.

Table 2. Strengths and weaknesses of the field verification

<b>Strength</b>	<b>Weakness</b>
It provides the true fault prevalence values	It is expensive and time consuming
It makes the results far more credible	Recruiting buildings for study is challenging

### **1.5. Data-Driven FDD for RTUs**

This study also proposes and demonstrates a data-driven fault detection and diagnostics (FDD) strategy for packaged rooftop units (RTUs) using statistical machine learning

classification methods. The fault detection and diagnostics task is formulated as a multi-class classification problem. Seven typical rooftop unit faults are discriminated against one another as well as the normal condition.

The performance of data-driven FDD is highly dependent on the quantity and quality of the available data. A persistent challenge has been the lack of reliable datasets to be used in the development of data-driven FDD methods (Granderson et al., 2020). Since experimental data for RTU systems is rare and expensive to obtain, we use a measurement data library with faulted and unfaulted systems at steady-state operation, generated with simulations based on Cheung & Braun (2013a, 2013b) to provide a rich training dataset for the classification models. The synthetic minority over-sampling technique is used to generate new artificial samples of the minority class in order to balance the dataset. We would like to emphasize that this data library is different than the FDD data we talked about in previous sections.

Nine classification methods including logistic regression (LR), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), K-nearest neighbors (KNN), bagging (BA), random forests (RF), AdaBoost (AD), XGBoost (XGB), and support vector machine (SVM) are applied to our dataset, and their performance is compared.

The performance of the classification methods for individual faults is also characterized using true positive rate and false positive rate statistical measures. The relative importance of input variables is also discussed.

## **1.6. Structure of the Thesis**

Chapter 2 gives an overview of the relevant research on the HVAC fault prevalence and FDD development for packaged rooftop units (RTUs).

Chapter 3 describes the methodologies used for analysis of FDD records to determine the HVAC fault prevalence in commercial buildings, field study verification, and development of machine learning based FDD methods for RTUs.

Chapter 4 presents the main results for the HVAC fault prevalence study using commercial FDD data and discusses the findings.

Chapter 5 presents the results of the evaluation of the commercial FDD tools using our filed study.

Chapter 6 discusses the performance of the various machine learning classification methods in detecting and diagnosing the typical faults in RTUs.

Chapter 7 summarizes the main conclusions and discusses the recommendations for future research.

## **CHAPTER 2. Literature Review**

In this chapter, studies on HVAC fault prevalence and fault detection and diagnostics (FDD) methods for packaged rooftop units (RTUs) are discussed in separate sections.

### **2.1. HVAC Fault Prevalence**

The goal of this literature review is to summarize studies which have characterized HVAC fault prevalence in commercial buildings.

Commercial buildings consume approximately 18% of total energy and 37% of electrical energy in the United States (EIA, 2018). HVAC systems are one primary end use in these buildings. Unfortunately, these systems often operate far from their optimal efficiencies because of design, installation, and operational problems. HVAC faults, or deviation from the expected operating conditions of an HVAC system or component, can increase a building's energy consumption and operational costs; may prevent the building from receiving needed services for HVAC; may negatively affect other interconnected energy systems; and could increase equipment maintenance or replacement costs (Ebrahimifakhar et al., 2020).

Fault detection and diagnostics (FDD) tools use building automation system (BAS) data to detect the presence of HVAC faults and support diagnosis of their root causes.

Applying FDD tools in commercial buildings and correction of the identified faults can save 9% of energy consumption (Kramer et al., 2020). Faults in the United States commercial buildings waste approximately 0.9–2.7 quads of energy annually (Frank et

al., 2019). However, this energy waste estimate is based on uncertain estimates of actual fault prevalence in the field. There is a lack of reliable data about which HVAC faults appear how frequently by building and system type. The purpose of this study is to fill the gap in the current state of knowledge about HVAC fault prevalence.

Researchers and FDD providers have largely focused on evaluating FDD performance building by building, and quantifying costs or other impacts. They often propose approaches that purport to improve the accuracy of fault detection, but by necessity will limit their investigations to simulated data (Li and O'Neill, 2019), a single building, or a small collection of buildings. A study exploring the use of automated methods for identifying “non-routine events” (possible faults) found success in streamlining measurement and verification processes, but recommended further work analyzing a larger set of buildings, including data from multiple real-world buildings and projects (Touzani et al., 2019). However, no unified dataset has been published on the observed prevalence of faults that could inform future studies. An exploratory study (limited to 12 buildings) that informed the current study was the first of its kind to attempt to harmonize FDD data from multiple buildings and identify the necessary steps and the barriers to doing so (Newman et al., 2020). One key challenge was the lack of a common taxonomy across individual studies. This was addressed by Chen et al. (2020, 2021) presenting a standardized taxonomy for HVAC faults related to air handling unit (AHU), air terminal unit (ATU), and rooftop unit (RTU) systems, which is described in chapter 3.

Several studies have been conducted for finding the frequency of faults in refrigeration and air conditioning systems. These studies collected data from two main data sources:

service records and field measurements. Service record sources include reports from service companies, insurance companies, building maintenance records, and manufacturers. Field measurements are obtained by system monitoring or technician inspection. Table 3 shows a list of the sampling methods used in the literature.

Table 3. Data sampling methods

<b>Service records</b>	Reports from service companies
	Reports from insurance companies
	Reports from building maintenance records
	Reports from manufacturers
<b>Field measurements</b>	System monitoring
	Technician inspection

Stouppe and Lau (1989) examined 15,760 failure records occurring between 1980 and 1987 on different air-conditioning (AC) and refrigeration systems in commercial buildings by analyzing insurance claims. They documented failures in compressors, fans, motors, and valves, and summarized major cause of failures for these elements. Based on different failures of service records, the percentage of failure among all failures can be inferred. However, failure prevalence at a certain point in time cannot be inferred because they do not mention the time when failures occurred. They found that in hermetic air conditioning systems 76.6% of faults were electrical, 18.9% of faults were mechanical,

and 4.5% of faults were attributed to a malfunction in the refrigerant circuit. The climate zone coverage of these service records is not given.

Hewett et al. (1992) studied energy savings that could be obtained by efficiency tune-ups on small commercial cooling systems in New England in the United States. They conducted field measurements on 25 cooling systems in 9 different sites. The focus of study was on refrigerant charge, duct leakage, and airflow faults. They found that 18 out of 25 units had refrigerant leakage fault. The fault prevalence can be calculated by dividing the number of faulty units by the total number of units.

Breuker and Braun (1998a) estimated the frequencies of occurrence and the service costs of different RTU faults by analysis of service records of a company from 1989 to 1995. About 6,000 service records were analyzed in order to determine the common faults in RTUs and estimate their energy impacts. The focus of study was on finding the percentage of a specific fault among all faults. They found that 60% of failures were electrical or control problems, while 40% of faults were mechanical. They also found that although compressor failures do not happen as frequently as other faults, they have the highest service costs in RTUs. The climate zone coverage of these service records is not also available.

Felts and Bailey (2000) monitored and analyzed over 250 RTUs installed in small commercial buildings in northern California in various climate zones. The entire monitoring period was three months in the summer, and each unit was monitored for three to five days. The measurement points were outdoor air, return air, mixed air, and supply air temperatures, power, and power factor. This study showed that 40% of the

RTUs were more than 25% oversized, and 10% of the RTUs were more than 50% oversized. It was also shown that economizers generally did not operate correctly. While the purpose of the study was to represent the whole 450,000 RTU customers in northern California, the sample size was not statistically representative.

Downey and Proctor (2002) collected and analyzed performance data on over 13,000 air conditioners in both residential and commercial buildings in California. Appropriate measurements were taken over 26 months, and the performance of the air conditioners against manufacturer's recommended refrigerant charge and evaporator airflow is evaluated. Their analysis concluded that 57% of the units had improper refrigerant charge, and 21% of the units had low airflow rate through the indoor coil.

Davis et al. (Davis, Baylon, et al. 2002; Davis, Francisco, et al. 2002) developed a field protocol to evaluate the performance of RTUs in small commercial buildings, and applied it to 30 RTUs in Oregon in the United State. The main focus of the protocol was on refrigerant charge, airflow, and economizer operation. Their field results showed that only 36% of the units had the correct amount of refrigerant charge. They also found that about 67% of the RTUs had evaporator airflow less than 350 SCFM/ton, and only less than 40% of economizers were fully functional.

Comstock et al. (2002) conducted a fault survey among four major American chiller manufacturers to determine the most common and expensive faults in chillers. Fault survey form included five categories: chiller type, service reason, fault type, corrective action, and service cost. A total of 509 service records were gathered for different types of chillers. The fault data were presented in forms of frequency of occurrence and repair

cost. They reported that most common faults happened in control box and starter sections. Refrigerant leakage was the second most commonly cited fault in chillers.

Cowan (2004) investigated data from 503 RTUs at 181 commercial buildings sites across five states, i.e., Oregon, Washington, Idaho, Montana, and California, gathered in four field studies. It was found that 46% of the units had improper refrigerant charge, 64% of the units had economizer problems, 42% of the units had airflow problems, 58% of the units had thermostat problems, and 20% of the units had sensor problems. By reviewing these four field studies results, they concluded that properly working thermostats and economizers offer the highest potential for energy saving.

Madani (2014) analyzed the fault reports provided to heat pump manufacturers and insurance companies in Sweden in both commercial and residential buildings. 8,659 fault reports from an insurance company, and about 37,000 fault reports from manufacturers were gathered. The results were presented in terms of the percentages of individual faults among all faults. This study showed that control and electronics faults are the most common and costliest faults in heat pump systems.

Liu et al. (1995) performed an optimization study of the HVAC operation at a seven-story building with a total of 123,000 ft<sup>2</sup> conditioned floor area in Texas in the United States. Field test was done on 3 AHUs and 210 terminal units. They found that terminal reheat leakage and excessive air flow are the main faults in the building. These faults increased energy consumption and thermal comfort complaints.

Yoshida et al. (1996) conducted a survey among HVAC experts in Japan to identify the ten most important faults in variable air volume (VAV) air handling systems based on their experience. The faults were ranked not only on frequency of occurrence, but also other factors such as environmental impacts, energy impacts, difficulty of detection, causing physical damage, and repair costs. The survey suggested that faults that occur in outdoor air damper and VAV box sections are fairly common.

Qin and Wang (2005) conducted a site survey in a large commercial building in Hong Kong with 1,251 pressure independent VAV terminal units over 14 days. Investigating the operation of the VAV terminal units showed that 261 VAV terminals (20.9%) were ineffective. Their investigation also showed that zone temperature sensor error and local direct digital control error are the most common faults in VAV terminals. In addition to presenting the percentage of the individual faults among all faults, the results showed that how many times each fault happened.

Gunay et al. (2019) developed a text-mining algorithm to extract information about fault frequency of HVAC systems from computerized maintenance management systems databases in Canada. The text mining algorithm was demonstrated using 26,992 service records gathered over seven years for a cluster of 44 buildings, and two years of service records for a central heating and cooling plant with four boilers and five chillers in a university campus. Analyzing the central heating and cooling plant dataset showed that the average annual warning/failure rate was 4.5 for a chiller, while it was 6.5 for a boiler. From the building cluster dataset, they found that approximately 50% of the warning/failure events were related to room/zone/floor level systems.

Shoukas et al. (2020) analyzed the fault data collected from FDD tools provided by four companies, representing over 28,000 RTUs, to determine the frequency of the reported faults. The fault data covered five different building types and multiple climate zones in the United States. Since different companies use different formats, fault definitions, diagnostics, and reporting, they were not able to compare between FDD tools, and results were presented separately for each data provider. They concluded that the frequency of the faults depends on the fault definitions and the diagnostics methods. They found that RTU faults occurred most commonly on economizer dampers, sensors, communications, and cooling systems.

Ebrahimifakhar et al. (2021) described a method to estimate the prevalence of HVAC faults in AHUs, ATUs, and RTUs. The study data collected from several fault detection and diagnostics (FDD) data providers, providing a large sample with a wide range of building types, geographical locations, and equipment types. They described how the data from different data providers can be processed and unified using a common taxonomy, and illustrated HVAC fault prevalence metrics that can provide insights using this type of data. They provided preliminary figures that illustrate their HVAC fault prevalence metrics.

Kim et al. (2021) performed an extensive literature review to summarize studies which have characterized fault prevalence in commercial buildings. They focused on three fault occurrence metrics in their review: fault prevalence, fault incidence, and percentage of fault among all faults. Table 4 shows the technical definitions of each fault occurrence metric. The provided data in this review can be listed as follows:

- Equipment type
- Fault type
- Fault occurrence metrics
- Fault impact
- Sampling method
- Building type
- Climate zone

Table 4. Fault occurrence metric definitions

<b>Metric</b>	<b>Definition</b>
Fault prevalence	Percentage of units with a given fault at a given severity at a single point in time.
Fault incidence	Frequency at which a fault occurs in a specific period of time.
Percentage of fault among all faults	Percentage of a specific fault as a subset of a greater collection of faults.

Their literature review showed that most of the building types defined in CBECS (EIA, 2002), and most of the climate zones are covered in the previous fault prevalence studies. They also found that fault occurrence metrics change significantly between different studies. For example, the prevalence of improper refrigerant charge fault was reported between 30% and 70% in different studies. An accurate comparison of fault prevalence

between different studies for this specific fault is challenging, since the fault intensity data is unknown and probably inconsistent. The review identified knowledge gaps in current literature on fault prevalence, and recommended a comprehensive study on HVAC fault prevalence in commercial buildings to fill these potential gaps:

- Fault prevalence data for different equipment types, e.g., coils, or dampers
- Fault prevalence data for different system types, e.g., AHUs, ATUs, or RTUs
- Fault prevalence data for different fault types
- Fault prevalence data for different building types
- Fault prevalence data for different climate zones
- Economic and energy impact data for different fault types

Table 5 summarizes the previous studies on HVAC fault prevalence. Our literature review shows that more work is required to understand the HVAC fault prevalence in commercial buildings. The previous studies have two primary limitations:

1. Small sample sizes: Most of the studies had small sample sizes, and focused on local regions and specific building types. These studies are not representative of the whole population of commercial buildings in the United States.
2. Out of date studies: Most of the studies were conducted before the year of 2010, and they are out of date. Since HVAC technologies have advanced in the past few years, new studies on HVAC fault prevalence are needed.

Table 5. Summary of HVAC fault prevalence studies

Study	Data Source	HVAC System	Sample Size	Coverage
Stouppe and Lau (1989)	Insurance company	AC and refrigeration systems	15,760 service records	Unavailable
Hewett et al. (1992)	Field inspection	Unitary cooling equipment	25 AC systems in 9 different sites	New England
Breuker and Braun (1998a)	Service company	RTUs	Around 6,000 service records	Unavailable
Felts and Bailey (2000)	System monitoring	RTUs	250 RTUs	Northern California
Downey and Proctor (2002)	Field measurements	AC systems	13,258 AC systems	California
Davis, Baylon, et al. (2002); Davis, Francisco, et al. (2002)	Field measurements	RTUs	30 RTUs	Oregon
Comstock et al. (2002)	Manufacturers	Chillers	509 service records	Unavailable
Cowan (2004)	Field measurements	RTUs	503 RTUs in 181 commercial buildings	Oregon, Washington, Idaho, Montana, and California

Madani (2014)	Manufacturers and insurance company	Heat pump systems	45,659 fault reports	Sweden
Liu et al. (1995)	Field measurements	AHU systems	3 AHUs and 210 terminal units	Texas
Yoshida et al. (1996)	HVAC expert survey	AHU systems	-	Japan
Qin and Wang (2005)	Site survey	VAV systems	261 variable air volume terminals	Hong Kong
Gunay et al. (2019)	Building maintenance records	Central heating and cooling plant	26,992 service records	Canada
Shoukas et al. (2020)	FDD tools	RTUs	Over 28,000 RTUs	Multiple US climate zones
Ebrahimifakhar et al. (2021)	FDD tools	AHUs, ATUs, and RTUs	964 AHUs, 18,896 ATUs, and 8,017 RTUs	Multiple US climate zones

## 2.2. RTU Fault Detection and Diagnostics

The purpose of this literature review is to summarize the existing fault detection and diagnostics (FDD) methods for packaged rooftop units (RTUs).

Packaged rooftop units are widely used in commercial buildings, serving approximately 52% of the total commercial building floor space in the United States (EIA, 2012). RTUs

are typically deployed in small commercial buildings, which means that they often don't have the high-level maintenance that chillers receive. They are believed to operate below their rated efficiencies because of faults introduced during installation or developed during operation (Feng et al., 2005). Therefore, significant energy is wasted annually because of the presence of operating faults in RTUs that go unnoticed by the owners or operators of the equipment. A solution to address this problem is to apply fault detection and diagnosis (FDD) tools to RTU systems so that important faults can be addressed promptly. FDD systems have the potential to reduce equipment downtime, energy costs, and maintenance costs, and improve occupant comfort and reliability (Braun, 1999; Li and Braun, 2007a, 2007b; Yuill and Braun, 2016, 2017).

Several FDD methods for RTU systems have been developed by researchers. Rossi and Braun (1997) developed a statistical, rule-based FDD method for vapor compression air conditioners with single-stage compressors, fixed-speed fans, and fixed orifice expansion valves. Their method only requires low-cost temperature and humidity measurements for detecting and diagnosing five common faults in air conditioners: (1) refrigerant leakage, (2) compressor valve leakage, (3) liquid-line restriction, (4) condenser fouling, and (5) evaporator fouling. A set of residuals is generated from the differences between measured values and predicted values obtained from a steady-state model for normal operation in the absence of faults. These residuals are used as inputs for both fault detection and diagnostic classifiers. The fault detection classifier uses the magnitude of the residuals to determine whether the system is normal or faulty, while the fault diagnostic classifier uses the directional change (sign) of the residuals to determine the type of fault.

Breuker and Braun (1998b) evaluated the performance of the FDD method developed by Rossi and Braun (1997). Their results show that the method can successfully detect and diagnose several faults over a wide range of operating conditions. However, the sensitivity of the method is affected by those operating conditions. They also quantified the minimum fault level at which the faults can first be detected and diagnosed.

Fault characteristics on RTU systems equipped with thermostatic expansion valves (TXVs) are different from those with fixed orifice expansion devices, and TXVs are used on most new systems. Therefore, Chen and Braun (2001) modified the original FDD approach proposed by Rossi and Braun (1997) to be applicable for RTU systems with TXVs. Their algorithm was designed to detect and diagnose seven typical faults in RTUs: (1) refrigerant leakage, (2) refrigerant overcharge, (3) compressor valve leakage, (4) liquid-line restriction, (5) condenser airflow reduction, (6) evaporator airflow reduction, and (7) non-condensable gas. To simplify the FDD approach, they proposed two easily implemented methods for isolating the faults: (1) sensitivity ratio method, and (2) simple rule-based method. Both methods required a smaller number of sensors and were successful in detecting and diagnosing faults at a range of operating conditions. However, neither method was able to distinguish between non-condensable and refrigerant overcharge faults.

Armstrong et al. (2006) developed an electrical signal based FDD technique for RTU systems. Their method requires high frequency current and voltage measurements. The electrical signals are measured and analyzed by a non-intrusive load monitor (NILM). Changes in the power signatures of the compressors and fans are used to detect and

diagnose common faults in air conditioners. This FDD method is minimally intrusive and complements other FDD approaches that are based on the temperature and humidity measurements.

Multiple simultaneous faults are probably common in packaged air conditioners, but none of the previous FDD approaches can deal with multiple faults that happen simultaneously. Li and Braun (2007a, 2007c, 2009a, 2009b, 2009c) developed a decoupling-based FDD method for multiple simultaneous faults in packaged air conditioners. The key to handling multiple simultaneous faults is to identify decoupled features that uniquely are affected by individual faults. However, direct measurement of some decoupling features is too expensive or otherwise problematic. Therefore, they developed virtual sensors that use simple models to estimate decoupling features from indirect low-cost measurements, such as temperature measurements of saturated fluids to estimate pressure.

Kim and Braun (2020) developed an FDD system for RTUs which incorporates virtual sensors and fault impact models for supporting the service decision making. The FDD system could detect the cause of the faults, and diagnose the fault intensities. Optimal fault detection thresholds were determined in such a way that maximize fault detection sensitivity and minimize false alarm rate. The decision for performing the RTU service could be based on assessing the capacity and COP degradation compared to the normal operation.

Ebrahimifakhar et al. (2020, 2021) proposed a data-driven strategy for fault detection and diagnostics in rooftop air conditioning units, based on machine learning classification

methods. Their strategy formulates the fault detection and diagnostics task as a multi-class classification problem. The focus of their study was on detecting and diagnosing the following common rooftop unit faults: refrigerant undercharge, refrigerant overcharge, compressor valve leakage, liquid-line restriction, condenser fouling, evaporator fouling, and non-condensable gas in the refrigerant. Ten-fold cross-validation technique was used to select tuning parameters for different classification methods. Their results demonstrate the potential of data-driven strategies to detect and diagnose common rooftop unit faults.

Bode et al. (2020) investigated whether machine learning FDD algorithms trained on an experimental dataset could be transferred to a real-world building dataset. They used experimental dataset of a unitary split system heat pump provided by the National Institute of Standards and Technology (NIST) to train their FDD algorithms. Their results showed that the trained FDD algorithms perform satisfactorily on the experimental dataset, but poorly on the real-world building dataset.

Most existing FDD approaches developed for packaged air conditioners use a rule-based approach. However, there is a large body of literature for data-driven FDD methods applied to other heating, ventilation, and air conditioning (HVAC) systems, especially chillers. Han et al. (2011a, 2011b) proposed a hybrid model that combines support vector machine (SVM) and genetic algorithm (GA) to detect and diagnose chiller faults. They adopted multi-class SVM as the FDD tool and selected GA for identifying the important sensors using a feature selection technique. Their FDD strategy was validated using experimental data from ASHRAE project 1043-RP (Comstock and Braun, 1999).

Zhao et al. (2013, 2014) proposed a pattern recognition-based method for detecting and diagnosing faults in chiller operations using support vector data description (SVDD), which is a novel one-class classification technique. Their results showed that the SVDD based FDD methods outperformed the principal component analysis (PCA) based FDD methods.

Li et al. (2016) developed a two-stage data-driven FDD approach to detect and diagnose chiller faults using linear discriminant analysis (LDA) method. The FDD task was transformed into a multi-class classification problem. At the first stage, a fault was detected and diagnosed, and at the second stage, its corresponding severity level was identified.

In summary, the majority of FDD methods developed and reviewed in this literature for RTUs are rule-based, and therefore, that is a need for developing “machine learning based” FDD for RTUs. Currently, there are only "two" research studies in this field: Ebrahimifakhar et al. (2020, 2021) and Bode et al. (2020) that both are reviewed in this section. We would like to mention that there are hundreds of machine learning based FDD studies but those are related to other HVAC systems like chillers, AHUs, ATUs, etc.

## **CHAPTER 3. Methodology**

### **3.1. Analysis of FDD Records as an Indicator of HVAC Fault Prevalence**

The primary source of data for this study comes from commercial fault detection and diagnostics (FDD) software tools. This is because commercial FDD software outputs can be obtained for a large number of buildings and HVAC systems at a relatively low cost. Since FDD software outputs are subject to error, i.e., they might have some level of false negative, false positive, and misdiagnosis rates, some of the FDD software outputs will be verified using manual inspection of buildings which will be explained in section 3.2.

#### **3.1.1. Data Overview**

The fault data received for this study is sourced from three commercial FDD software tools. The study dataset includes at least twelve months of data for each building. The study includes three classes of system: air handling unit (AHU), air terminal unit (ATU) and packaged rooftop unit (RTU), and includes analysis of components of these systems, such as a supply air temperature sensor for an AHU, or an economizer for an RTU.

Table 6 shows the number of buildings, HVAC systems, and daily fault records for each of the data providers. A “daily fault record” constitutes the presence of a specific fault on a unique piece of equipment on a single day. A fault flagged multiple times in a single day constitutes one daily fault. For example, an RTU flagged with a stuck economizer damper fault every hour in 2019 would generate 365 daily fault records in the study

dataset. During that same time period the same RTU could generate other daily fault records relating to other fault types.

Table 6. FDD data sources

<b>Data Source</b>	<b># of Buildings</b>	<b># of AHUs</b>	<b># of ATUs</b>	<b># of RTUs</b>	<b># of Daily Fault Records</b>
Provider A	131	964	18,896	0	4,473,881
Provider B	29	709	13,812	13	4,352,792
Provider C	131	0	0	2,162	2,861,910

Tables 7 and 8 show that the sample space of data obtained from these three providers represents multiple building types and climate zones.

