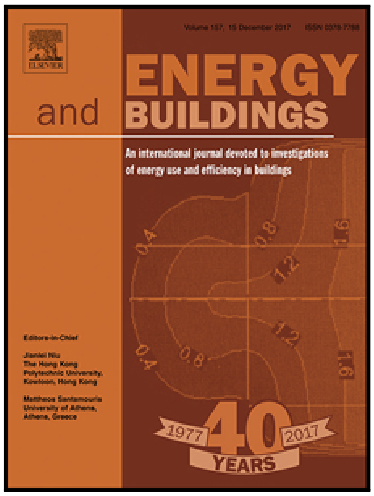
Accepted Manuscript



Semi-supervised Learning for Early Detection and Diagnosis of Various Air Handling Unit Faults

Ke Yan, Chaowen Zhong, Zhiwei Ji, Jing Huang

PII: S0378-7788(18)32025-5
DOI: https://doi.org/10.1016/j.enbuild.2018.10.016
Reference: ENB 8845

To appear in: *Energy & Buildings*

Received date: 4 July 2018
Revised date: 24 August 2018
Accepted date: 15 October 2018

Please cite this article as: Ke Yan, Chaowen Zhong, Zhiwei Ji, Jing Huang, Semi-supervised Learning for Early Detection and Diagnosis of Various Air Handling Unit Faults, *Energy & Buildings* (2018), doi: https://doi.org/10.1016/j.enbuild.2018.10.016

Highlights

- The work introduces a novel semi-supervised approach to detect and diagnose faults for AHUs.
- 80% accuracy rate is reached using a training set with 8000 normal samples and only around 30 samples for each fault type.
- This work addresses the tradeoff between the initial number of faulty samples and the final classification accuracy.
- This work addresses the tradeoff between the initial number of faulty samples and the computational cost.
- This work addresses the tradeoff between the threshold of confidently levels and the final classification accuracy.

Semi-supervised Learning for Early Detection and Diagnosis of Various Air Handling Unit Faults

Ke Yan[a], Chaowen Zhong[a