Table 7. Number of buildings by building type

<b>Building Type</b>		<b>Health Care</b>	<b>Other</b>	<b>Office</b>	<b>Education</b>	<b>Mercantile</b>
<b># of Buildings</b>	<b>Provider A</b>	77	33	20	1	0
	<b>Provider B</b>	29	0	0	0	0
	<b>Provider C</b>	0	0	0	0	131

Table 8. Number of buildings by Building America climate zone

Climate Zone		Marine	Hot-Dry	Cold	Hot-Humid	Mixed-Humid	Mixed-Dry
# of Buildings	Provider A	60	56	8	5	2	0
	Provider B	0	21	7	0	0	1
	Provider C	12	18	26	37	37	1

### 3.1.2. Data Preparation

Curating and analyzing data from a number of different sources is complicated by the diversity of data formats, fault naming conventions, and metadata and file structures that the FDD software tools employ. The first, and most intensive, step is to prepare the data by cleaning and normalizing it by mapping it to a common fault taxonomy. Data preparation includes the following steps:

- Cleaning data to identify and resolve missing, mislabeled, empty fields, erroneous data, etc.
- Anonymizing data to ensure that any sensitive information that may identify buildings or data partners is removed.
- Normalizing data to a standard format using a common fault taxonomy.

Since data preparation and cleaning steps might change by FDD data provider, there is likely partner-specific customization that will be required at this step. However, once this

step is done, it is expected that all next steps will operate using a common analysis process.

Fault data from each partner are converted to a standard format, which is called binary daily fault (BDF) data. Table 9 shows a sample of BDF data. HVAC fault prevalence metrics are calculated from the BDF data.

Table 9. Standard binary daily fault (BDF) data

<b>Fault Record</b>	<b>Building ID</b>	<b>Equipment ID</b>	<b>Equipment Type</b>	<b>Date</b>	<b>Fault Name Mapped</b>
1	A1	A-AHU1	AHU	20190101	AHU-Heating- Coil_valve-Leakage
2	A15	A-AHU5	AHU	20190101	AHU-Cooling- Coil_valve-Stuck
3	B2	B-ATU17	ATU	20190102	ATU-Discharge_air- Damper-Stuck
4	B21	B-ATU24	ATU	20190102	ATU-Discharge_air- Airflow-Abnormal
5	C32	C-RTU3	RTU	20190103	RTU-Outdoor_air- Airflow-Abnormal
6	C114	C-RTU6	RTU	20190104	RTU-Mixed_air- Temperature_sensor- Frozen

### 3.1.3. Standardized Taxonomy for HVAC Faults

Each fault detection and diagnostics (FDD) tool uses different fault names to refer to the same fault in an HVAC system. For example, in one commercial FDD tool, an “economizer damper hunting” fault is reported to show a malfunctioning damper control, but in another tool, this fault may be reported as an “economizer damper short cycling” fault or an “unstable economizer damper” fault. Therefore, a unifying taxonomy for HVAC faults in AHUs, ATUs, and RTUs in commercial buildings was developed (Chen et al., 2020, 2021). The developed fault taxonomy contains 123, 44, and 107 unique fault names for AHUs, ATUs, and RTUs, respectively. Mapping functions were created for each FDD tool to convert their fault reports to this unifying taxonomy. Tables 10, 11, and 12 shows a selection of some of the HVAC faults in the taxonomy for AHUs, ATUs, and RTUs, respectively.

Table 10. Example list of the AHU faults in the developed taxonomy

Equipment	Component	Fault Name	Fault ID	Fault Type*
AHU	Air economizer	Outdoor air	AHU-Outdoor_air-	BB
		damper hunting	Damper-Hunting	
	Cooling coil	Cooling coil	AHU-Cooling-	CB
	valve	valve stuck	Coil_valve-Stuck	
	Outdoor air	Outdoor air	AHU-Outdoor_air-	CB
	temperature sensor	temperature sensor bias	Temperature_sensor-Bias	

\*BB=Behavior-based, CC=Condition-based

Table 11. Example list of the ATU faults in the developed taxonomy

Equipment	Component	Fault Name	Fault ID	Fault Type
ATU	Reheat coil	Reheat coil	ATU-Reheat-	CB
	valve	valve leakage	Coil_valve-Leakage	
	Discharge air	Discharge air	ATU-Discharge_air-	BB
	damper	damper hunting	Damper-Hunting	
	Discharge air	Discharge air	ATU-Discharge_air-	CB
	temperature	temperature	Temperature_sensor-	
	sensor	sensor drift	Drift	

Table 12. Example list of the RTU faults in the developed taxonomy

Equipment	Component	Fault Name	Fault ID	Fault Type
RTU	Air economizer	Economizer	RTU-Economizer-	CB
		damper stuck	Damper-Stuck	
	Supply air	Supply air	RTU-Supply_air-	CB
	temperature	temperature	Temperature_sensor-	
	sensor	sensor frozen	Frozen	
	Compressor	Compressor short cycling	RTU-	BB
			Refrigerant_circuit-	
			Compressor- Short_cycling	

There are three different fault categories based on how the faults are presented: condition-based, behavior-based, and outcome-based (Frank et al., 2019). Condition-based faults are improper or undesired physical conditions in HVAC systems such as stuck dampers, leaky valves, and biased sensors. Behavior-based faults present improper or undesired behavior during the operation of HVAC systems. Examples of behavior-based faults are economizer damper hunting, and simultaneous heating and cooling. Outcome-based faults are states in which an outcome or performance of the HVAC systems deviates from expected values, such as excessive energy consumption or insufficient ventilation rate. The HVAC fault taxonomy applied in the current project mainly includes condition-based and behavior-based faults, since they are most commonly used in FDD software tools.

An important feature of the taxonomy is that it supports flexible analysis based upon multiple levels of equipment class. For example, prevalence can be calculated for specific faults related to RTU supply air temperature sensors, supply air temperature sensors in general, temperature sensors in general, or sensors in general. Similarly, prevalence could be calculated for all heating faults, all damper faults, all stuck damper faults, and so on.

#### **3.1.4. Metric Definitions**

There are many different ways to express fault prevalence. To determine the priority HVAC fault prevalence metrics to be calculated in this study, we identified several questions that we expect to be of most interest to the study's target audience of FDD providers, users, regulators, and researchers. These questions include:

- What percentage of units are observed to be faulted at any given point in time?
- Which faults are most often observed to be present?
- How many faults are observed to be present each month for a given building?

To quantitatively characterize the HVAC fault prevalence, the following metrics are defined.

#### 3.1.4.1. Monthly Fault Presence (Metric 1)

This metric gives the percentage of equipment that experiences the presence of fault type ‘x’ on one or more days, for each month of the year, and is expressed as a percentage of all equipment. For a given piece of equipment, if fault ‘x’ is present for at least one day in a given month, that month is denoted as a “1” binary value, and considered one “fault\_month”. If the fault is observed to be present in multiple years for a given piece of equipment (e.g., present in February 2018 and in February 2019), each case will be considered a distinct value for this metric (e.g., February 2018 = 1, and February 2019 = 1, a total of two “fault\_months” for February).

This metric is calculated by:

$$monthly\_fault\_presence = \frac{fault\_months}{equipment\_months} \quad (1)$$

where fault\_months is the accumulated number of monthly fault occurrences for one type of fault in a calendar month across different years, and equipment\_months is the number of monitored pieces of a specific type of equipment in one calendar month, or in a

calendar month over a range of years. For example, if 100 dampers are monitored for two full years, the damper equipment\_months count for June would be 200.

The fault\_months is calculated by:

$$fault\_months = \sum_{i=1}^{num\_calender\_year} \sum_{j=1}^{equipment\_months} Fault\_OC_{month} \quad (2)$$

where  $Fault\_OC_{month}$  is the monthly fault occurrence. If there is at least one fault record in the FDD report within the month, then  $Fault\_OC_{month} = 1$ . The num\_calender\_year is the number of all years that may cover the time range of interest (e.g., the month of January appears in our dataset for a piece of equipment across two years, hence num\_calender\_year would equal 2).

Figure 1 illustrates the calculation of  $Fault\_OC_{month}$  and fault\_months under a selected time period. There are three AHUs, each monitored for two years. In January three out of six pieces of equipment had a fault flagged at least once during the month (so that  $Fault\_OC_{month} = 1$  for these three), hence there is a total of three fault\_months for January of six possible. This represents a monthly fault presence of 50 percent for January.

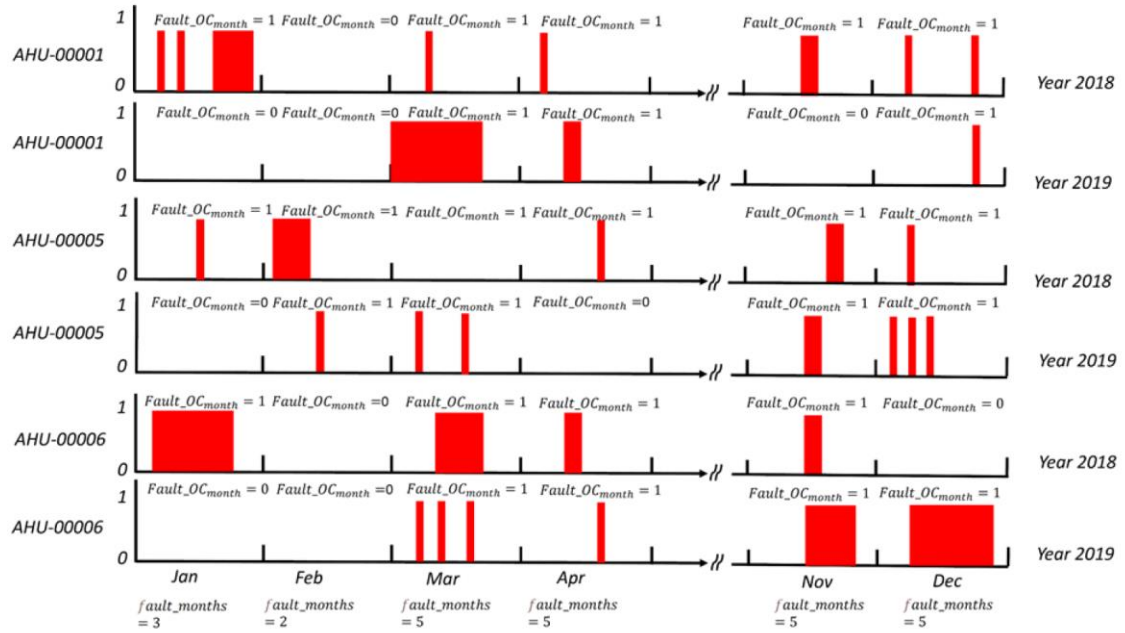


Figure 1. Graphical depiction of monthly fault presence (Metric 1)

### 3.1.4.2. Average Monthly Fault Presence (Metric 2)

Metric 2 is closely related to metric 1, and shows the percentage of equipment that experiences the presence of a given fault type on one or more days in a month, averaged across all months (whereas metric 1 presents a different fault presence value for each month). This metric shows which fault types are most often present in the data.

### 3.1.4.3. Mean Number of Faults per Building per Month (Metric 3)

This metric shows how many faults are observed to be present (at the building level) each month, among the set of faults considered in this study. The calculation steps of this metric are:

- 1) Establish total unique faults for each month, for one building.
- 2) Calculate mean value across all months for that building.
- 3) Repeat for all buildings.
- 4) Calculate mean of all building-specific mean values.

### 3.2. Field Study Verification

Since FDD software results inherently have a certain degree of error, these results are complemented with a verification process. The verification is based on a direct field inspection of buildings, and will determine the typical rates of correct and incorrect fault classification (false negatives and false positives).

Two buildings from among the buildings of the FDD provider C are selected for the field testing. These buildings have a total of 49 RTUs. Details of the site and equipment descriptions are shown in Table 13.

Table 13. Field study site and equipment descriptions

<b>Building ID</b>	<b>Building Type</b>	<b>Building Age</b>	<b># of RTUs</b>	<b>Capacity (Tons)</b>	<b>Make</b>
C1	Mercantile	18	27	2-13	AAON
C2	Mercantile	11	22	7.5-17.5	LENNOX

Field testing involves four types of data gathering:

- Collecting trend data from the BAS system of the subject buildings.
- Site visit measurements and observations (single point in time).
- Short term data logging.
- Proactive diagnostics (e.g., commanding a damper open to see if it moves).

The selected buildings have a BAS that monitors and controls their operation. There are 15 measurement points in the BAS, which are sampled every 1 minute. These measurements are shown in Table 14.

A monitoring plan is developed for site visits to collect data for RTUs which includes:

- Measurement points
- Measurement period for each point (about two weeks)
- Time interval measurement (1 minute)
- Measurement units

Table 15 shows the measurement points and their corresponding units in the monitoring plan.

Table 14. BAS measurements

<b>Data Point</b>	<b>Measurement Unit</b>
Outdoor air temperature	°F
Outdoor air relative humidity	
Return air temperature	°F
Mixed air temperature	°F
Supply air temperature	°F
Zone air temperature	°F
Zone air relative humidity	
Cooling setpoint	°F
Heating setpoint	°F
Occupancy	
Cooling command	
Heating command	
Economizer command	
Supply air fan command	
CO <sub>2</sub>	ppm

Table 15. RTU site visit data

<b>Data Point</b>	<b>Measurement Unit</b>
Outdoor air temperature	°F
Outdoor air relative humidity	
Return air temperature	°F
Mixed air temperature	°F
Supply air temperature	°F
Zone air temperature	°F
Zone air relative humidity	
Outdoor air damper position	
Compressor discharge pressure	psig
Compressor status	
Condenser temperature	°F
Supply fan current	Amps
Supply fan belt tension	Lbs

In this field study, we focused on detecting the following faults:

- Sensor faults
- Economizer damper stuck
- Non-condensable gas
- Abnormal supply fan belt tension
- No unoccupied temperature setback

In the following sections, the methods for detecting these faults are explained. Our goal is to compare our field study findings with the reports of FDD software tool already installed in the buildings.

Our literature review presented in section 2.2 showed that several methods have been developed for diagnosing the RTU faults during the past years. Most of these methods are rule-based and some of them are data-driven. There is still more work required to develop FDD methods that can be easily used in field studies. The faults selected for our field study are those that we find an easy and straightforward method for diagnosing them in our site visits from the literature. Since the purpose of our field study was for commercial FDD tools verification, it was important to us to select FDD methods that can correctly identify the presence or absence of these RTU faults.

### **3.2.1. Sensor Faults**

Reliable measurement is a key factor to ensure good performance of the RTUs. The RTU control depends on the sensor measurements. Unfortunately, sensor faults (bias, drift, frozen) are usually inevitable after the RTUs work for a while. Inaccurate measurements might lead to increase the system energy consumption or decrease the indoor air quality, although there are appropriate control algorithms. Therefore, detecting and diagnosing the RTU sensor faults is a significant task. In order to detect and diagnose the sensor faults, values in the time series of measurements from the building BAS data are compared to values in a corresponding time series of measurements from our data loggers (considered as real values). If the difference between the values is nearly constant over

time, the fault is categorized as a sensor bias. If the difference between the values is not nearly constant over time, the fault is categorized as a sensor drift. If BAS value is constant over time but real value changes, the fault is categorized as a sensor frozen. Supply air, mixed air, return air, and zone temperature sensors are checked in this study. Figure 2 shows our installed sensor beside the zone temperature and relative humidity sensor.



Figure 2. Our installed sensor beside the zone temperature and relative humidity sensor

### 3.2.2. Economizer Damper Stuck

RTU economizer includes duct/damper arrangement along with automatic control that allows to use outdoor air to decrease or eliminate the mechanical cooling. When there is a

cooling demand, and if the outdoor air condition is appropriate for economizing, unconditioned outdoor air will be introduced into the building to provide all of the cooling demand or supplement the mechanical cooling.

A key factor of the RTU economizers is the high limit switch that determines whether the outdoor air condition is favorable for economizing and enables or disables the economizer based on that. These are the most common high limit economizer controls:

- Fixed dry-bulb temperature
- Differential dry-bulb temperature
- Fixed enthalpy
- Differential enthalpy

RTU economizers often times fail to function properly because of damper, sensor, and other faults in the unit. Economizer faults might go completely unnoticed for long periods. If this happens, it can increase the system energy consumption. In this study, economizer damper stuck fault is detected by commanding the damper to open and close from the unit control, and visually check the damper operation. Figure 3 shows the economizer section of the one of the RTUs in the field.

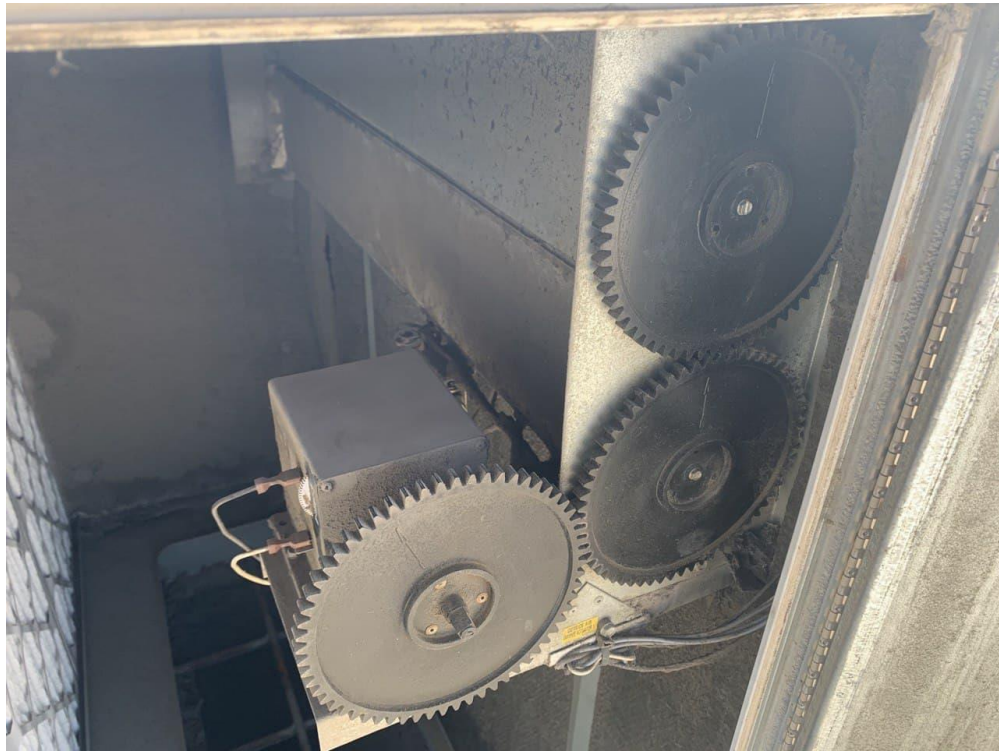


Figure 3. Economizer section of the one of the RTUs in the field

### 3.2.3. Non-Condensable Gas

Non-condensable gases, such as air and nitrogen, might enter a packaged rooftop unit (RTU) refrigeration cycle during installation or servicing if the refrigerant circuit is not fully evacuated prior to refrigerant charging. When the RTU is turned off, the non-condensable gas tends to accumulate in the unit's condenser (Li and Braun, 2007c). Since these gases do not condense, they can decrease the system efficiency. When non-condensable gas mixes with a two-phase refrigerant, it exerts an additional partial pressure. Therefore, the relationship between the refrigerant saturation temperature and pressure changes.

In this study, when the RTU is off and two-phase refrigerant exists within the unit's condenser, the difference between the measured condensing temperature and saturation temperature calculated from the measured compressor discharge pressure is used for detecting the presence of non-condensable gas.

#### **3.2.4. Abnormal Supply Fan Belt Tension**

Reduced tension in an RTU supply fan belt can lead the belt to slip and transfer less mechanical energy from the motor to the fan. Therefore, for a given motor speed, the airflow rate will be lower. Over tensioning decrease the belt and bearing life.

A belt tension checker is used for detecting and diagnosing this fault, and the following data is collected in the field:

- Belt cross section (e.g., BX)
- Belt Span
- Smallest sheave diameter
- Belt deflection force

#### **3.2.5. No Unoccupied Temperature Setback**

If the RTU systems are properly controlled during the unoccupied mode, it can significantly reduce the energy costs in commercial buildings. In order to detect and diagnose the deficiencies in occupancy scheduling, unoccupied temperature setback, the zone cooling and heating setpoints are checked using the building automation system

(BAS) data. During unoccupied hours, the zone cooling setpoint should be increased, and the zone heating setpoint should be decreased to achieve saving opportunities.

### **3.3. Machine Learning based FDD for RTUs**

Faults in RTU systems can be divided into two classes: (1) hard failures, and (2) soft faults. Hard failures happen abruptly and cause the RTU to stop functioning. Soft faults degrade the system performance, but permit continued operation of the system. Hard failures such as compressor failure frequently occur in RTU systems and are expensive to repair. However, they can be easily detected and diagnosed by inexpensive sensors. For example, a compressor failure can be easily detected and diagnosed by checking the inlet and outlet temperatures of the compressor. Hard failures are typically caused by extended periods of operation with a soft fault. Detecting and diagnosing soft faults, such as refrigerant leakage or heat exchanger fouling, is more challenging. These faults not only cause premature failure of components, but also reduce the operating efficiency or capacity of the system (Braun, 2003; Breuker and Braun, 1998a).

The focus of this study is on detecting and diagnosing the following soft faults:

- Refrigerant undercharge (UC)
- Refrigerant overcharge (OC)
- Compressor valve leakage (VL)
- Liquid-line restriction (LL)
- Condenser airflow reduction (CA)

- Evaporator airflow reduction (EA)
- Non-condensable gas (NC)

Fault detection means determining whether faults have happened in the system (Isermann, 2006; Katipamula and Brambley, 2005). An early detection of faults may provide valuable warning on arising problems in the system, and can be used to signal required attention by a technician. Fault diagnosis means determining the type, magnitude and location of the fault (Isermann, 2006; Katipamula and Brambley, 2005). Fault diagnosis is essential for eliminating or counteracting the faults. In some applications, fault diagnosis follows the fault detection, while in other applications, fault detection and diagnosis are performed in a single step (Katipamula and Brambley, 2005). In this study, we present a data-driven FDD approach that simultaneously detects and diagnoses faults in a single step using machine learning classification methods. Different fault types are discriminated against one another as well as the normal condition.

Over the past few decades, many classification methods have been developed and widely used in FDD for a wide range of engineering applications (Bishop, 2006; Fernández-Delgado et al., 2014; James et al., 2013). We used the following classic and state-of-the-art classifiers in this study:

- Logistic regression (LR)
- Linear discriminant analysis (LDA)
- Quadratic discriminant analysis (QDA)
- K-nearest neighbors (KNN)

- Bagging (BA)
- Random forests (RF)
- AdaBoost (AD)
- XGBoost (XGB)
- Support vector machine (SVM)

Data-driven FDD methods require a large and comprehensive database of training data that contain both normal and faulted conditions. However, experimental data for air conditioners is rare and difficult to obtain. This problem limits the size of the database available for a data-driven FDD approach. Therefore, a simulated database of model faults at steady state operation generated by Cheung and Braun (2013a, 2013b) is used to provide a rich training data for the machine learning classifiers. They used inverse modeling to generate a database of system performance under both faulted and normal conditions for vapor compression systems. Their simulation results have been validated against experimental data from previous research projects and found to perform well (Yuill et al., 2014).

In this study, a nominal three-ton RTU is used to test the proposed data-driven FDD strategy. The specifications of the RTU are shown in Table 16. The thermodynamic state of the system under normal or faulted conditions is characterized by the following comprehensive set of variables (features):

- Return air (evaporator inlet) dry bulb temperature ( $T_{RA}$ )
- Return air (evaporator inlet) wet bulb temperature ( $WB_{RA}$ )

- Supply air (evaporator outlet) dry bulb temperature ( $T_{SA}$ )
- Supply air (evaporator outlet) wet bulb temperature ( $WB_{SA}$ )
- Ambient air dry bulb temperature ( $T_{amb}$ )
- Liquid-line pressure ( $P_{LL}$ )
- Liquid-line temperature ( $T_{LL}$ )
- Suction pressure ( $P_{suc}$ )
- Suction temperature ( $T_{suc}$ )
- Compressor discharge pressure ( $P_{dischg}$ )
- Compressor discharge temperature ( $T_{dischg}$ )
- Condenser exiting air temperature ( $T_{air,ce}$ )
- Refrigerant saturation temperature in the evaporator ( $T_{sat,e}$ )
- Refrigerant saturation temperature in the condenser ( $T_{sat,c}$ )
- Compressor power ( $Power_{comp}$ )

Table 16. Specifications of the RTU system

System Type	Nominal Capacity [kW]	Refrigerant	Expansion Device	Condenser Type	Compressor Type	Operating Mode
RTU	10.6	R410A	Fixed Orifice (FXO)	Fin-tube	Scroll	Cooling

A schematic of a typical vapor compression refrigeration cycle with the selected state variables is shown in Figure 4.

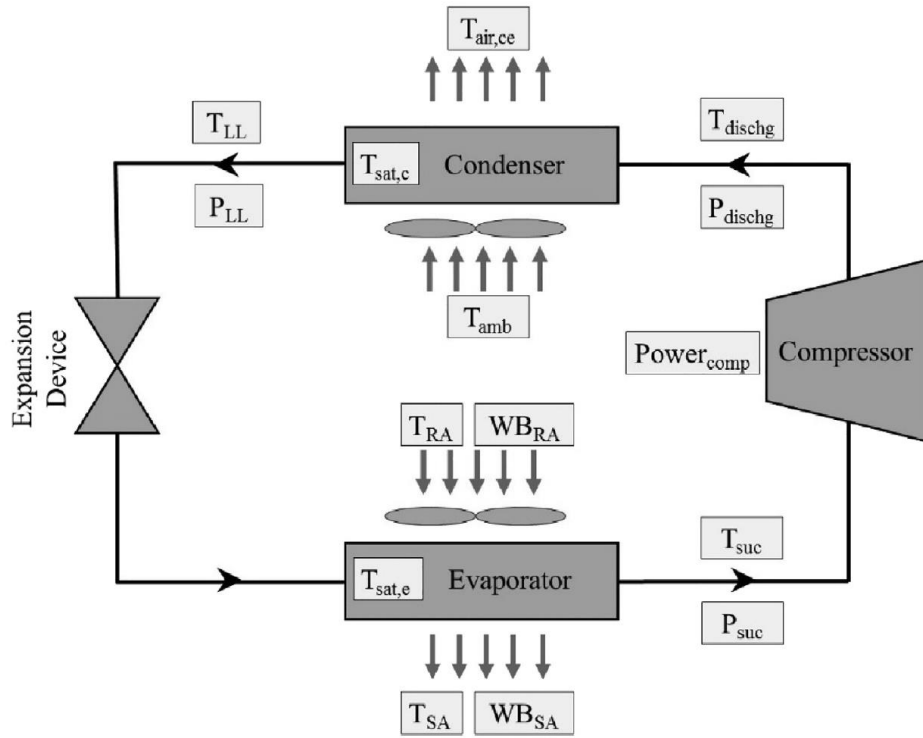


Figure 4. Schematic of a typical vapor compression refrigeration cycle

Classification involves building a statistical model for predicting a qualitative (categorical) output variable based on the input variables (features) (James et al., 2013). Our full dataset consists of fifteen input variables (such as  $T_{RA}$ ,  $WB_{RA}$ , and more) for 2851 observations (samples), and the output variable takes on eight possible categorical values: (1) UC, (2) OC, (3) VL, (4) LL, (5) CA, (6) EA, (7) NC, and (8) NF (no fault). This input variable set is comprehensive, containing a larger number of measured variables than is typically practical for an FDD application. However, the current work is intended to explore the overall feasibility of machine learning based fault classification for RTUs. Further work is required to examine the importance of the variables for classification to determine the most cost-effective reduced set.

Some statistical descriptors of the input variables of the total data set (the 2851 observations that contain both normal and faulted conditions) are shown in Table 17. Input variables often need to be scaled to avoid particular variables dominating the classification algorithm (Chiang et al., 2001). Therefore, we standardized the input variables so that each of them has a mean of zero and a standard deviation of one. This standardization ensures that each input variable is given equal weight before the application of the classification algorithm.

An examination of the minima and maxima in Table 17 shows that there appear to be outliers in the dataset. For example, the liquid line temperature extremes could not exist because they are outside the bounds of the ambient temperature. Although the original modeling effort included removal of many of the outliers that could be justified based upon physics or numerical solver troubles, the data set still contains errors. A desirable capability of the classifier in this particular application is robustness in the presence of unreliable data sets, since perfectly reliable modeling of fault impacts on measurable variables in these systems is not currently attainable.

By applying several different classification methods to our dataset, we intend to answer two important questions: which classification method is most effective for this task; and what is the potential of data-driven methods, generally for detection and diagnosis of faults. We randomly divided the 2851 observations into two parts, a training set containing  $2/3$  of the data points (1901 observations), and a test set containing the remaining 950 observations. Then we fitted a classification model using the training set, and evaluated its performance on the test set. In general, the accuracy of the classifier's

predictions applied to the training set is not important. We instead are concerned with how well the classifier predicts on the test set. A good classifier is one for which the test accuracy is largest. The main characteristics of the datasets are summarized in Table 18, which shows the name of the dataset, number of observations, and number within each class in the dataset.

Table 17. The statistics of the input variables of the total data set

<b>Input Variable</b>	<b>Unit</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Minimum</b>	<b>Maximum</b>
$T_{RA}$	°C	25.0	2.9	21.1	28.9
$WB_{RA}$	°C	17.6	4.3	12.8	23.9
$T_{SA}$	°C	14.9	3.7	1.0	27.4
$WB_{SA}$	°C	12.2	4.7	0.1	23.3
$T_{amb}$	°C	32.4	9.0	18.3	46.1
$P_{LL}$	kPa	2748.2	619.1	1603.0	4654.6
$T_{LL}$	°C	38.4	10.4	11.8	68.9
$P_{suc}$	kPa	1031.0	129.2	226.8	1368.6
$T_{suc}$	°C	13.2	6.1	-3.8	33.0
$P_{dischg}$	kPa	2842.7	613.2	1658.2	4767.7
$T_{dischg}$	°C	71.6	16.6	27.7	117.6
$T_{air,ce}$	°C	42.8	9.4	23.8	70.5
$T_{sat,e}$	°C	8.1	4.6	-34.1	18.1
$T_{sat,c}$	°C	44.5	9.6	23.9	68.9
$Power_{comp}$	W	2537.3	618.2	869.0	5021.1

Table 18. Fault class scenarios of the datasets

<b>Dataset</b>	<b># Observations</b>	<b># UC</b>	<b># OC</b>	<b># VL</b>	<b># LL</b>	<b># CA</b>	<b># EA</b>	<b># NC</b>	<b># NF</b>
Training Data	1901	252	353	367	233	263	239	163	31
Test Data	950	146	159	166	103	143	127	89	17
Total Data	2851	398	512	533	336	406	366	252	48

Each classification method may have several tuning parameters. Proper tuning parameter selection is an important issue for good predictive performance. Tuning parameters are usually selected with the k-fold cross-validation (CV) technique (Bishop, 2006; James et al., 2013). It involves randomly splitting the available training data into k folds of approximately equal size. The first fold is used for the validation set (hold-out set), and the remaining k-1 folds are used for fitting the model. The accuracy of the fitted model is then calculated on the validation set. This process is repeated for all k possible choices for the validation set. The accuracy values from the k runs are then averaged to calculate the k-fold CV accuracy. k-fold CV accuracy provides an estimate of the test accuracy associated with a given classification model.

We used 10-fold CV technique to find optimal tuning parameters for each classification method. We selected a grid of tuning parameter values, and calculated the 10-fold CV accuracy for each set of values, as described earlier. We then selected the tuning parameter values for which the 10-fold CV accuracy is largest. After the optimal tuning parameters values were found, we refitted the model on the full training dataset to generate the final classifier (James et al., 2013). For bagging (BA) and random forests

(RF) classification methods, we used out-of-bag (OOB) accuracy (James et al., 2013) to find optimal values for tuning parameters. This is a technique to estimate the test accuracy of a classification model, without the need to do k-fold CV.

To evaluate the performance of classification models in detecting and diagnosing faults, the prediction results can be shown as a two-dimensional confusion matrix (Witten and Frank, 2005), which compares the actual and predicted classifications. Each matrix element indicates the number of test observations, with the actual (true) class in rows and the predicted class in columns. The diagonal elements of a confusion matrix show the correct predictions, while the off-diagonal elements show the incorrect predictions and how they were misclassified.

Table 19 shows the confusion matrix for a three-class classifier (as an example), where TP (true positive) denotes the number of the cases for which a specific class actually happened and the classifier predicted that, TN (true negative) denotes the number of the cases for which a specific class did not happen and the classifier did not predict that, FP (false positive) denotes the number of the cases for which a specific class did not happen but the classifier predicted that, and FN (false negative) denotes the number of the cases for which a specific class actually happened but the classifier did not predict that.

Table 19. An illustrative confusion matrix with three classes

		<b>Predicted Class</b>		
		Class 1	Class 2	Class 3
<b>Actual (True) Class</b>	Class 1	$a_{11}$ (TN)*	$a_{12}$ (FP)	$a_{13}$ (TN)
	Class 2	$a_{21}$ (FN)	$a_{22}$ (TP)	$a_{23}$ (FN)
	Class 3	$a_{31}$ (TN)	$a_{32}$ (FP)	$a_{33}$ (TN)

\* Those in () are for class 2 (as an example)

The overall accuracy rate (OAR) is typically used as a simple measure for assessing the overall performance of a classifier. It is the number of correct predictions (the diagonal elements of the confusion matrix) divided by the total number of observations. However, class-specific performance is also important in fault detection and diagnosis, where the terms true positive rate (TPR) and false positive rate (FPR) characterize the performance of a classifier for individual classes. For a given class, TPR is the percentage of the happened observations that are correctly predicted, and FPR is the percentage of the non-happened observations that are predicted as happened. TPR and FPR can be calculated as follows:

$$TPR = \frac{TP}{TP + FN} \quad (3)$$

$$FPR = \frac{FP}{FP + TN} \quad (4)$$

## **CHAPTER 4. HVAC Fault Prevalence Results & Discussion**

A total of 11,688,583 daily fault records of AHUs, ATUs, and RTUs were analyzed from the three FDD providers. Values for metrics 1 to 3 have been generated.

### **4.1. Monthly Fault Presence (Metric 1) Results**

As we explained earlier, metric 1 gives the percentage of equipment that experiences the presence of fault type ‘x’ on one or more days, for each month of the year, and is expressed as a percentage of all equipment. The metric 1 results for AHUs, ATUs, and RTUs are presented in the following.

#### **4.1.1. AHU Results**

FDD provider A has 964 AHUs, and 46 unique AHU faults are successfully mapped to our fault taxonomy. FDD provider B has 709 AHUs, and 28 unique AHU faults are mapped.

Figure 5 shows the monthly fault presence (metric 1) for “AHU simultaneous heating and cooling” which is a behavior-based fault for FDD provider A. The prevalence rate has a range between 4% and 8%. This fault has lower overall rate in summer, and is likely to be correlated to season.

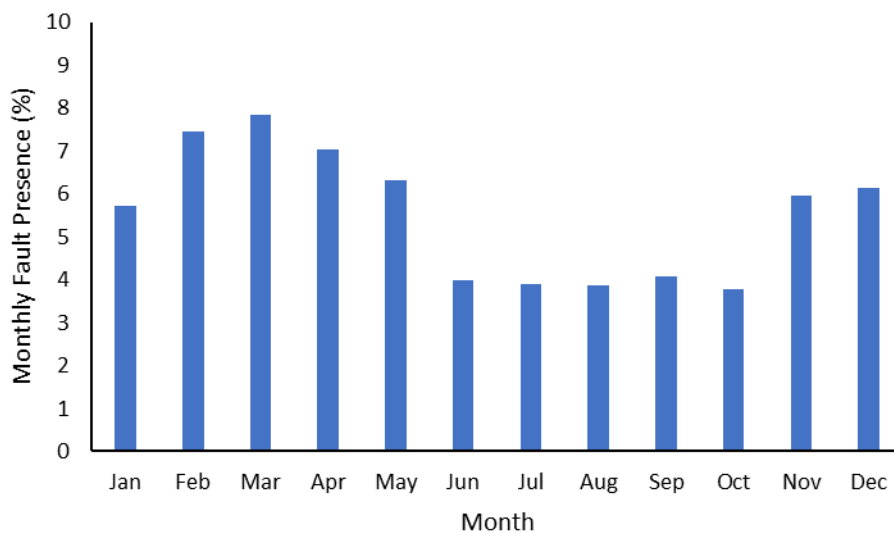


Figure 5. Monthly fault presence (metric 1) for AHU simultaneous heating and cooling for FDD provider A

Figure 6 shows the monthly fault presence for “AHU mixed air temperature abnormal” for FDD provider A. This fault is also behavior-based. The prevalence rate has a range between 18% and 26%. There is no obvious seasonal trend for this fault.

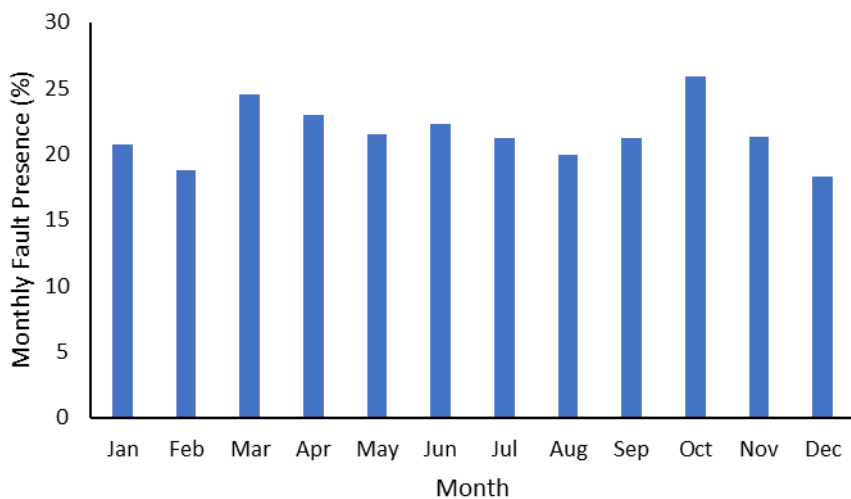


Figure 6. Monthly fault presence (metric 1) for AHU mixed air temperature abnormal for FDD provider A

Figure 7 shows the monthly fault presence for “AHU heating coil valve leakage” which is a condition-based fault for FDD provider B. The prevalence rate has a range between 1% and 4%. This fault has lower overall rate in summer, and this could be because of reduced usage of heating systems in summer.

Figure 8 shows the monthly fault presence for “AHU heating coil valve hunting” for FDD provider B. This is a behavior-based fault. The prevalence rate has a range between 1% and 10%. There is also a seasonal trend for this fault.

More examples of AHU faults for FDD providers A and B are given in appendix A.

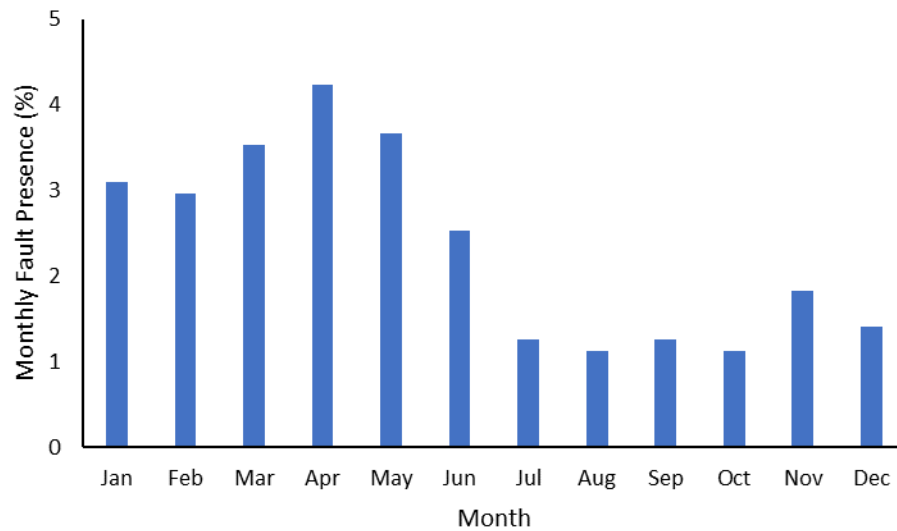


Figure 7. Monthly fault presence (metric 1) for AHU heating coil valve leakage for FDD provider B

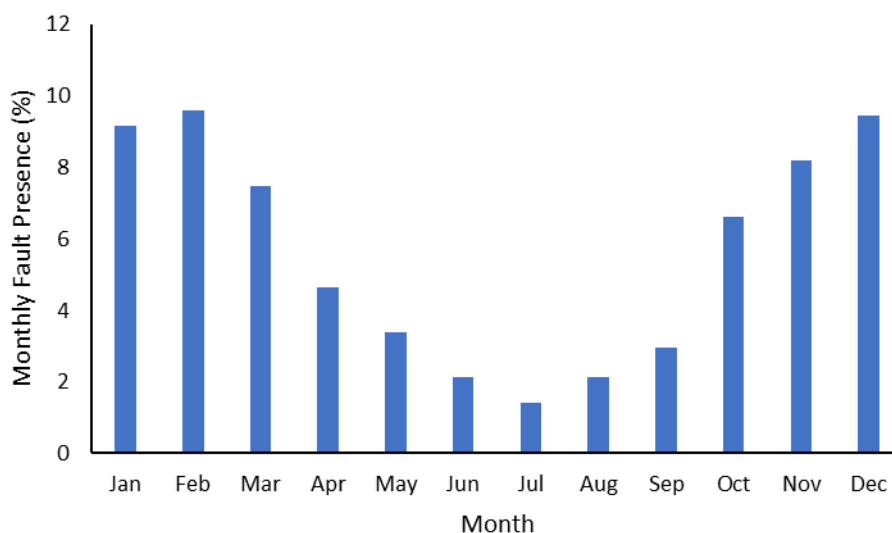


Figure 8. Monthly fault presence (metric 1) for AHU heating coil valve hunting for FDD provider B

#### 4.1.2. ATU Results

FDD provider A has 18,896 ATUs, and 17 unique ATU faults are successfully mapped to our fault taxonomy. FDD provider B has 13,812 ATUs, and 13 unique ATU faults are mapped.

Monthly fault presence (metric 1) for “ATU reheat coil valve stuck” for FDD provider A is shown in Figure 9. This is a condition-based fault. The prevalence rate has a range between 6% and 8%. This fault does not show a seasonal trend.

Monthly fault presence for “ATU discharge airflow abnormal” for FDD provider A is shown in Figure 10. This is a behavior-based fault. The prevalence rate has a range between 10% and 13%. This fault also does not show a seasonal trend.

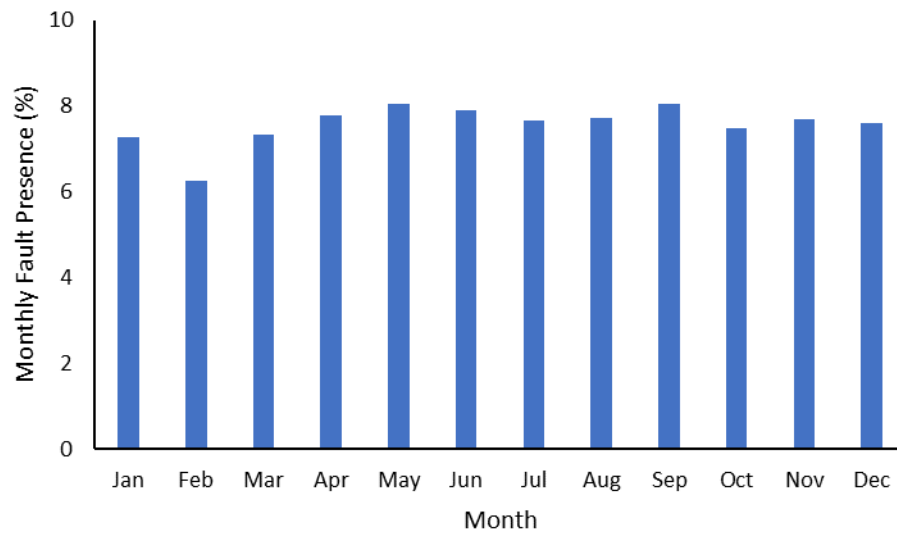


Figure 9. Monthly fault presence (metric 1) for ATU reheat coil valve stuck for FDD provider A

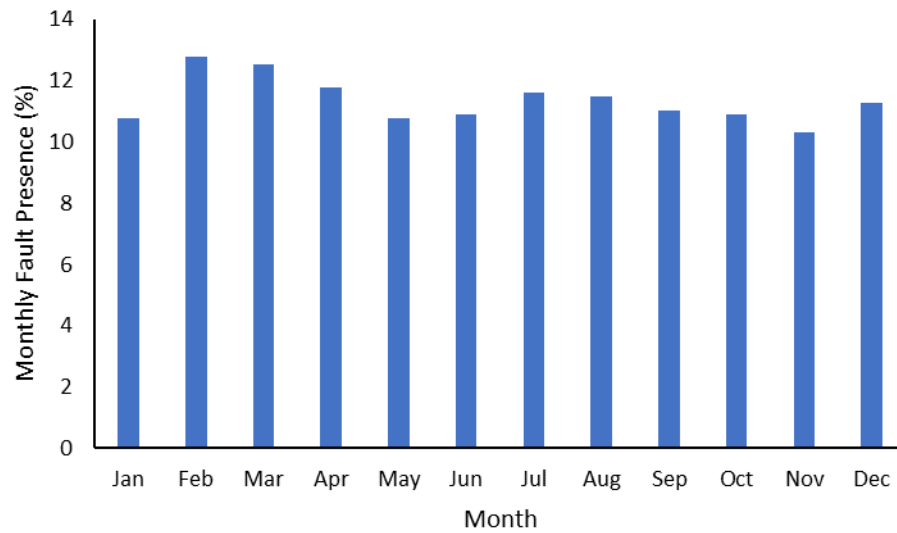


Figure 10. Monthly fault presence (metric 1) for ATU discharge airflow abnormal for FDD provider A

Monthly fault presence for “ATU discharge air damper stuck” for FDD provider B is shown in Figure 11. This is a condition-based fault. The prevalence rate has a range between 7% and 13%. This fault shows an apparent seasonal trend.

Monthly fault presence for “ATU reheat coil valve hunting” for FDD provider B is shown in Figure 12. This is a behavior-based fault. The prevalence rate has a range between 12% and 19%. This fault shows a clear seasonal trend.

More examples of ATU faults for FDD providers A and B are given in appendix B.

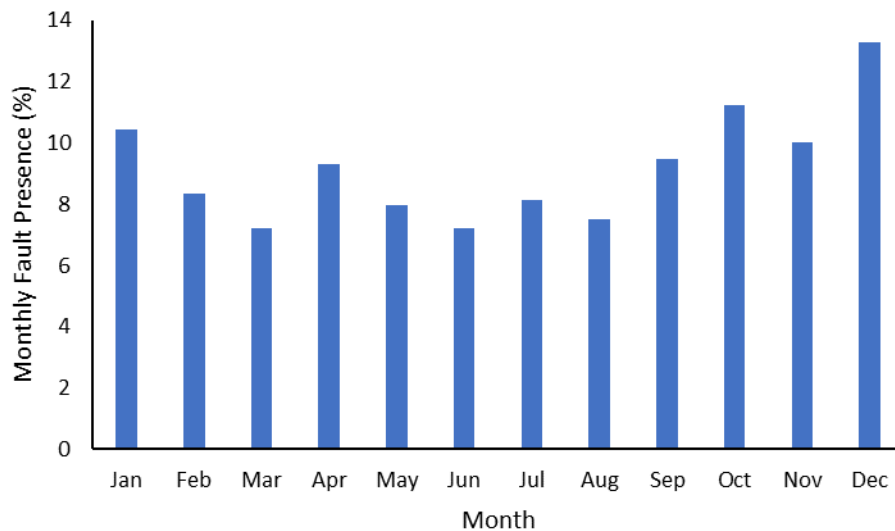


Figure 11. Monthly fault presence (metric 1) for ATU discharge air damper stuck for FDD provider B

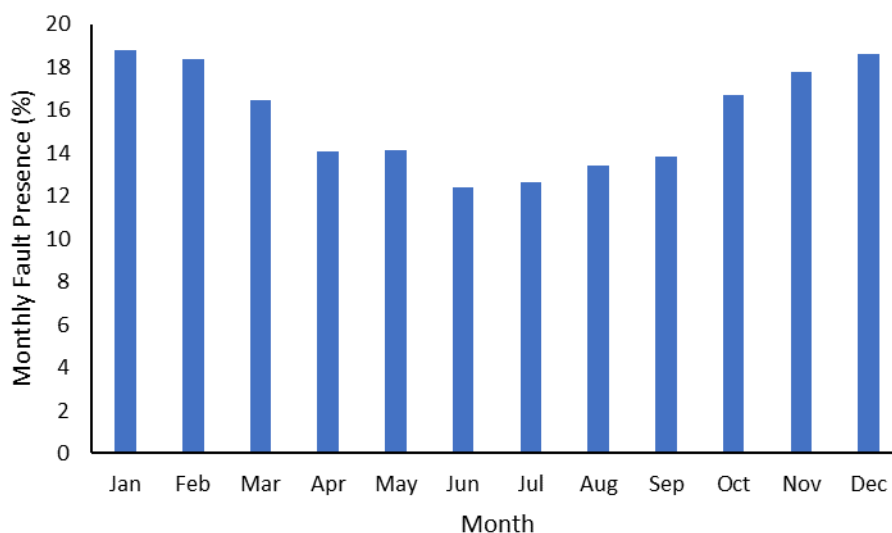


Figure 12. Monthly fault presence (metric 1) for ATU reheat coil valve hunting for FDD provider B

#### 4.1.3. RTU Results

FDD provider B has only 13 RTUs, and 3 unique RTU faults are successfully mapped to our fault taxonomy. FDD provider C has 2,162 RTUs, and 38 unique RTU faults are mapped.

Figure 13 shows the monthly fault presence (metric 1) for “RTU cooling failure” for FDD provider C. This is an outcome-based fault. The prevalence rate has a range between 4% and 19%. This fault shows a genuine seasonal trend.

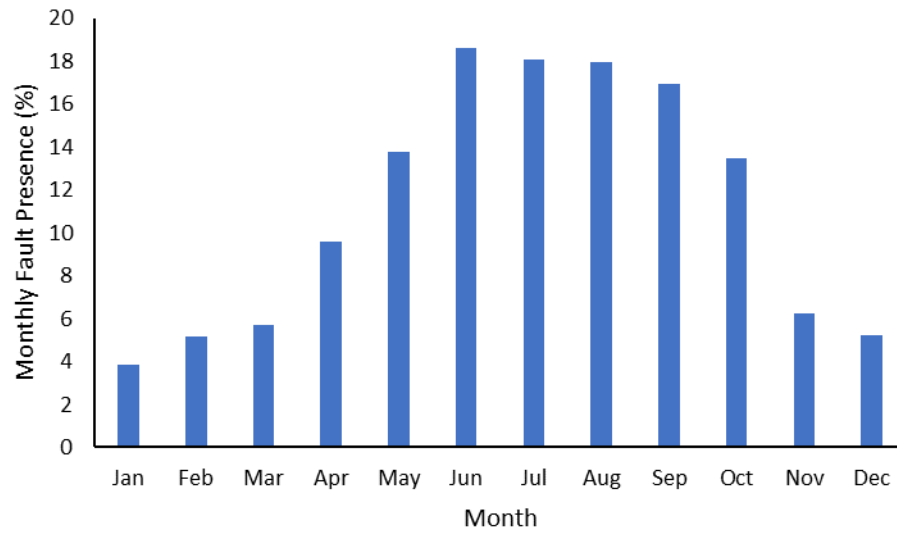


Figure 13. Monthly fault presence (metric 1) for RTU cooling failure for FDD provider C

Figure 14 shows the monthly fault presence for “RTU heating failure” for FDD provider C. This is an outcome-based fault. The prevalence rate has a range between 0% and 12%. This fault also shows a genuine seasonal trend.

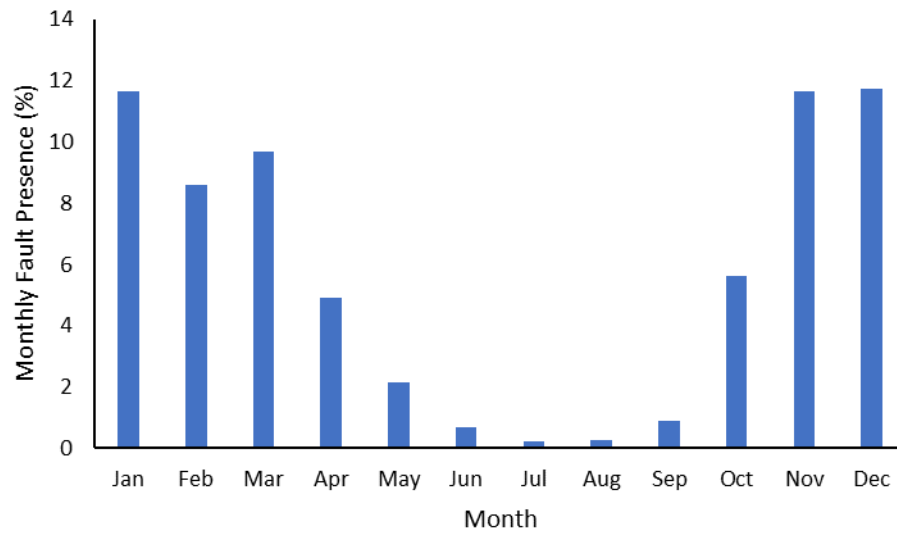


Figure 14. Monthly fault presence (metric 1) for RTU heating failure for FDD provider C

## **4.2. Average Monthly Fault Presence (Metric 2) Results**

As we explained earlier, metric 2 is closely related to metric 1, and shows the percentage of equipment that experiences the presence of a given fault type on one or more days in a month, averaged across all months. This metric is a useful way to sort the relative prevalence of all individual fault types, and can also help in understanding the most problematic system components (e.g., dampers, sensors) or functional elements (e.g., cooling, heating). The metric 2 results for AHUs, ATUs, and RTUs are presented in the following.

### **4.2.1. AHU Results**

Figure 15 shows the average monthly fault presence (metric 2) for AHU faults for FDD provider A. There are 46 unique AHU faults in total. “Missed control optimization opportunity”, “sensor frozen”, and “mismatch between supply air temperature and its setpoint” are the most common faults from the FDD provider A representing faults in 28%, 27%, and 26% of the AHUs, respectively. “Missed control optimization opportunity” is not strictly a fault, and relates to missing the following opportunities:

- Supply air temperature reset
- Static pressure reset
- Setback schedule

Figure 16 shows the average monthly fault presence for AHU faults for FDD provider B. There are 28 unique AHU faults in total. “Missed control optimization opportunity”,

“mismatch between supply air temperature and its setpoint”, and “mixed air temperature sensor fault” are the most common faults from the FDD provider B representing faults in 58%, 40%, and 35% of the AHUs, respectively. As can be seen, FDD provider A and B have two common faults between their three most prevalent faults. “Missed control optimization opportunity” relates to missing the following opportunities:

- Supply air temperature reset
- Supply air pressure reset
- Minimum outdoor airflow reset

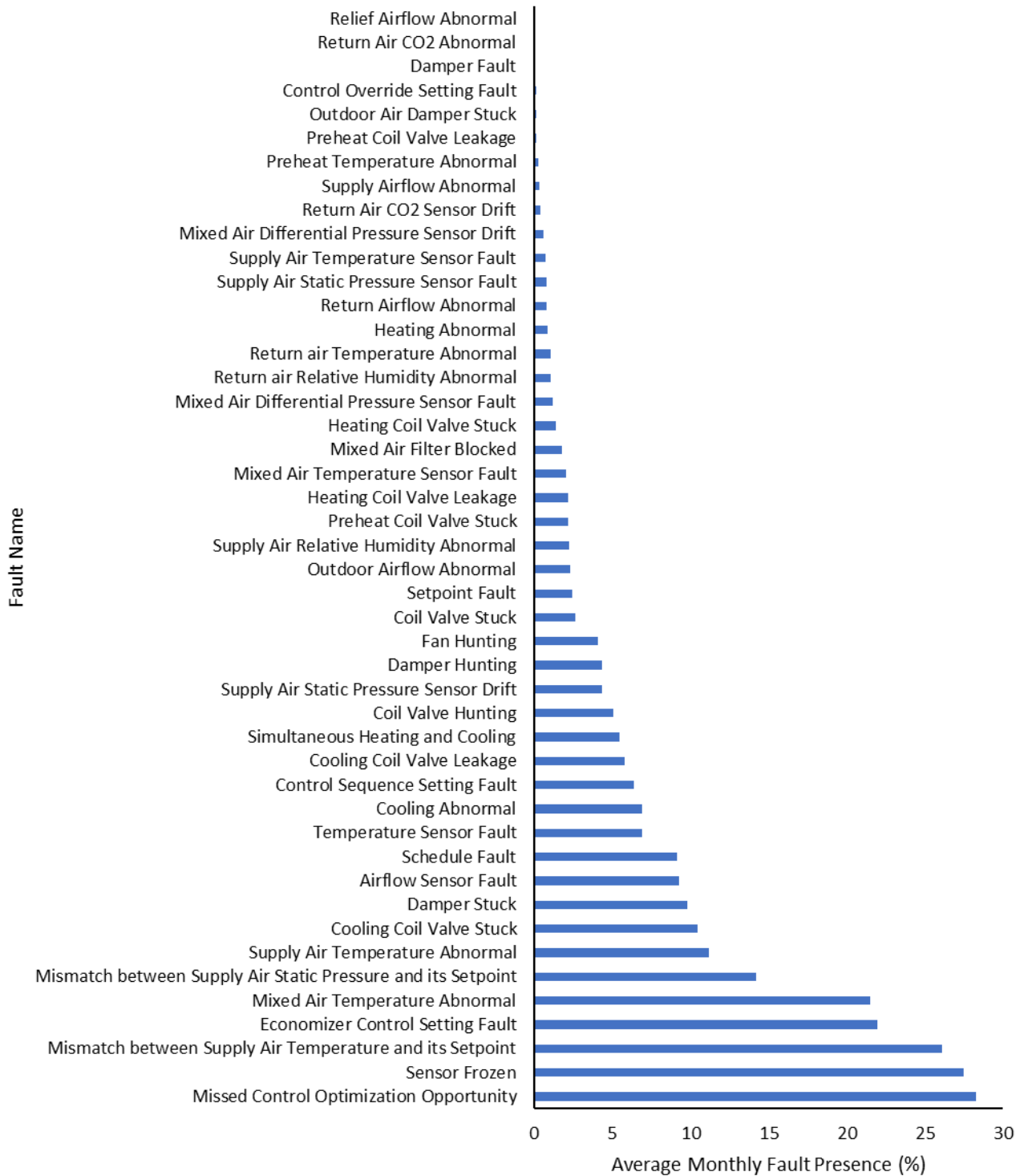


Figure 15. Average monthly fault presence (metric 2) for AHU faults from FDD provider A

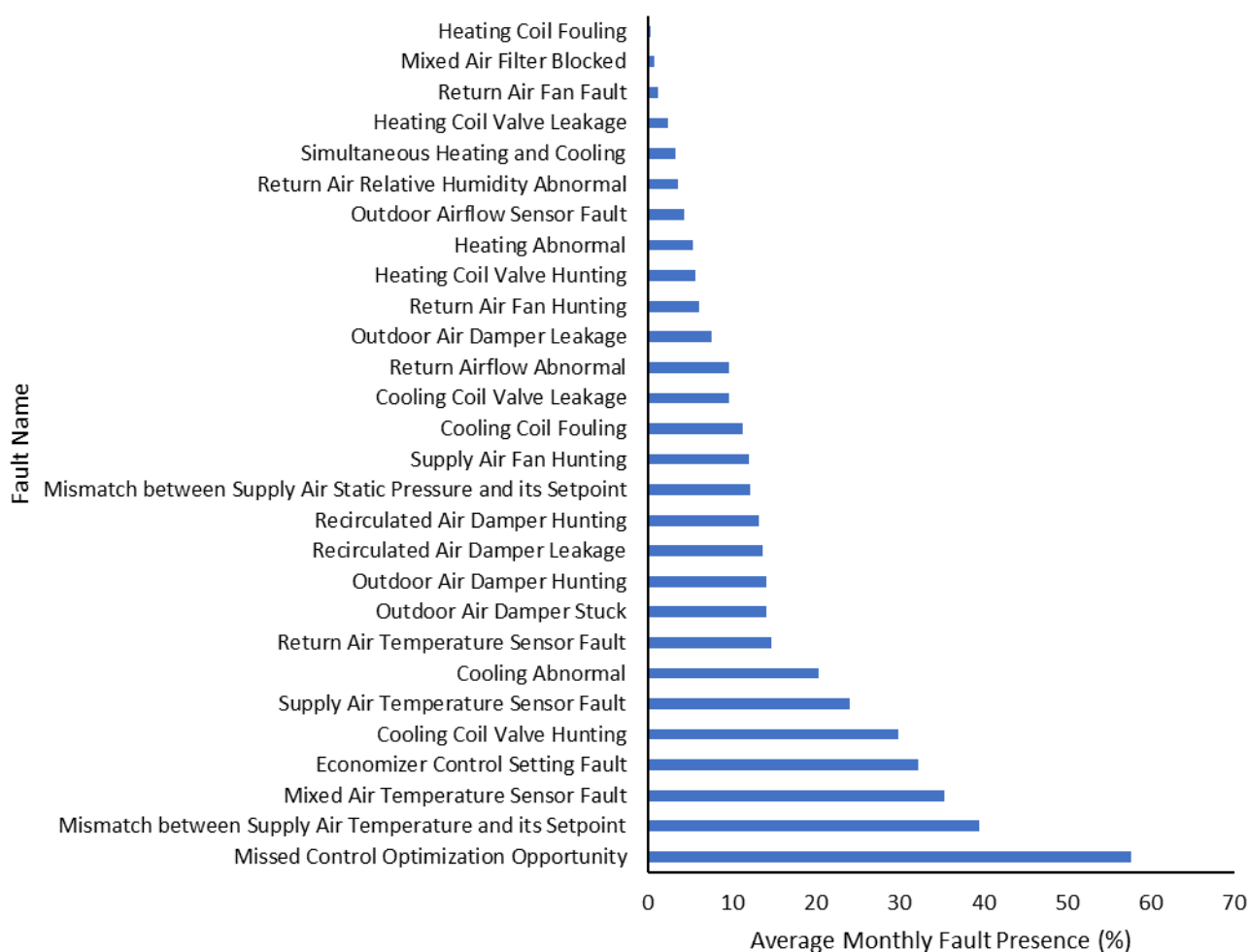


Figure 16. Average monthly fault presence (metric 2) for AHU faults from FDD provider B

Figure 17 shows the comparison of average monthly fault presence for AHU faults between FDD providers A and B. While for each fault there is a wide range provider to provider, this varied fault to fault. FDD provider A has systematically lower prevalence rates. Difference in scale could be due to FDD thresholding or maintenance vigor. Also, there are several potential fault prevalence drivers such as building type and climate zone

that are not the same between the FDD providers. As we explained in chapter 2, since different companies use different formats, fault definitions, diagnostics, and reporting, shoukas et al. (2020) were not able to compare between FDD tools, and they presented their results for RTU fault prevalence separately for each data provider. Standard fault definitions like what we used in this study can help to overcome this barrier.

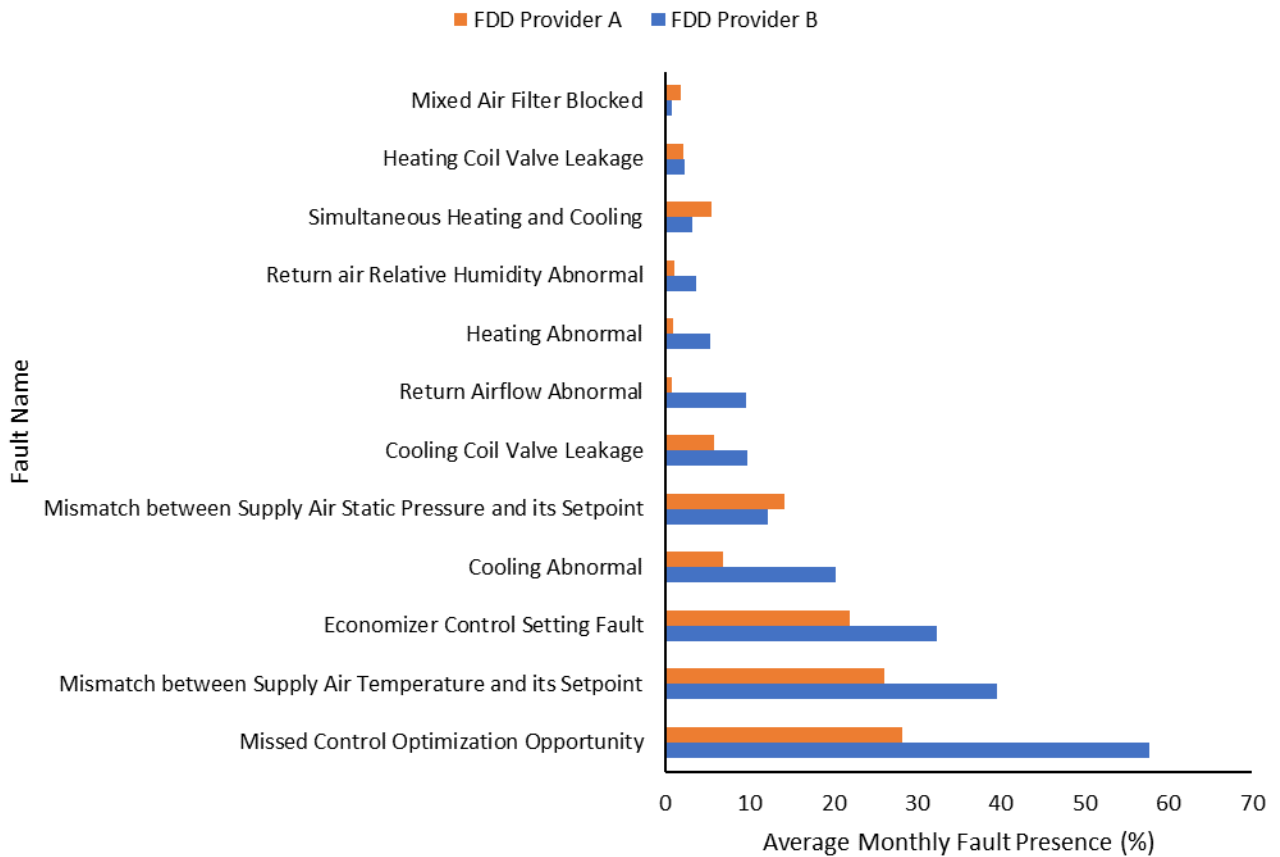


Figure 17. Comparison of average monthly fault presence for AHU faults between FDD providers A and B

#### 4.2.2. ATU Results

Average monthly fault presence (metric 2) for ATU faults for FDD provider A is shown in Figure 18. There are 17 unique ATU faults. “Zone temperature abnormal”, “sensor frozen”, and “discharge air temperature abnormal” are the most common faults from the FDD provider A representing faults in 23%, 14%, and 14% of the ATUs, respectively.

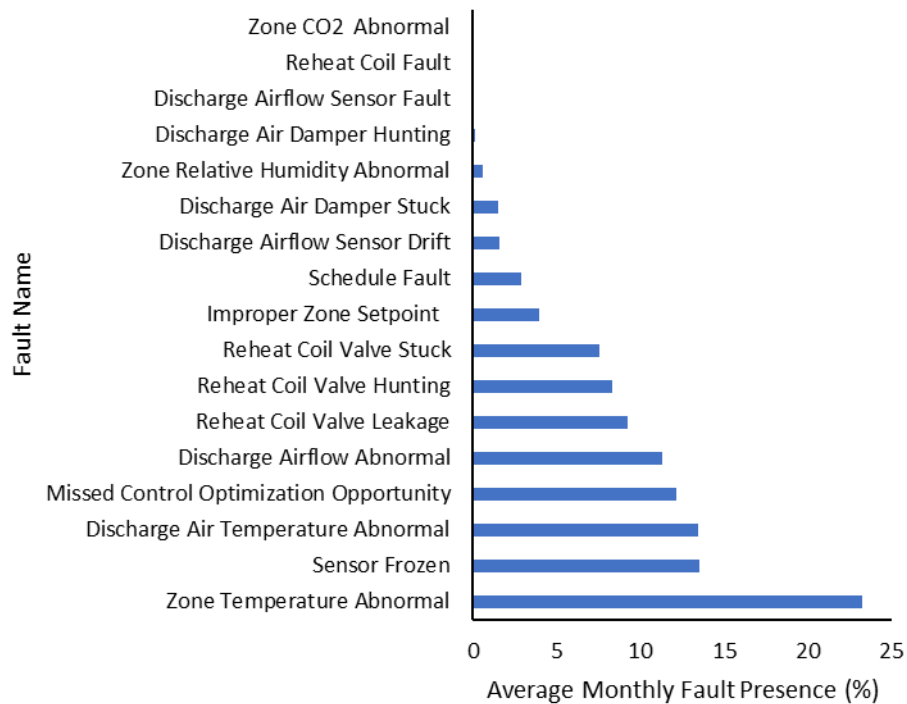


Figure 18. Average monthly fault presence (metric 2) for ATU faults from FDD provider A

Average monthly fault presence (metric 2) for ATU faults for FDD provider B is shown in Figure 19. There are 13 unique ATU faults. “Zone temperature abnormal”, “control sequence setting fault”, and “discharge airflow abnormal” are the most common faults

from the FDD provider B representing faults in 50%, 29%, and 23% of the ATUs, respectively. As can be seen, “zone temperature abnormal” is the most common ATU fault for both FDD providers A and B, although the prevalence rate is different.

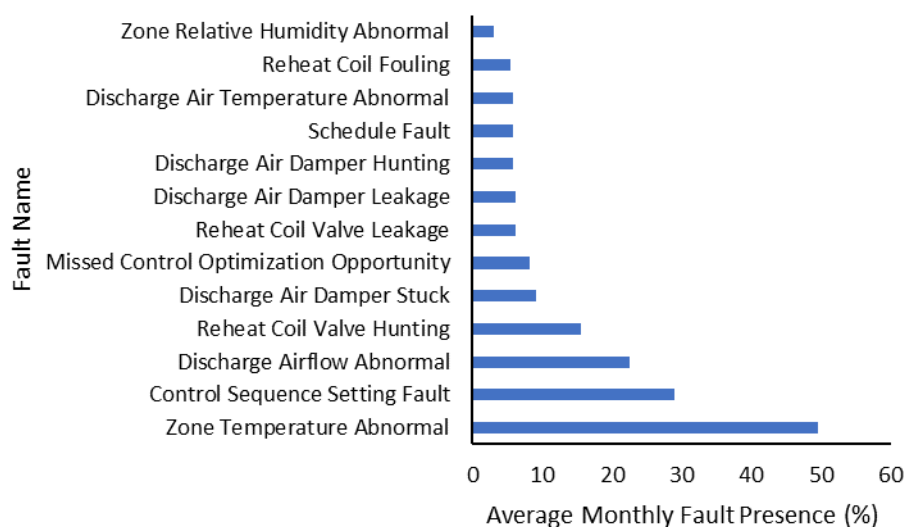


Figure 19. Average monthly fault presence (metric 2) for ATU faults from FDD provider B

Figure 20 shows the comparison of average monthly fault presence for ATU faults between FDD providers A and B. For most of the faults, FDD provider A has lower prevalence rates. As we explained for AHU faults, this scale variation is not surprising to us. Adding more FDD data providers to the study would help to have a better understanding of the fault prevalence values.

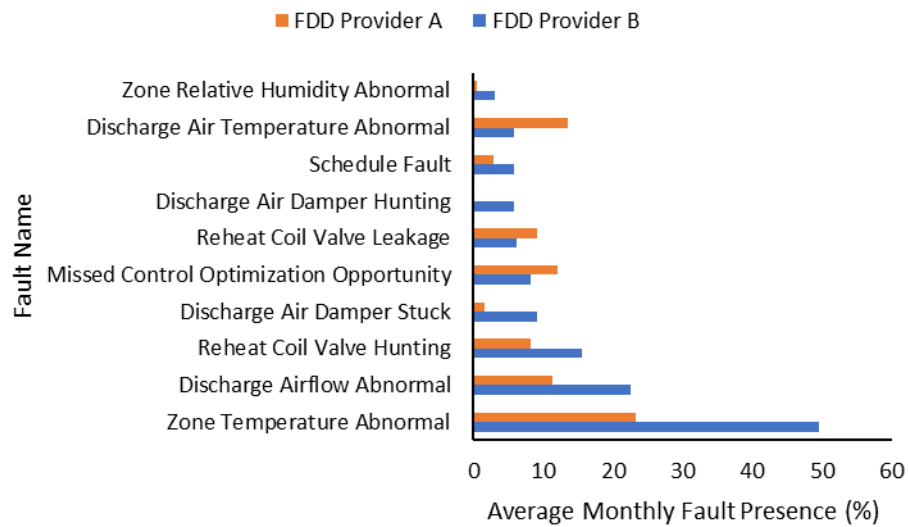


Figure 20. Comparison of average monthly fault presence for ATU faults between FDD providers A and B

#### 4.2.3. RTU Results

Figure 21 shows the average monthly fault presence (metric 2) for RTU faults for FDD provider C. There are 38 unique RTU faults. “Zone relative humidity sensor frozen”, “outdoor air temperature sensor frozen”, and “zone temperature sensor frozen” are the most common faults from the FDD provider C representing faults in 55%, 52%, and 47% of the RTUs, respectively. As can be seen, 9 faults among the 10 most common faults are either sensor frozen or abnormal readings. All of these faults have high prevalence rates. Our field study verification revealed that these are not real faults, and most of them are false positives of the FDD tool. This topic is explained in more detail in chapter 5.

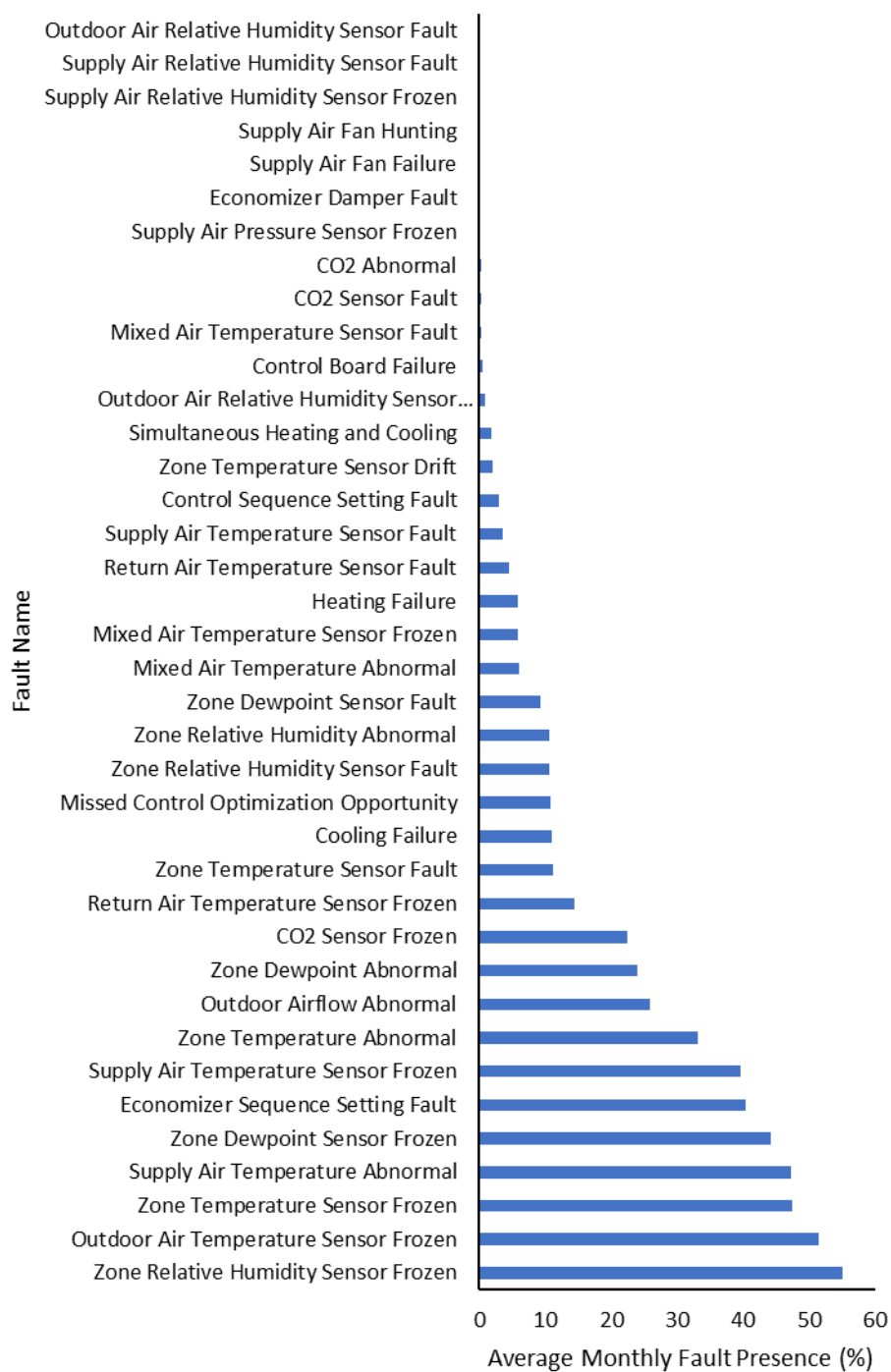


Figure 21. Average monthly fault presence (metric 2) for RTU faults from FDD provider C

Figure 22 shows the average monthly fault presence for RTU faults for FDD provider B. There are only 3 unique RTU faults. Since FDD provider B has only 13 RTUs, it is expected these fault prevalence values cannot be generalized.

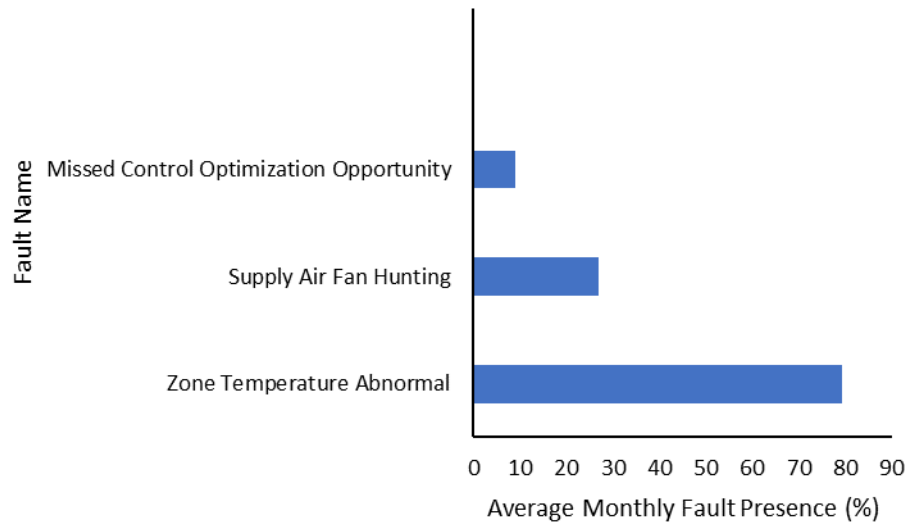


Figure 22. Average monthly fault presence (metric 2) for RTU faults from FDD provider B

#### 4.3. Mean Number of Faults per Building per Month (Metric 3) Results

Metric 3 indicates how many faults are observed to be present (at the building level) each month, among the set of faults considered in this study. The metric 3 results for the three FDD providers are presented in the following.

#### **4.3.1. FDD provider A Results**

Figure 23 shows the distribution of the mean number of faults per building per month for FDD provider A. As can be seen, 48% of the buildings were in the range of 0-100 faults per month, 18% were in the range of 100-200, 15% were in the range of 200-300, 7% were in the range of 300-400, and 12% had higher than 400 faults per month. It should be noted that the number of faults in each building includes all the AHU and ATU faults. As we expected, buildings with higher quantities of equipment had higher quantities of faults. One health care building in a hot-dry climate zone with 38 AHUs and 834 ATUs had 1,043 faults per month which was the highest number among all the buildings. The average and median values are 176 and 102 faults per building per month, respectively. This is an example of a metric where it could make more sense to normalize.

Figure 24 shows the distribution of the mean number of faults per building per equipment per month for FDD provider A. The average and median values are 1.5 and 1.4 faults per building per equipment per month, respectively. One health care building in a hot-dry climate zone with 10 AHUs had 5.5 faults per equipment per month which was the highest number among all the buildings.

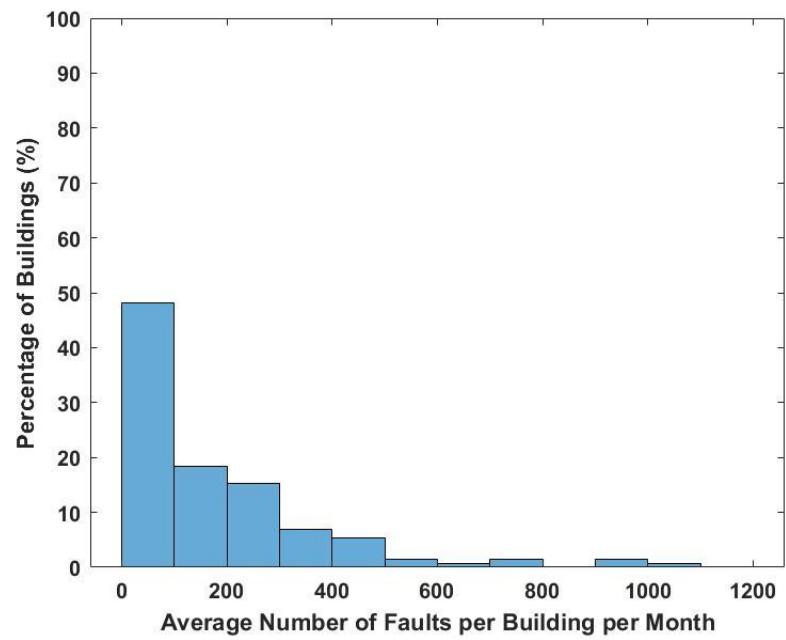


Figure 23. Mean number of faults per building per month distribution for FDD provider A

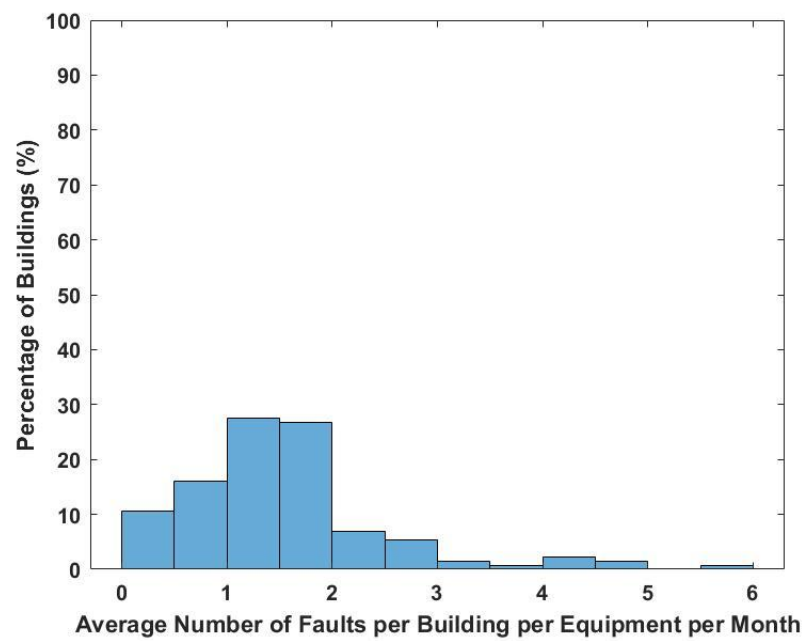


Figure 24. Mean number of faults per building per equipment per month distribution for FDD provider A

#### **4.3.2. FDD provider B Results**

Figure 25 shows the distribution of the mean number of faults per building per month for FDD provider B. As can be seen, 34% of the buildings were in the range of 0-300 faults per month, 17% were in the range of 300-600, 14% were in the range of 600-900, 3% were in the range of 900-1200, and 32% had higher than 1200 faults per month. The number of faults in each building includes all the AHU, ATU and RTU faults. One health care building in a hot-dry climate zone with 58 AHUs and 1,720 ATUs had 3,559 faults per month which was the highest number among all the buildings. The average and median values are 922 and 589 faults per building per month, respectively.

Figure 26 shows the distribution of the mean number of faults per building per equipment per month for FDD provider B. The average and median values are 2.1 and 2.1 faults per building per equipment per month, respectively. One health care building in a hot-dry climate zone with 30 AHUs and 445 ATUs had 3.6 faults per equipment per month which was the highest number among all the buildings.

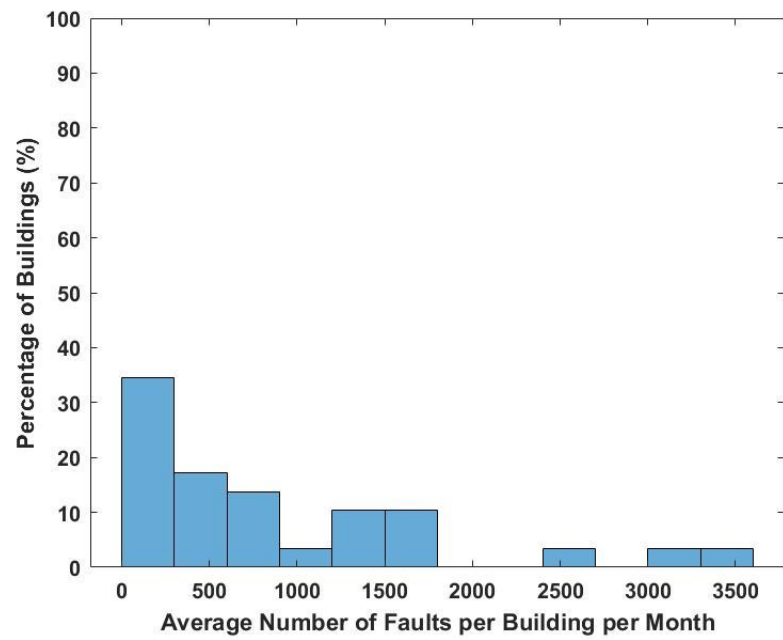


Figure 25. Mean number of faults per building per month distribution for FDD provider B

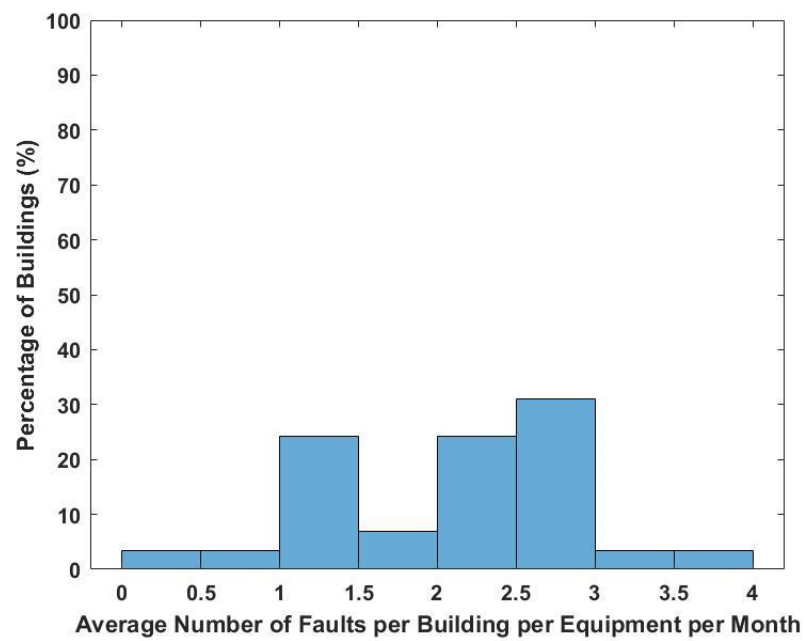


Figure 26. Mean number of faults per building per equipment per month distribution for FDD provider B

#### **4.3.3. FDD provider C Results**

Figure 27 shows the distribution of the mean number of faults per building per month for FDD provider C. 13% of the buildings were in the range of 0-50 faults per month, 52% were in the range of 50-100, 29% were in the range of 100-150, and 6% were in the range of 150-200. The number of faults in each building includes all the RTU faults. One mercantile building in a cold climate zone with 29 RTUs had 191 faults per month which was the highest number among all the buildings. The average and median values are 90 and 83 faults per building per month, respectively.

Figure 28 shows the distribution of the mean number of faults per building per equipment per month for FDD provider C. The average and median values are 5.5 and 5.7 faults per building per equipment per month, respectively. These values are higher than the corresponding values of the other two FDD providers and could be related to the false positives of this FDD tool we found in our field study. One mercantile building in a hot-humid climate zone with 14 RTUs had 8.8 faults per equipment per month which was the highest number among all the buildings.

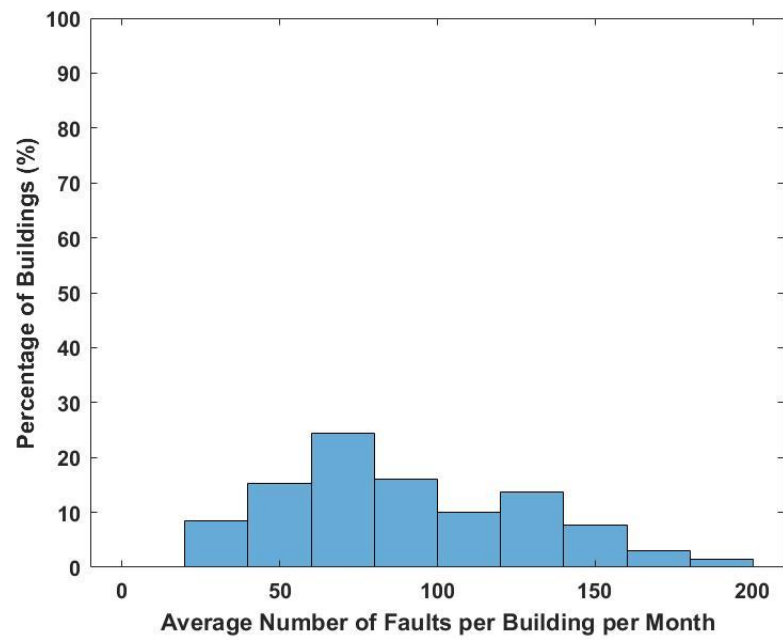


Figure 27. Mean number of faults per building per month distribution for FDD provider C

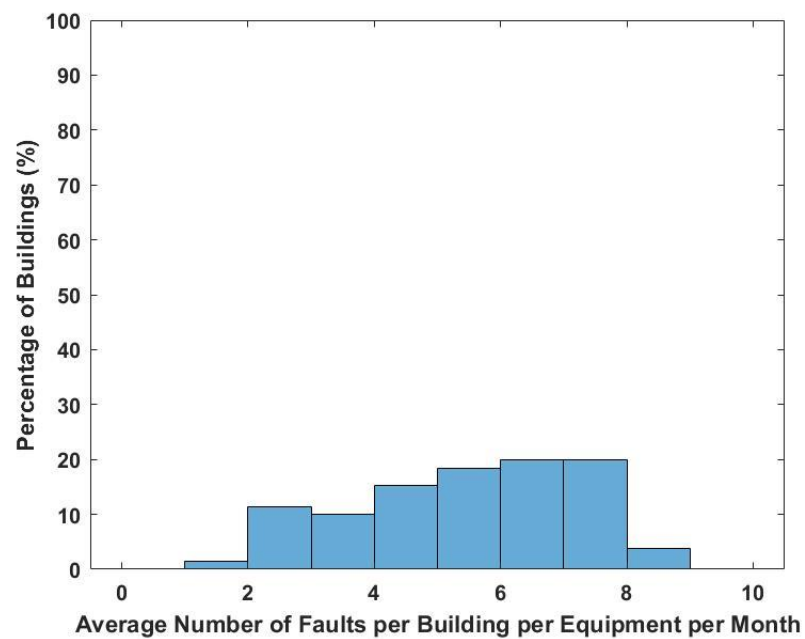


Figure 28. Mean number of faults per building per equipment per month distribution for FDD provider C

## **CHAPTER 5. Field Study Results & Discussion**

As we mentioned earlier, since the commercial FDD software results inherently contain a certain level of error, these results are complemented with a field verification to evaluate the performance of the FDD software tools. For this purpose, two buildings from among the buildings of the FDD provider C are selected. RTUs of these two buildings are monitored for about two weeks using our installed data loggers. Using the fault detection and diagnostics methods explained in section 3.2, the actual RTU faults in these buildings are identified. The results of our field study are compared with the FDD provider C fault reports to find the false negatives and false positives.

### **5.1. First Building Results**

The first building has a total of 27 RTUs. The RTUs are monitored from March 4, 2021 to March 12, 2021. Table 20 shows the FDD provider C fault report for the same period.

As can be seen, the fault prevalence rates are very high. We took a closer look at the BAS data to find the reason. We noticed that there are a couple of columns with zero values in the BAS data. This shows that sensor values are not correctly communicated to BAS.

FDD provider C software considers these zero values as sensor frozen faults. These faults are false alarms (false positives) of the FDD software, since a fault is identified while no fault is present. This is the reason why we mentioned in chapter 4 that the sensor fault prevalence results from the FDD provider C are not reliable.

Table 20. FDD provider C fault report for the first building

<b>Fault Name</b>	<b>Number of RTUs</b>	<b>Number of Faulted RTUs</b>	<b>Fault Prevalence (%)</b>
Mixed air temperature sensor frozen	27	27	100
Return air temperature sensor frozen	27	27	100
Return air CO <sub>2</sub> sensor frozen	27	27	100
Zone relative humidity sensor frozen	27	21	78

Table 21 shows the actual faults identified using the collected data in our field visit and FDD methods explained in section 3.2.

These are the actual faults that our field verification identified in the first building. None of these faults were detected by the provider C FDD software, and these are missed detections (false negatives) of the software. It should be noted that the FDD software was not designed to diagnose the non-condensable gas and abnormal supply fan belt tension faults.

Table 21. Actual faults identified for the first building

<b>Fault Name</b>	<b>Number of RTUs Checked</b>	<b>Number of Faulted RTUs</b>	<b>Fault Prevalence (%)</b>
Zone air temperature sensor bias	11	3	27
Economizer damper stuck	10	2	20
Non-condensable gas	19	1	5
Abnormal supply fan belt tension	10	7	70

Figure 29 shows an example of both healthy and faulted zone air temperature sensors. To detect and diagnose the zone air temperature sensor bias, values in the time series of measurements from the building BAS data are compared to values in a corresponding time series of measurements from our data loggers (considered as real values). If the difference between the values is higher than 2 °F (detection threshold) and nearly constant over time, the fault is categorized as a sensor bias.

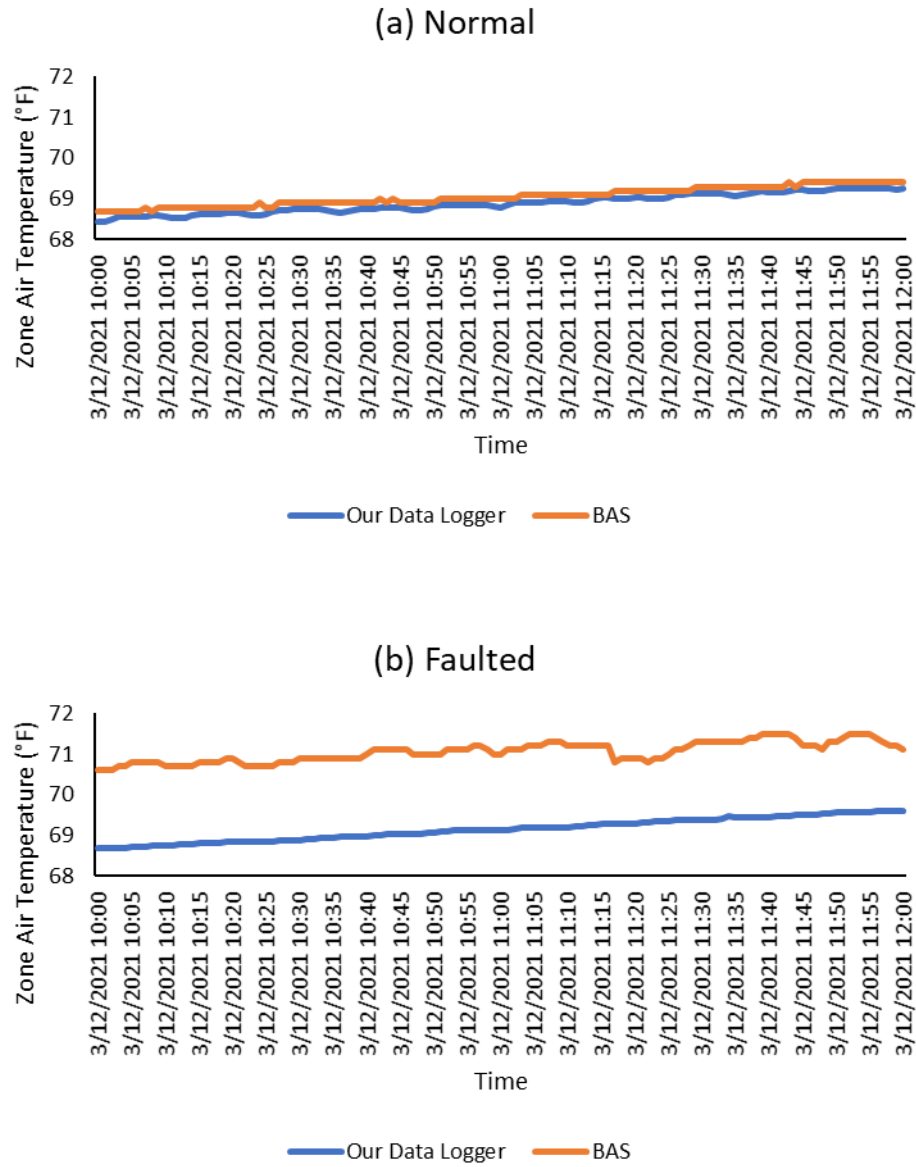


Figure 29. Normal and faulted zone air temperature sensors

Table 22 shows some examples of the RTUs with and without non-condensable gas inside their refrigerant circuits. When the unit is off and two-phase refrigerant exists within the condenser, the difference between the measured condensing temperature and

saturation temperature calculated from the measured compressor discharge pressure is used for detecting the non-condensable gas fault (Li and Braun, 2007c). The detection threshold is selected to be 3 °F. All RTUs have R410A as refrigerant.

Table 22. RTUs with and without non-condensable gas

Unit	Refrigerant	Saturation	Condenser	$\Delta T$ (°F)	Diagnostics
	Pressure (psig)	Temperature (°F)	Temperature (°F)		
RTU 09	111.6	65.2	59.6	5.6	Faulted
RTU 13	102.2	60.3	59.5	0.8	Normal

Table 23 shows some examples of RTUs with normal and abnormal supply fan belt tension. A belt tension checker is used for detecting and diagnosing this fault.

Table 23. RTUs with normal and abnormal supply fan belt tension

Unit	Belt Cross Section	Smallest	Rpm	Actual	Normal	Diagnostics
		Sheave Diameter (Inches)		Belt Deflection Force (lbs)	Belt Deflection Force (lbs)	
RTU 06	BX	4.2	1735	3.5	4.9-7.2	Under Tension
RTU 09	BX	4.2	1735	6.2	4.9-7.2	Normal

Figure 30 shows an example of inclusion of temperature setbacks in the building during unoccupied periods. Using temperature setbacks during unoccupied times can result in reducing building energy consumption. Temperature setbacks were implemented in all the RTUs in this building.

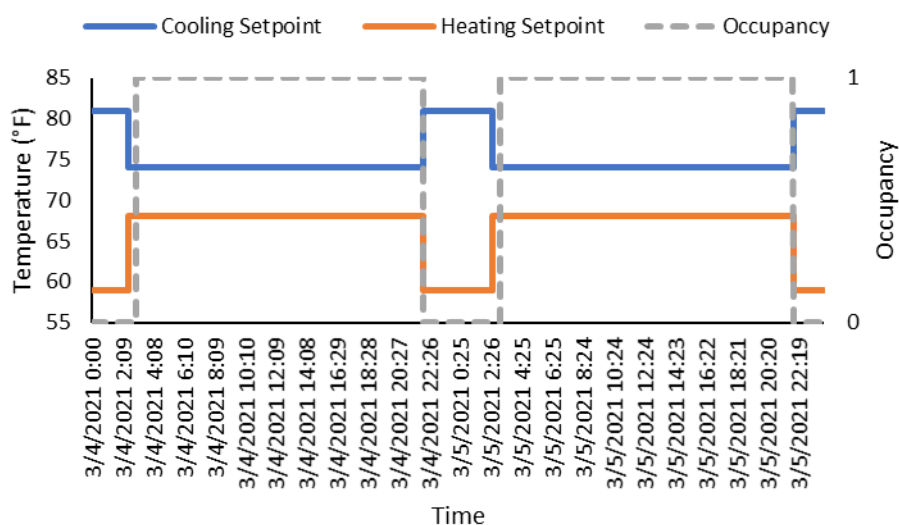


Figure 30. Unoccupied temperature setback

## 5.2. Second Building Results

The second building has 22 RTUs. Monitoring time is from March 26, 2021 to April 13, 2021. Table 24 shows the FDD provider C fault report for this time period.

Table 24. FDD provider C fault report for the second building

<b>Fault Name</b>	<b>Number of RTUs</b>	<b>Number of Faulted RTUs</b>	<b>Fault Prevalence (%)</b>
Supply air			
temperature sensor	22	14	64
frozen			
Return air temperature			
sensor frozen	22	11	50
Return air CO <sub>2</sub> sensor			
frozen	22	22	100
Zone dewpoint sensor			
frozen	22	15	68
Economizer damper			
stuck	22	3	14

Similar to first building, the fault prevalence rates are very high. As we explained earlier, the reason is the presence of a couple of columns with zero values in the BAS data. FDD tool identifies the zero values as sensor frozen faults.

Table 25 shows the results of our field study in the second building. Most RTUs have several refrigerant circuits. For the non-condensable gas fault, all refrigerant circuits are checked. That is why there are 30 units (> 22 units) in the table. None of these faults except three economizer damper stuck faults were detected by the provider C FDD tool.

Table 25. Field study results for the second building

<b>Fault Name</b>	<b>Number of RTUs Checked</b>	<b>Number of Faulted RTUs</b>	<b>Fault Prevalence (%)</b>
Zone air temperature sensor bias	11	0	0
Economizer damper stuck	22	6	27
Non-condensable gas	30	0	0
Abnormal supply fan belt tension	18	10	56

### 5.3. Confidence Interval for Fault Prevalence

One interesting goal of this field study could be finding the prevalence of the specific faults in the whole population of RTUs in the US commercial buildings. There is a discrete distribution in the population which means  $\theta$  percent of RTUs have a specific fault and  $1-\theta$  percent of RTUs do not have that specific fault. We are looking for to find the  $\theta$  which is a constant unknown. If  $\theta = 0.4$ , it means that 40% of RTUs are faulted.

In order to find the  $\theta$ , we need to select a random sample from the whole population of RTUs. It should be noted that our sample which includes 49 RTUs is not completely random, and therefore we have some biases. An estimate of the  $\theta$  is  $\hat{\theta}$  which is the fault prevalence based on the sample selected. Unlike  $\theta$ ,  $\hat{\theta}$  has a distribution, since if we repeat

the sampling several times, we get a new value for  $\hat{\theta}$  each time. It can be shown that the  $\hat{\theta}$  has a variance of  $\theta(1-\theta)/n$ , where  $n$  is the sample size. Therefore, the standard error of  $\hat{\theta}$  is  $\{\hat{\theta}(1-\hat{\theta})/n\}^{1/2}$ . It also can be shown that  $(\hat{\theta}-\theta)/\{\hat{\theta}(1-\hat{\theta})/n\}^{1/2}$  converges to a standard normal variable (Davison, 2003), and therefore a  $(1-2\alpha)$  confidence interval for  $\theta$  has the endpoints  $\hat{\theta} - z_{1-\alpha}\{\hat{\theta}(1-\hat{\theta})/n\}^{1/2}$  and  $\hat{\theta} + z_{\alpha}\{\hat{\theta}(1-\hat{\theta})/n\}^{1/2}$ . If we want a 95% confidence interval for  $\theta$ , then  $z_{0.975} = -z_{0.025} = 1.96$ . Table 26 shows the 95% confidence interval for the prevalence of the RTU faults considered in our field study. As can be seen, since our sample size is small, we have a wide 95% confidence interval.

Table 26. 95% confidence interval for the prevalence of RTU faults

<b>Fault Name</b>	<b>Number of RTUs Checked</b>	<b>Number of Faulted RTUs</b>	<b>Fault Prevalence (%)</b>	<b>95% Confidence Interval (%)</b>
Zone air				
temperature sensor	22	3	14	(0,28)
bias				
Economizer				
damper stuck	32	8	25	(10,40)
Non-condensable				
gas	49	1	2	(0,6)
Abnormal supply				
fan belt tension	28	17	61	(43,79)

## CHAPTER 6. Data-Driven FDD for RTUs Results & Discussion

To explore the performance of the various machine learning classification methods in detecting and diagnosing the normal and seven fault modes of operation in RTUs, we applied each classification method to our simulation data library. This data library is generated with simulations based on Cheung & Braun (2013a, 2013b) to provide a rich training dataset for the classifiers. We would like to emphasize that this data library is different than the FDD data we talked about in chapters 4 and 5. All statistical models were implemented using R, a statistical software environment (R Core Team, 2019). A list of R packages and functions used for each classification method is shown in Table 27. As discussed in chapter 3, optimal tuning parameters for each classification method are selected using either CV or OOB techniques.

Table 27. A full list of R packages and functions used

<b>Classification Method</b>	<b>Function</b>	<b>Package</b>
LR	multinom ()	nnet
LDA	lda ()	MASS
QDA	qda ()	MASS
KNN	knn ()	class
BA	randomForest ()	randomForest
RF	randomForest ()	randomForest
AD	boosting ()	adabag
XGB	xgboost ()	xgboost
SVM	svm ()	e1071

Figure 31 shows the 10-fold CV accuracy (or OOB accuracy) and test accuracy for nine classifiers. The 10-fold CV accuracy (or OOB accuracy) provides a reasonable approximation to the true test accuracy. For LDA, RF, AD, BA, XGB, LR, and SVM

classifiers, it overestimates the true test accuracy, while for KNN and QDA classifiers, it underestimates the true test accuracy. Based on the true test accuracy, which is the real quantity of interest, SVM and LR are the best classifiers, with overall accuracy rates of 96.2% and 93.6% respectively, and KNN and LDA are the weakest classifiers, with overall accuracy rates of 83.6% and 76.2%, respectively. Confusion matrices for these four classifiers on the test data are shown in Figure 32.

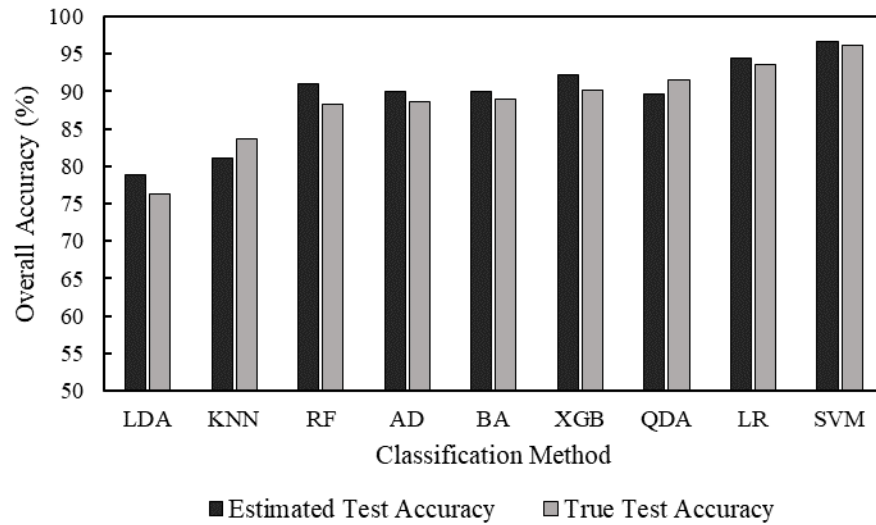


Figure 31. Estimated and true test accuracy for different classification methods

The SVM classification method has two tuning parameters: cost and gamma (James et al., 2013). The cost parameter determines the cost of a violation to the margin. For large values of cost, the margin will be small, and few support vectors will be on the margin or will violate the margin. The gamma parameter shows how far the influence of a single training sample reaches. If gamma parameter is very small, the region of influence of any

support vector includes the whole training set. If gamma parameter is too large, the region of influence of the support vectors only includes the support vector itself. 10-fold CV was performed to select the best values for these parameters, with results as shown in Figure 33. Using values of  $\text{cost}=10^7$ , and  $\text{gamma}=10^{-4}$  resulted in the highest 10-fold CV accuracy rate. The optimal tuning parameter values for each of the classification methods are shown in Table 28.

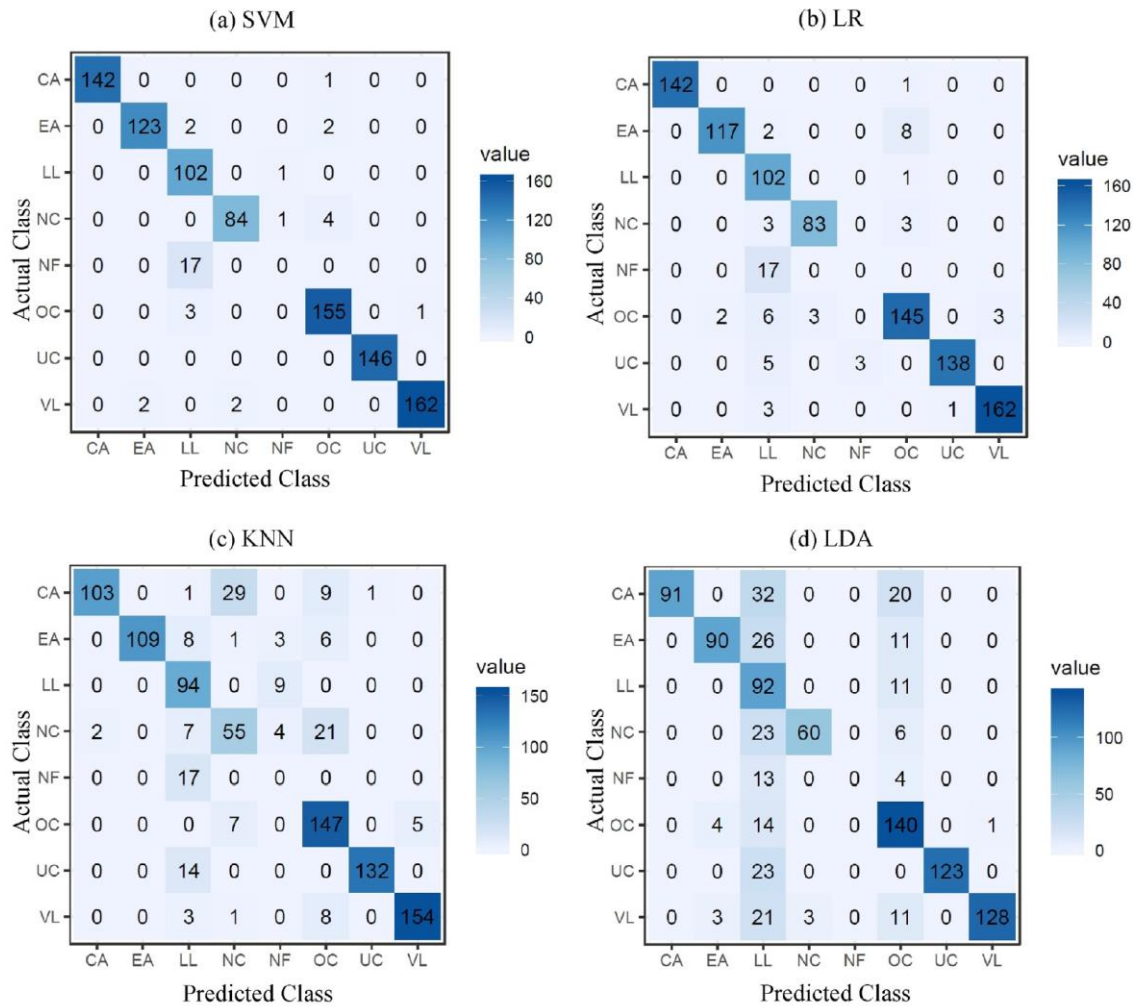


Figure 32. Confusion matrices for four classification methods: (a) SVM, (b) LR, (c) KNN, (d) LDA

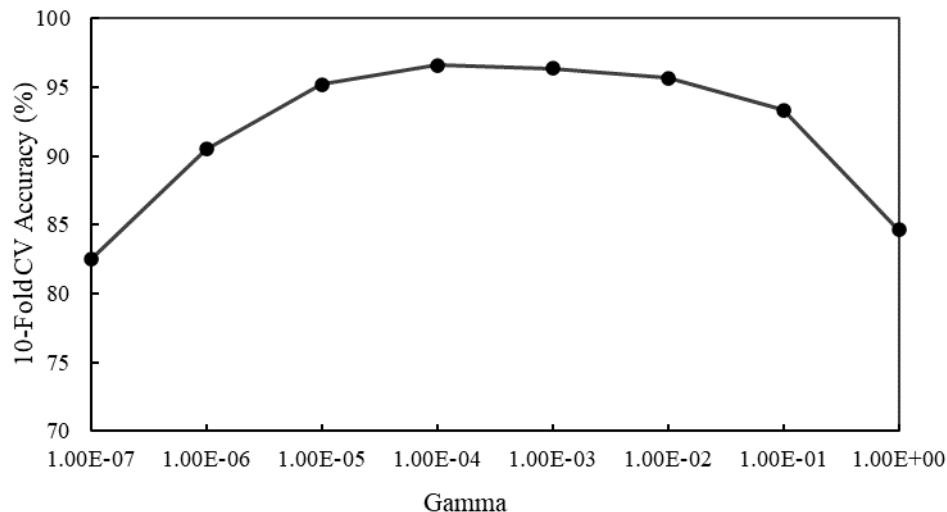


Figure 33. SVM 10-fold CV accuracy rates as a function of gamma values while cost= $10^7$

Table 28. The list of tuning parameters for each of the classification methods

Classification Method	Tuning Parameters	Range of Values	Optimal Values
LR	-	-	-
LDA	-	-	-
QDA	-	-	-
KNN	Number of nearest neighbors, K	[1, 2, 3, ..., 30]	1
BA	Number of trees, ntree	[50, 100, 150, ..., 1000]	500
RF	Number of variables randomly sampled as candidates at each split, mtry	[1, 2, 3, ..., 14]	2
	Number of trees, ntree	[50, 100, 150, ..., 1000]	550
AD	Maximum depth of each tree, maxdepth	[1, 2, 3, ..., 8]	8
	Number of trees, mfinal	[100, 200, 300, ..., 1000]	900
XGB	Learning rate, eta	[0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 0.9]	0.5
	Maximum depth of each tree, max_depth	[1, 2, 3, ..., 10]	5
	Number of trees, nrounds	[100, 200, 300, ..., 1000]	300
SVM	Cost of constraints violation, cost	[ $10^0$ , $10^1$ , $10^2$ , ..., $10^8$ ]	$10^7$
	Parameter needed for radial kernel, gamma	[ $10^{-7}$ , $10^{-6}$ , $10^{-5}$ , ..., $10^0$ ]	$10^{-4}$

Figure 34 shows the TPR of each class for all the classification methods. A TPR of 100% means that all the samples of that class are correctly classified. None of the classifiers, except QDA, could correctly predict the unfaulted (NF) class samples (TPR of 0%). Even QDA can only correctly classify 58.8% of the NF samples. This is an unfortunate result, because in practical application the most important task of FDD is to avoid “false alarms” (Yuill and Braun, 2017).

The reason for this low performance is that our dataset is highly imbalanced, and only 48 samples out of the total 2851 samples belong to the NF class. This problem clearly shows the effect of class distribution on classifier learning (Weiss and Provost, 2001). The imbalance is caused by the fact that the simulation matrix contains simulations at each combination of operating conditions for several fault intensities, but for the NF class, there can only be one intensity. We address the issue of class imbalance later in this study. The SVM classification method has a very high TPR for all the classes except the NF class. In this method, the TPR of UC, OC, VL, LL, CA, EA, NC, and NF classes are 100%, 97.5%, 97.6%, 99.0%, 99.3%, 96.9%, 94.4%, and 0.0%, respectively.

Interestingly, all of the UC faults are correctly predicted. As can be seen from the confusion matrix shown in Figure 32(a), all 17 NF samples in the test data are misclassified as LL. For the LDA classifier, the TPR of UC, OC, VL, LL, CA, EA, NC, and NF classes are 84.2%, 88.1%, 77.1%, 89.3%, 63.6%, 70.9%, 67.4%, and 0.0%, respectively. The confusion matrix in Figure 32(d) shows that 13 NF samples were misclassified as LL and 4 NF samples are misclassified as OC.

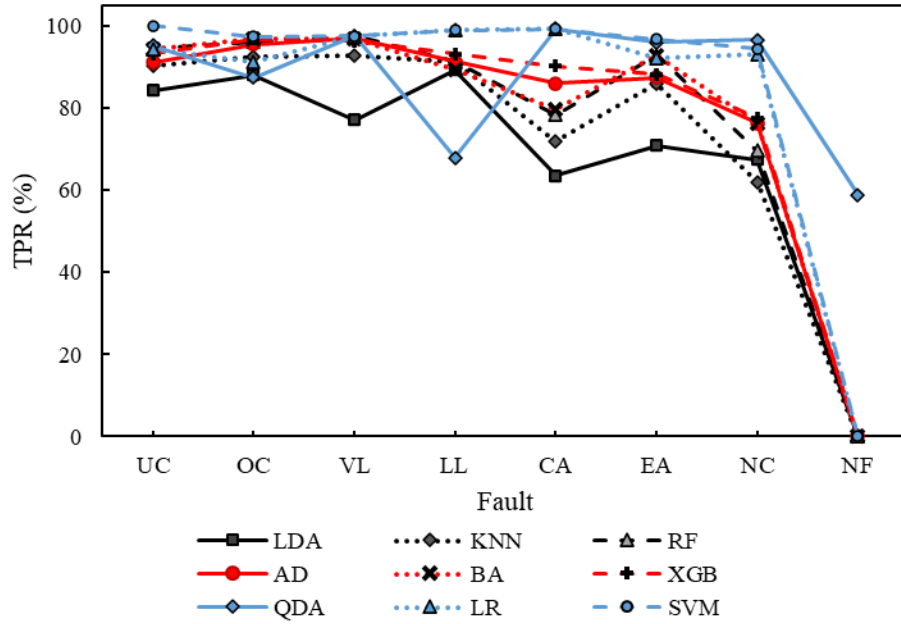


Figure 34. TPR values for each class for all nine classification methods

The FPR of each class for all classification methods is shown in Figure 35. A FPR of 0% means that none of the other classes is classified into that class. The figure shows that LL and OC have high FPR values, while UC has the lowest FPR value. The SVM classifier has a very low FPR for all classes. For this classifier, the FPR of UC, OC, VL, LL, CA, EA, NC, and NF classes are 0.0%, 0.9%, 0.1%, 2.6%, 0.0%, 0.2%, 0.2%, and 0.2%, respectively. For the LDA classification method, the FPR of UC, OC, VL, LL, CA, EA, NC, and NF classes are 0.0%, 8.0%, 0.1%, 17.9%, 0.0%, 0.9%, 0.3%, and 0.0%, respectively. Of all fault types, LL has the highest FPR value for this classifier. The confusion matrix in Figure 32(d) shows that 23 UC samples, 14 OC samples, 21 VL samples, 32 CA samples, 26 EA samples, 23 NC samples, and 13 NF samples are misclassified as members of class LL.

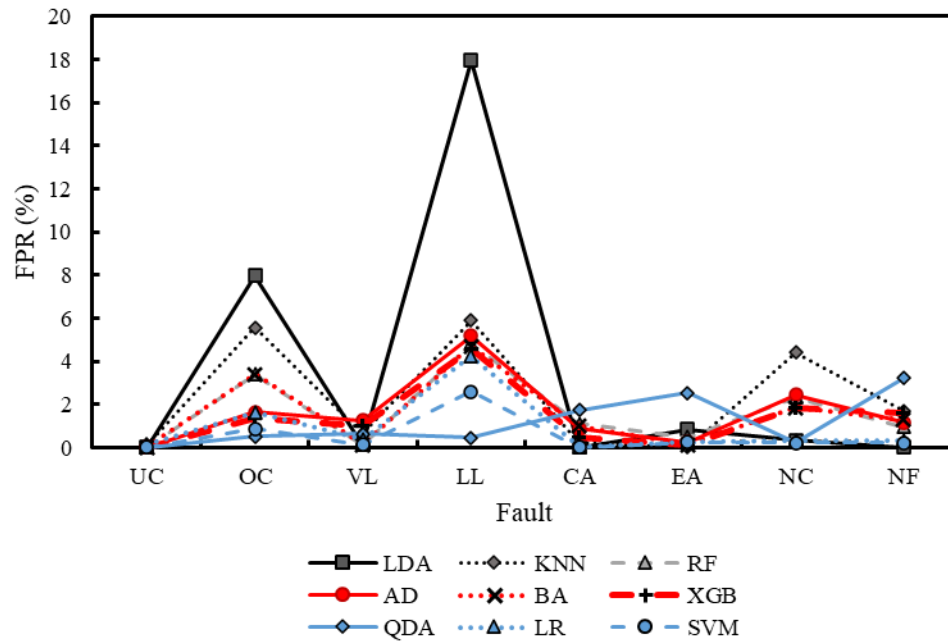
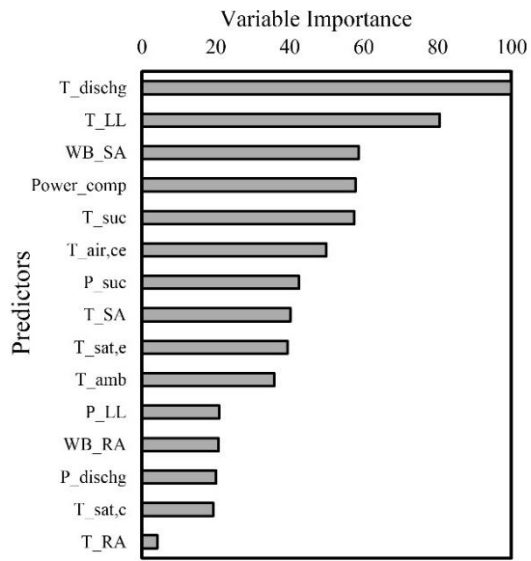
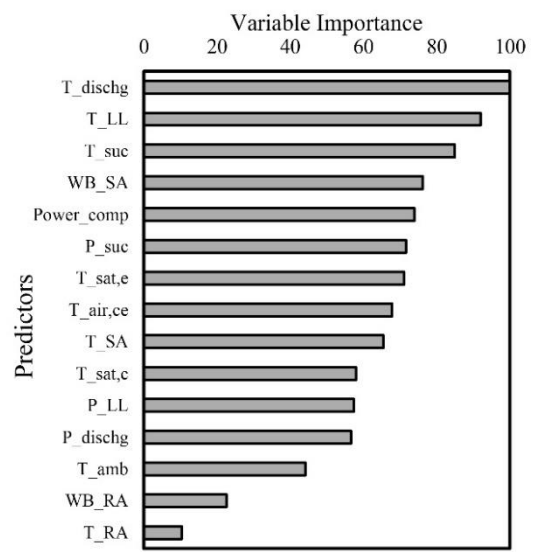


Figure 35. FPR values for each class for all nine classification methods

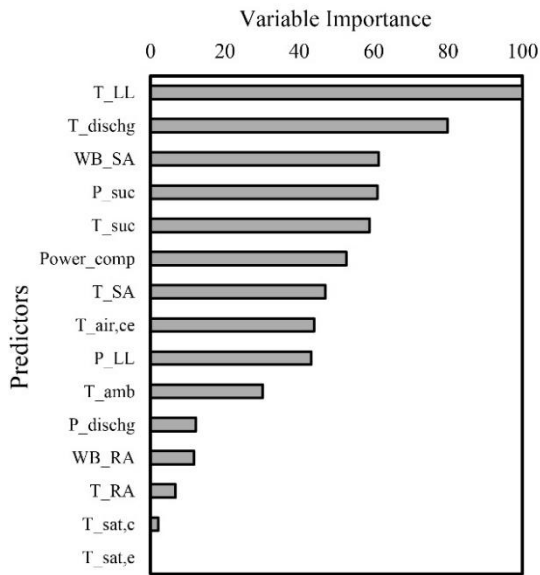
For classification methods BA, RF, AD, and XGB, we can compute the total decrease in Gini index (James et al., 2013) from splitting over a given predictor variable, averaged over all trees. A large value shows an important predictor variable. Gini index is a measure of node impurity; a small value indicates that a node mostly contains samples from a single class. Figure 36 shows a graphical representation of the importance of each predictor variable in the classification task. Based on the results obtained from these four classification methods,  $T_{LL}$  and  $T_{dischg}$  are the most important predictor variables in the fault detection and diagnosis process. However, there is no clear drop off in importance to divide essential predictors from non-essential predictors. For example, at least five predictors score above 50% importance for each of the classifiers shown in the figure.



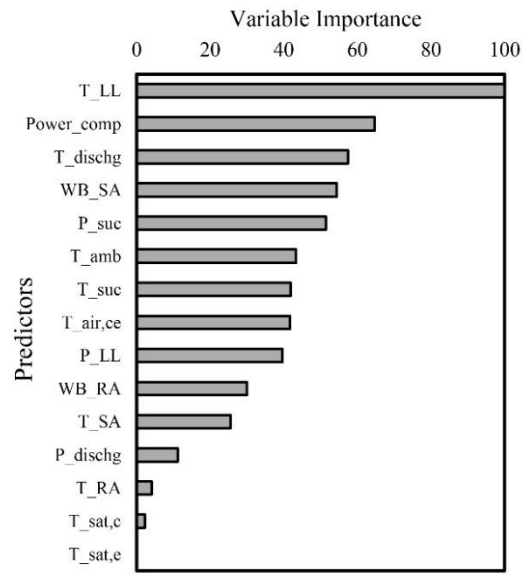
(a) BA



(b) RF



(c) AD



(d) XGB

Figure 36. Variable importance plots for different classification methods: (a) BA, (b) RF, (c) AD, (d) XGB

In order to alleviate the problem of class imbalance in our original dataset, we used the synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002) to balance the classes. In this oversampling (Japkowicz, 2000) approach, new samples of the minority class (NF class, as discussed above) are artificially generated using the K nearest neighbors of each minority class (NF) sample. The SMOTE function with  $K=5$  from the R package DMwR was used to oversample the minority (NF) class. As can be seen from Table 18, the original training data has 31 NF samples. 155 new synthetic NF samples (500% of original size) were generated using the oversampling technique. After applying the oversampling, the test accuracy of the SVM method slightly decreases from 96.2% to 95.5%. The test accuracy of the LR method also slightly decreases from 93.6% to 92.6%. However, there is a significant reduction in false negative rate ( $FNR=1-TPR$ ) for the NF class. Updated confusion matrices for these two classifiers on the test data are shown in Figure 37.

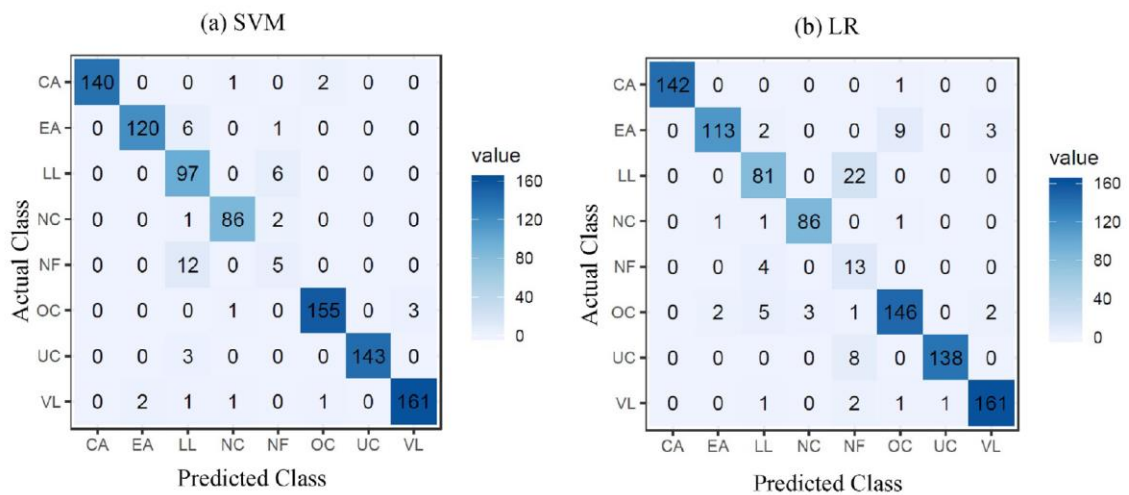


Figure 37. Confusion matrices for two classification methods after applying oversampling: (a) SVM, (b) LR

Figure 38 shows the TPR of each class before and after applying the oversampling method for the SVM and LR classification methods. Before applying the oversampling method, the SVM classifier could not correctly predict any of the 17 NF samples (TPR of 0%). After applying the oversampling method, the SVM classification method correctly predicted 5 NF samples out of 17 NF samples (TPR of 29.4%). The TPR of other classes also changed very slightly for SVM. For the LR classifier, the TPR of the NF class increased from 0% to 76.5% (it correctly predicted 13 of 17 NF samples). However, the TPR of the LL class decreased from 99.0% to 78.6%, and the TPR of other classes also changed very slightly. These results show that the oversampling approach improves the performance on the minority class (NF class) for the LR classifier more than it does for the SVM classifier.

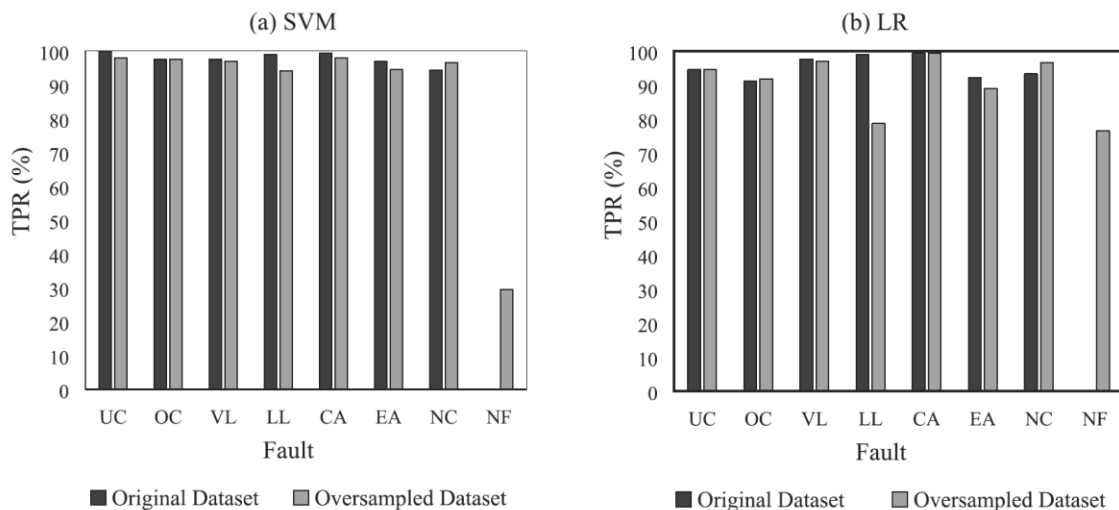


Figure 38. TPR values for each fault class before and after applying the oversampling: (a) SVM, (b) LR

The FPR of each class before and after applying the oversampling method for the SVM and LR classification methods is shown in Figure 39. For the SVM method, the FPR of the minority class (NF) increases from 0.2% to 1.0% after applying the oversampling method. For LR classifier, the FPR of the minority class (NF class) increases from 0.3% to 3.5%. The results show that the oversampling approach improves the performance on the minority class (NF class) in the expense of increasing the FPR value. However, as noted above, in practical application, the cost of the FPR for the NF class is low compared to the cost of the FNR.

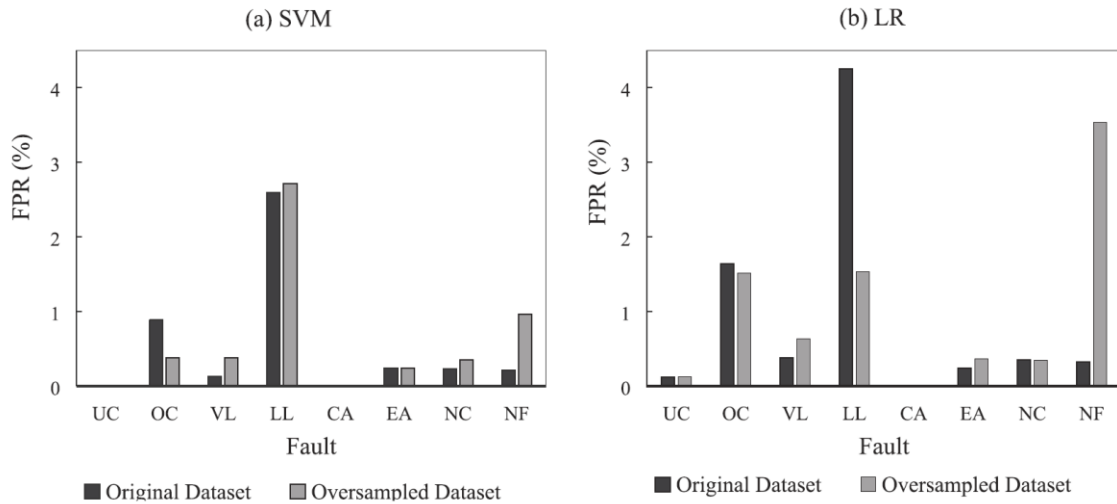


Figure 39. FPR values for each fault class before and after applying the oversampling: (a) SVM, (b) LR

## **CHAPTER 7. Conclusions and Recommendations for Future Research**

In this chapter, the conclusions and recommendations for future research are summarized.

### **7.1. HVAC Fault Prevalence Summary**

In this study, a multi-year dataset including a total of 11,688,583 daily fault records of AHUs, ATUs, and RTUs in commercial buildings is analyzed to determine a range of HVAC fault prevalence metrics. The fault data received for this study is sourced from three commercial FDD providers. Fault data from each provider are converted to a standard format, which is called binary daily fault (BDF) data. Since each FDD provider uses different fault names to refer to the same fault in an HVAC system, a unifying taxonomy for HVAC faults is used. Mapping functions were created for each FDD provider to convert their fault reports to this unifying taxonomy. To quantitatively characterize the HVAC fault prevalence, the following metrics are defined: monthly fault presence (metric 1), average monthly fault presence (metric 2), and mean number of faults per building per month (metric 3). Based on our results, the following conclusions can be drawn:

- (1) While some faults (e.g., “RTU cooling failure” from FDD provider C) shows a genuine seasonal trend, others (e.g., “ATU discharge airflow abnormal” from FDD provider A) do not have an obvious seasonal trend.
- (2) “Missed control optimization opportunity”, “sensor frozen”, and “mismatch between supply air temperature and its setpoint” with 28%, 27%, and 26%

prevalence rate, respectively, are the most common AHU faults from the FDD provider A.

- (3) “Missed control optimization opportunity”, “mismatch between supply air temperature and its setpoint”, and “mixed air temperature sensor fault” with 58%, 40%, and 35% prevalence rate, respectively, are the most common AHU faults from the FDD provider B.
- (4) FDD provider A and B have two common AHU faults between their three most prevalent AHU faults. FDD provider A has systematically lower prevalence rates for AHU faults.
- (5) “Zone temperature abnormal”, “sensor frozen”, and “discharge air temperature abnormal” with 23%, 14%, and 14% prevalence rate, respectively, are the most common ATU faults from the FDD provider A.
- (6) “Zone temperature abnormal”, “control sequence setting fault”, and “discharge airflow abnormal” with 50%, 29%, and 23% prevalence rate, respectively, are the most common ATU faults from the FDD provider B.
- (7) “Zone temperature abnormal” is the most common ATU fault for both FDD providers A and B, although the prevalence rate is different.
- (8) FDD provider C has high prevalence rates for RTU faults. Our field study showed that this is because of the false positives of the FDD tool.
- (9) The average number of faults per building per equipment per month is 1.5, 2.1, and 5.5 for FDD providers A, B, and C, respectively.

The following recommendations for future research are made:

- (1) Collecting fault data from more FDD providers.
- (2) Evaluating how the HVAC fault prevalence metrics change with potential drivers such as building type and climate zone.
- (3) Implementing additional HVAC fault prevalence metrics that can provide new insights about the data.
- (4) Collecting fault data for other HVAC systems, e.g., chillers, boilers, cooling towers, etc.

## **7.2. Field Study Summary**

Since the commercial FDD software outputs inherently contain a certain degree of error, these results are complemented with a field verification to evaluate the performance of the commercial FDD software tools. Two buildings from among the buildings of the FDD provider C are selected. RTUs of these two buildings are monitored for about two weeks using our data loggers. Using our fault detection and diagnostics methods, the actual RTU faults in these buildings are identified. The results of our field study are compared with the FDD provider C fault reports to find the false negatives and false positives. The following conclusions are made:

- (1) In the first building, we found that “zone air temperature sensor bias”, “economizer damper stuck”, “non-condensable gas”, and “abnormal supply fan belt tension” faults have 27%, 20%, 5%, and 70% prevalence rate, respectively.

- (2) In the second building, we found that “zone air temperature sensor bias”, “economizer damper stuck”, “non-condensable gas”, and “abnormal supply fan belt tension” faults have 0.0%, 27%, 0.0%, and 56% prevalence rate, respectively.
- (3) FDD provider C fault report shows high prevalence rates for RTU sensor frozen faults. Our investigation showed that these are the false alarms (false positives) of the FDD tool.

The following recommendations for future research are made:

- (1) Monitoring more buildings from FDD provider C from different climate zones.
- (2) Conducting field study for other FDD providers, and check other HVAC systems, e.g., AHUs and ATUs.

### **7.3. Data-Driven FDD for RTUs Summary**

A data-driven RTU fault detection and diagnostics strategy was presented in this study. The proposed strategy formulates the FDD task as a multi-class classification problem. Several statistical machine learning classification methods are applied to our dataset in order to detect and diagnose the seven typical faults in RTU systems using fifteen input variables. This approach is validated using a fault data library of simulated measurements from faulted and unfaulted RTU generated based upon the methods of Cheung and Braun (2013a, 2013b). The results show that the classification algorithms can detect and diagnose the seven typical faults in RTU systems with varying but generally acceptable levels of success. Based on our results, the following conclusions can be drawn:

(1) The SVM classification method has the highest overall accuracy rate of 96.2%, and LDA classification method has the lowest overall accuracy rate of 76.2%.

(2) None of the classification methods, except QDA, could correctly predict the NF class samples (TPR of 0%). Even the QDA method only correctly predicted 58.8% of the NF samples. This is because our original dataset is highly imbalanced, and only 48 samples out of the total 2851 samples belong to the NF class (minority class).

(3) The results obtained from the BA, RF, AD, and XGB classification methods show that  $T_{LL}$  and  $T_{dischg}$  are the most important predictor variables in the RTU fault detection and diagnostics process. However, several additional predictors are necessary.

(4) Using the synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002) can alleviate the problem of class imbalance in our original dataset. After applying the oversampling, the overall accuracy of the SVM and LR methods slightly decreases, but their performance for predicting the minority class (NF class) improves significantly.

Overall, machine learning based FDD shows sufficient potential for further investigation.

Future work to build upon these results should include:

(1) Application to data sets from additional RTU, to test how generalizable the resulting classifications are.

- (2) Study of the tradeoffs between the number of types of input (temperatures, pressures, etc.), and the effectiveness of the classifier.
- (3) Consideration of multiple simultaneous faults in the dataset as additional categories. Many FDD tools struggle with accurate diagnosis when multiple faults are present, so it would be beneficial to know whether machine learning based FDD has the potential to be more effective than status quo methods. Newly available data from tests with multiple simultaneous faults (Hu and Yuill, 2021; Hu et al., 2021) may facilitate development in this area.
- (4) Changes or tuning of fault intensity thresholds. Some of the fault levels in the training data set may not be severe enough to warrant the cost of repairing. These cases could be removed or reclassified as unfaulted for training purposes. This step could potentially help to address a shortcoming of the proposed classification-based FDD method, which is that it does not provide a fault severity assessment.
- (5) Repeat this work for split systems, which are even more common than RTU and have similar behaviors with respect to fault effects.

## References

- Armstrong, P. R., Laughman, C. R., Leeb, S. B., & Norford, L. K. (2006). Detection of rooftop cooling unit faults based on electrical measurements. *HVAC and R Research*, 12(1), 151–175. <https://doi.org/10.1080/10789669.2006.10391172>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Bode, G., Thul, S., Baranski, M., & Müller, D. (2020). Real-world application of machine-learning-based fault detection trained with experimental data. *Energy*, 198, 117323. <https://doi.org/10.1016/J.ENERGY.2020.117323>
- Braun, J. E. (1999). Automated fault detection and diagnostics for the HVAC&R industry. *HVAC and R Research*, 5(2), 85–86. <https://doi.org/10.1080/10789669.1999.10391225>
- Braun, J. E. (2003). Automated fault detection and diagnostics for vapor compression cooling equipment. *Journal of Solar Energy Engineering, Transactions of the ASME*, 125, 266–274. <https://doi.org/10.1115/1.1591001>
- Breuker, M. S., & Braun, J. E. (1998a). Common faults and their impacts for rooftop air conditioners. *HVAC and R Research*, 4(3), 303–318. <https://doi.org/10.1080/10789669.1998.10391406>
- Breuker, M. S., & Braun, J. E. (1998b). Evaluating the performance of a fault detection and diagnostic system for vapor compression equipment. *HVAC and R Research*, 4(4), 401–425. <https://doi.org/10.1080/10789669.1998.10391412>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Chen, Y., Crowe, E., Lin, G., & Granderson, J. (2020). What's in a name? Developing a standardized taxonomy for HVAC system faults. *Lawrence Berkeley National Laboratory*, Berkeley, CA, USA.
- Chen, B., & Braun, J. E. (2001). Simple Rule-Based Methods for Fault Detection and Diagnostics Applied to Packaged Air Conditioners. *ASHRAE Transactions*, 847–857.
- Chen, Y., Lin, G., Crowe, E., & Granderson, J. (2021). Development of a Unified Taxonomy for HVAC System Faults. *Energies*, 14(17), 5581. <https://doi.org/10.3390/EN14175581>
- Cheung, H., & Braun, J. E. (2013a). Simulation of fault impacts for vapor compression systems by inverse modeling. Part I: Component modeling and validation. *HVAC and R Research*, 19(7), 892–906. <https://doi.org/10.1080/10789669.2013.824800>

- Cheung, H., & Braun, J. E. (2013b). Simulation of fault impacts for vapor compression systems by inverse modeling. Part II: System modeling and validation. *HVAC&R Research*, 19(7), 907–921. <https://doi.org/10.1080/10789669.2013.819769>
- Chiang, L. H., Russell, E. L., & Braatz, R. D. (2001). *Fault Detection and Diagnosis in Industrial Systems*. <https://doi.org/10.1007/978-1-4471-0347-9>
- Commercial Buildings Energy Consumption Survey (CBECS), *US Energy Information Administration*, (2002).
- Commercial Buildings Energy Consumption Survey (CBECS), *US Energy Information Administration*, (2012).
- Commercial Buildings Energy Consumption Survey (CBECS), *US Energy Information Administration*, (2018).
- Comstock, M. C., Braun, J. E., & Groll, E. A. (2002). A survey of common faults for chillers. *ASHRAE Transactions*, 108, 819-825.
- Comstock, M. C., & Braun, J. E. (1999). Experimental data from fault detection and diagnostic studies on a centrifugal chiller. *ASHRAE Research Project RP-1043*, Report #4036-1.
- Cowan, A. (2004). Review of recent commercial roof top unit field studies in the Pacific Northwest and California. *Northwest Power and Conservation Council and Regional Technical Forum*.
- Davis, R., Baylon, D., Hart, R., & Water. E. (2002). Identifying energy savings potential on rooftop commercial units. *In Proc. ACEEE Summer Study on Energy Efficiency in Buildings*.
- Davis, R., Francisco, P., Kennedy, M., Baylon, D., & Manclark. B. (2002). Enhanced Operations & Maintenance Procedures for Small Packaged Rooftop HVAC Systems.
- Davison, A. C. (2003). *Statistical Models*, Cambridge University Press, Cambridge, U. K.
- Department of Energy (DOE). (2011). *Buildings Energy Data Book*, Washington, D.C.
- Downey, T., & Proctor, J. (2002). What can 13,000 air conditioners tell us? *Proceedings of the 2002 ACEEE Summer Study on Energy Efficiency in Buildings*, 1, 53–67.
- Ebrahimifakhar, A., Kabirikopaei, A., & Yuill, D. (2020). Data-driven fault detection and diagnosis for packaged rooftop units using statistical machine learning classification methods. *Energy and Buildings*, 225, 110318. <https://doi.org/10.1016/j.enbuild.2020.110318>
- Ebrahimifakhar, A., Yuill, D., & Kabirikopaei, A. (2021). Application of Machine Learning Classification Methods in Fault Detection and Diagnosis of Rooftop Units. *18<sup>th</sup> International Refrigeration and Air Conditioning Conference at Purdue*, West Lafayette, IN.

- Ebrahimifakhar, A., Yuill, D., Smith, A., Granderson, J., Crowe, E., Chen, Y., & Reeve, H. (2021). Analysis of Automated Fault Detection and Diagnostics Records as an Indicator of HVAC Fault Prevalence: Methodology and Preliminary Results. *6<sup>th</sup> International High Performance Buildings Conference at Purdue*, West Lafayette, IN.
- Felts, D., & Bailey, P. (2000). The state of affairs - packaged cooling equipment in California. *Proceedings of the 2000 ACEEE Summer Study on Energy Efficiency in Buildings*, 3, 137-147.
- Feng, M. Y., Roth, K. W., Westphalen, D., & Brodrick, J. (2005). Packaged Rooftop Units: Automated Fault Detection and Diagnostics. *ASHRAE Journal*, 47(4), 68–70.
- Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15, 3133–3181.
- Frank, S., Lin, G., Jin, X., Singla, R., Farthing, A., Zhang, L., & Granderson, J. (2019). Metrics and methods to assess building fault detection and diagnosis tools.
- Granderson, J., Kramer, H., Lin, G., Curtin, C., & Crowe, E. (2020). Proving the business case for building analytics. *Lawrence Berkely National Laboratory*, Berkeley, CA, USA.
- Granderson, J., Lin, G., Harding, A., Im, P., & Chen, Y. (2020). Building fault detection data to aid diagnostic algorithm creation and performance testing. *Scientific Data*, 7(1), 1–14. <https://doi.org/10.1038/s41597-020-0398-6>
- Gunay, H. B., Shen, W., & Yang, C. (2019). Text-mining building maintenance work orders for component fault frequency. *Building Research & Information*, 47(5), 518–533. <https://doi.org/10.1080/09613218.2018.1459004>
- Han, H., Gu, B., Kang, J., & Li, Z. R. (2011a). Study on a hybrid SVM model for chiller FDD applications. *Applied Thermal Engineering*, 31(4), 582–592. <https://doi.org/10.1016/j.applthermaleng.2010.10.021>
- Han, H., Gu, B., Wang, T., & Li, Z. R. (2011b). Important sensors for chiller fault detection and diagnosis (FDD) from the perspective of feature selection and machine learning. *International Journal of Refrigeration*, 34(2), 586–599. <https://doi.org/10.1016/j.ijrefrig.2010.08.011>
- Hewett, M., Bohac, D., Landry, R., Dunsworth, T., Englander, S., & Peterson, G. (1992). Measured Energy and Demand Impacts of Efficiency Tune-Ups for Small Commercial Cooling Systems. in *ACEEE Conference Proceedings*.
- Hu, Y., & Yuill, D. P. (2021). Effects of multiple simultaneous faults on characteristic fault detection features of a heat pump in cooling mode. *Energy and Buildings*, 251, 111355. <https://doi.org/10.1016/J.ENBUILD.2021.111355>
- Hu, Y., Yuill, D. P., Rooholghodos, S. A., Ebrahimifakhar, A., & Chen, Y. (2021).

- Impacts of simultaneous operating faults on cooling performance of a high efficiency residential heat pump. *Energy and Buildings*, 242, 110975. <https://doi.org/10.1016/J.ENBUILD.2021.110975>
- Isermann, R. (2006). *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.
- Japkowicz, N. (2000). The Class Imbalance Problem: Significance and Strategies. *Proceedings of the 2000 International Conference on Artificial Intelligence*, 111–117.
- Katipamula, S., & Brambley, M. R. (2005). Review article: Methods for fault detection, diagnostics, and prognostics for building systems—A review, part I. *HVAC and R Research*, 11(1), 3–25. <https://doi.org/10.1080/10789669.2005.10391123>
- Kim, J., Trenbath, K., Granderson, J., Chen, Y., Crowe, E., Reeve, H., Newman, S., & Ehrlich, P. (2021). Research challenges and directions in HVAC fault prevalence. *Science and Technology for the Built Environment*, 27(5), 624–640. <https://doi.org/10.1080/23744731.2021.1898243>
- Kim, W., & Braun, J. E. (2020). Development, implementation, and evaluation of a fault detection and diagnostics system based on integrated virtual sensors and fault impact models. *Energy and Buildings*, 228, 110368. <https://doi.org/10.1016/J.ENBUILD.2020.110368>
- Li, D., Hu, G., & Spanos, C. J. (2016). A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis. *Energy and Buildings*, 128, 519–529. <https://doi.org/10.1016/j.enbuild.2016.07.014>
- Li, H., & Braun, J. E. (2007a). A methodology for diagnosing multiple simultaneous faults in vapor-compression air conditioners. *HVAC and R Research*, 13(2), 369–395. <https://doi.org/10.1080/10789669.2007.10390959>
- Li, H., & Braun, J. E. (2007b). An overall performance index for characterizing the economic impact of faults in direct expansion cooling equipment. *International Journal of Refrigeration*, 30(2), 299–310. <https://doi.org/10.1016/j.ijrefrig.2006.07.026>
- Li, H., & Braun, J. E. (2007c). Decoupling features and virtual sensors for diagnosis of faults in vapor compression air conditioners. *International Journal of Refrigeration*, 30(3), 546–564. <https://doi.org/10.1016/j.ijrefrig.2006.07.024>
- Li, H., & Braun, J. E. (2007d). Economic evaluation of benefits associated with automated fault detection and diagnosis in rooftop air conditioners. *ASHRAE Transactions*, 113 PART 2, 200–210.
- Li, H., & Braun, J. E. (2009a). Decoupling features for diagnosis of reversing and check

- valve faults in heat pumps. *International Journal of Refrigeration*, 32(2), 316–326. <https://doi.org/10.1016/j.ijrefrig.2008.05.005>
- Li, H., & Braun, J. E. (2009b). Development, evaluation, and demonstration of a virtual refrigerant charge sensor. *HVAC and R Research*, 15(1), 117–136. <https://doi.org/10.1080/10789669.2009.10390828>
- Li, H., & Braun, J. E. (2009c). Virtual refrigerant pressure sensors for use in monitoring and fault diagnosis of Vapor-Compression equipment. *HVAC and R Research*, 15(3), 597–616. <https://doi.org/10.1080/10789669.2009.10390853>
- Li, Y., & O'Neill, Z. (2019). An innovative fault impact analysis framework for enhancing building operations. *Energy and Buildings*, 199, 311–331. <https://doi.org/10.1016/j.enbuild.2019.07.011>
- Liu, M., Athar, A., Zhu, Y., & Claridge, D. E. (1995). Reduce Building Energy Consumption by Improving the Supply Air Temperature Schedule and Recommissioning the Terminal Boxes.
- Madani, H. (2014). The common and costly faults in heat pump systems. *Energy Procedia*, 61, 1803–1806. <https://doi.org/10.1016/j.egypro.2014.12.217>
- Newman, S., Lerond, J., Reeve, H., Vrabie, D., Belew, S., & Tucker, J. (2020). Pilot Study for Determining HVAC Fault Prevalence from Fault Monitoring Data.
- Qin, J., & Wang, S. (2005). A fault detection and diagnosis strategy of VAV air-conditioning systems for improved energy and control performances. *Energy and Buildings*, 37(10), 1035–1048. <https://doi.org/10.1016/j.enbuild.2004.12.011>
- R Core Team (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. URL <https://www.R-project.org/>.
- Rossi, T. M., & Braun, J. E. (1997). A statistical, Rule-Based fault detection and diagnostic method for vapor compression air conditioners. *HVAC and R Research*, 3(1), 19–37. <https://doi.org/10.1080/10789669.1997.10391359>
- Shoukas, G., Bianchi, M., & Deru, M. (2020). Analysis of Fault Data Collected from Automated Fault Detection and Diagnostic Products for Packaged Rooftop Units. <https://doi.org/10.2172/1665808>
- Stouppe, D. E., & Lau, Y. S. (1989). Air conditioning and refrigeration equipment failures. *National Engineer*, 93(9), 14–17.
- Touzani, S., Ravache, B., Crowe, E., & Granderson, J. (2019). Statistical change detection of building energy consumption: Applications to savings estimation. *Energy and Buildings*, 185, 123–136. <https://doi.org/10.1016/j.enbuild.2018.12.020>
- Weiss, G. M., & Provost, F. (2001). The effect of class distribution on classifier learning: an empirical study. <https://doi.org/10.7282/T3-VPFW-SF95>

- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
- Yoshida, H., Iwami, T., Yuzawa, H., & Suzuki, M. (1996). Typical faults of air-conditioning systems, and fault detection by ARX model and extended Kalman filter. *ASHRAE Transactions* 102(1), 557-564.
- Yuill, D. P., & Braun, J. E. (2016). Effect of the distribution of faults and operating conditions on AFDD performance evaluations. *Applied Thermal Engineering*, 106, 1329–1336. <https://doi.org/10.1016/j.applthermaleng.2016.06.149>
- Yuill, D. P., & Braun, J. E. (2017). A figure of merit for overall performance and value of AFDD tools. *International Journal of Refrigeration*, 74, 649–659. <https://doi.org/10.1016/j.ijrefrig.2016.11.015>
- Yuill, D. P., Cheung, H., & Braun, J. E. (2014). Validation of a Fault-Modeling Equipped Vapor Compression System Model Using a Fault Detection and Diagnostics Evaluation Tool. *International Refrigeration and Air Conditioning Conference*.
- Zhao, Y., Wang, S., & Xiao, F. (2013). Pattern recognition-based chillers fault detection method using Support Vector Data Description (SVDD). *Applied Energy*, 112, 1041–1048. <https://doi.org/10.1016/j.apenergy.2012.12.043>
- Zhao, Y., Xiao, F., Wen, J., Lu, Y., & Wang, S. (2014). A robust pattern recognition-based fault detection and diagnosis (FDD) method for chillers. *HVAC and R Research*, 20(7), 798–809. <https://doi.org/10.1080/10789669.2014.938006>

## APPENDIX A - Monthly Fault Presence (Metric 1) for AHU Faults

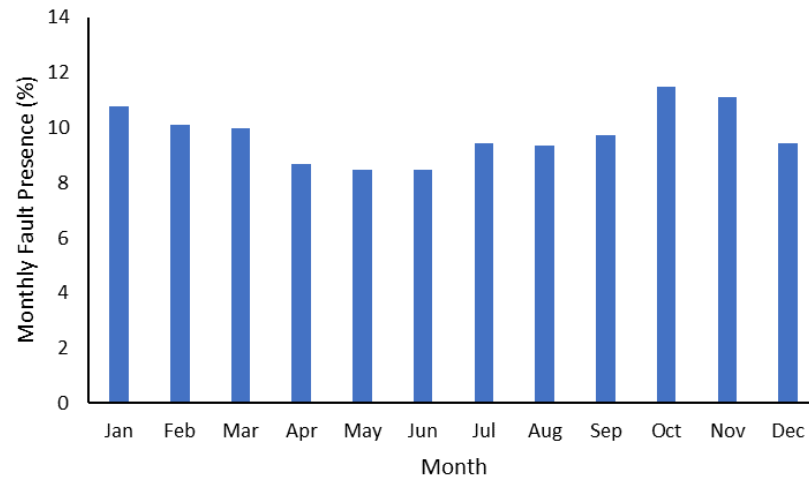


Figure A-1: Monthly fault presence (metric 1) for AHU damper stuck for FDD provider A

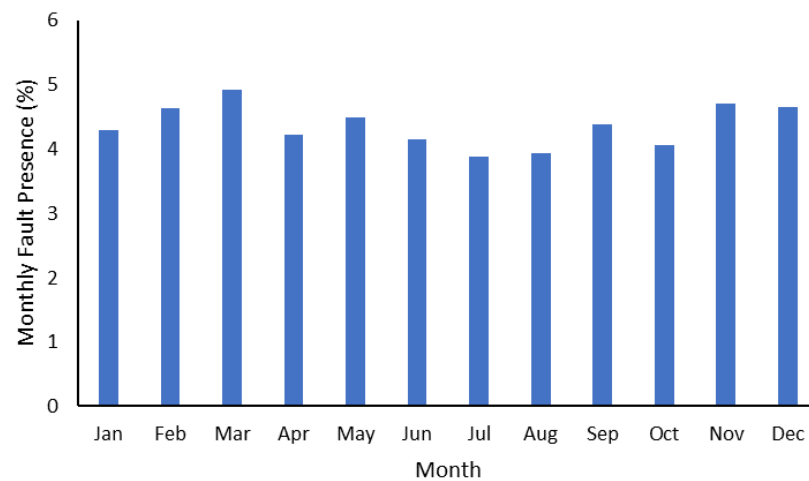


Figure A-2: Monthly fault presence (metric 1) for AHU supply air static pressure sensor drift for FDD provider A

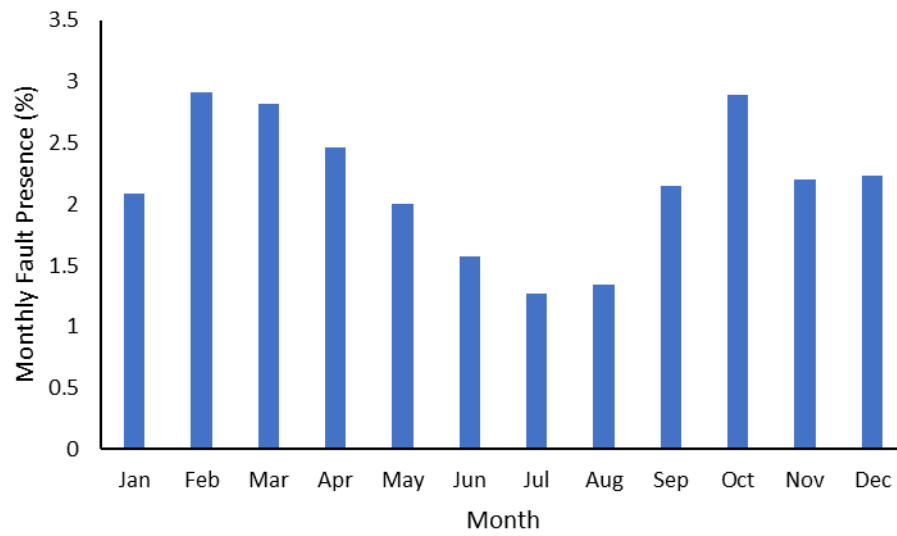


Figure A-3: Monthly fault presence (metric 1) for AHU heating coil valve leakage for FDD provider A

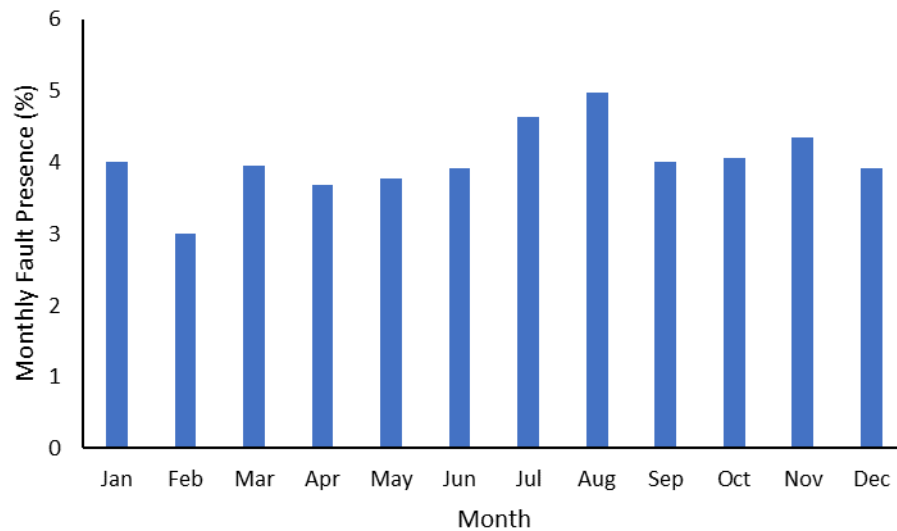


Figure A-4: Monthly fault presence (metric 1) for AHU fan hunting for FDD provider A

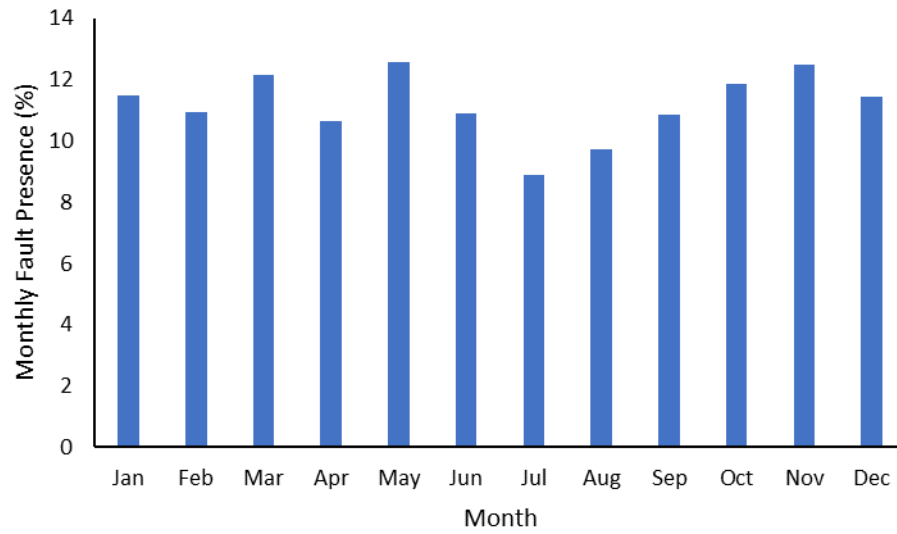


Figure A-5: Monthly fault presence (metric 1) for AHU supply air temperature abnormal for FDD provider

A

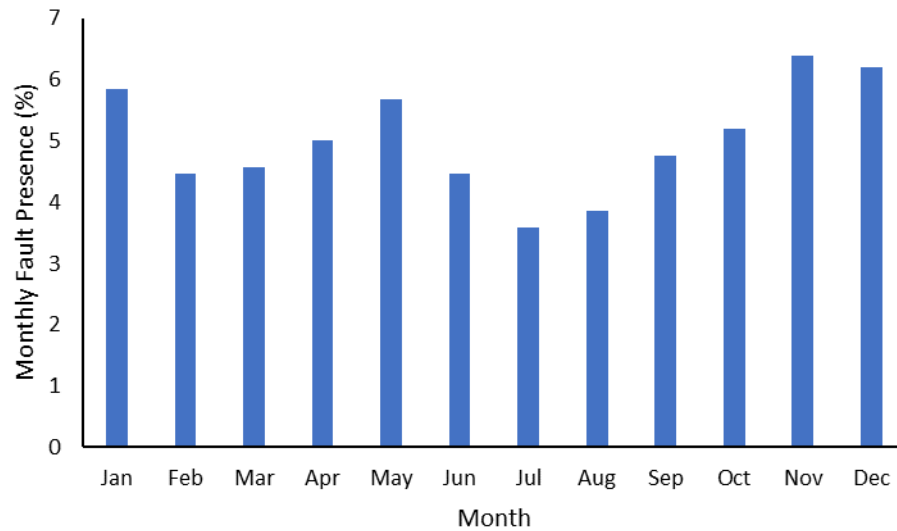


Figure A-6: Monthly fault presence (metric 1) for AHU coil valve hunting for FDD provider A

## APPENDIX B - Monthly Fault Presence (Metric 1) for ATU Faults

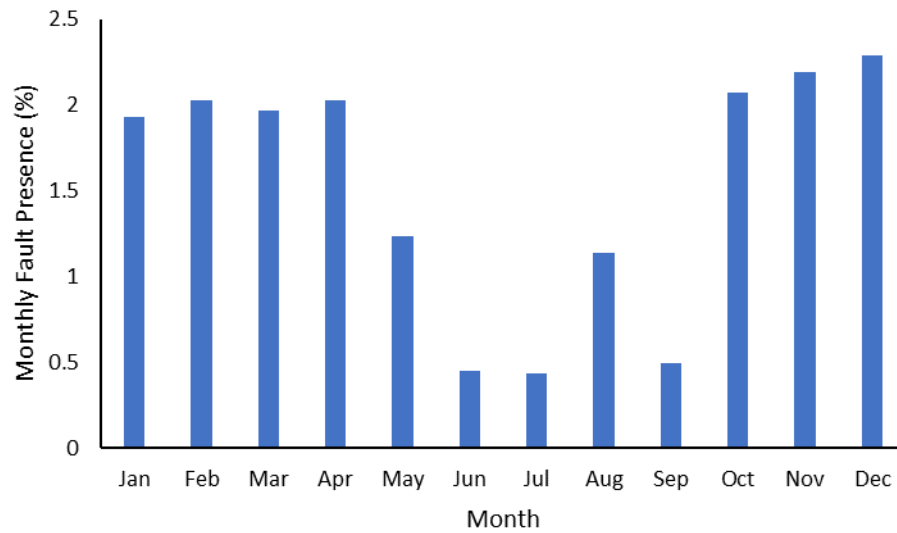


Figure B-1: Monthly fault presence (metric 1) for ATU discharge air damper stuck for FDD provider A

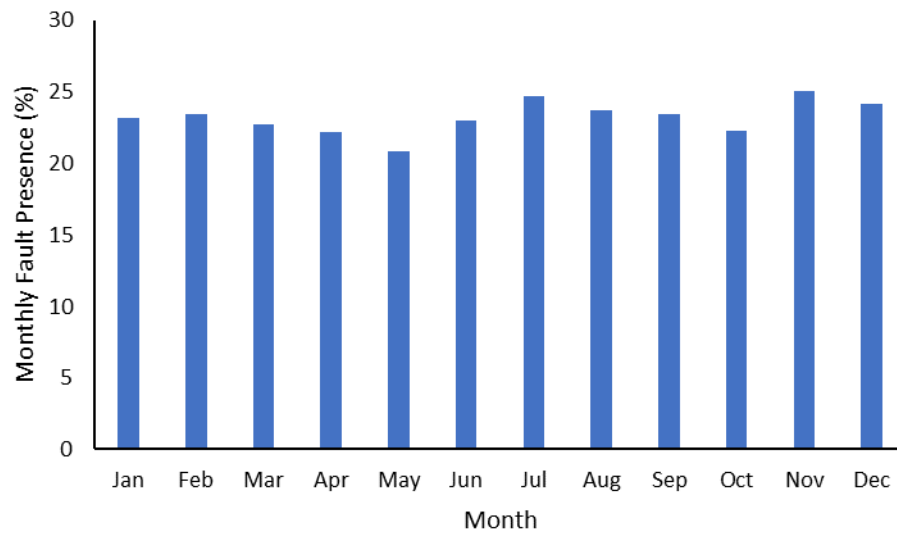


Figure B-2: Monthly fault presence (metric 1) for ATU zone temperature abnormal for FDD provider A

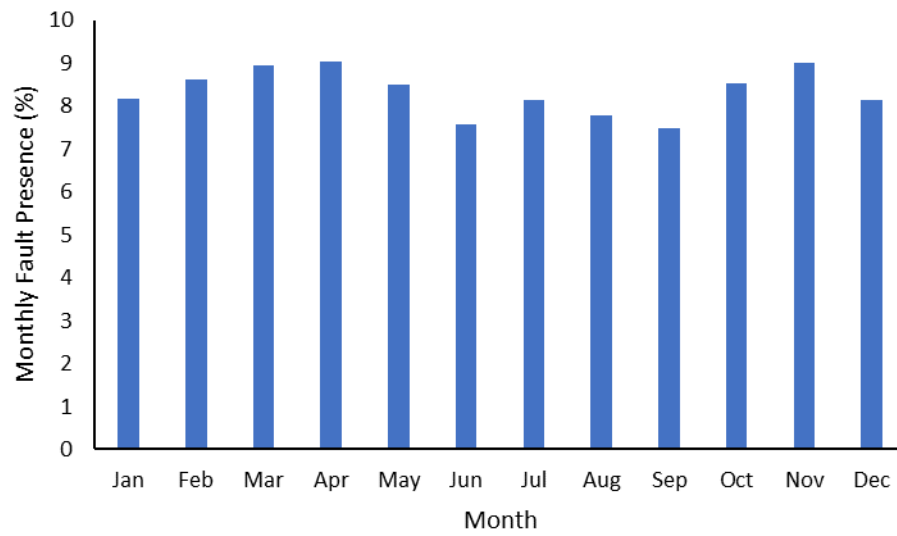


Figure B-3: Monthly fault presence (metric 1) for ATU reheat coil valve hunting for FDD provider A

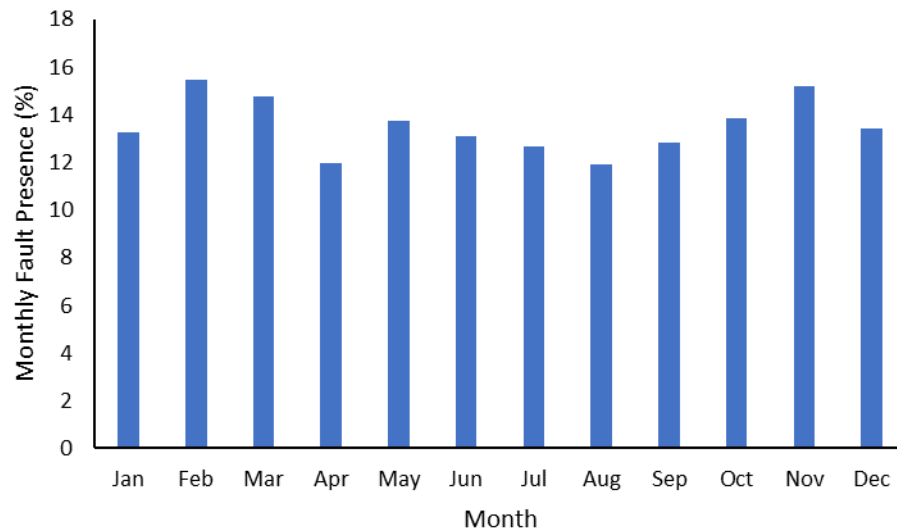


Figure B-4: Monthly fault presence (metric 1) for ATU discharge air temperature abnormal for FDD provider A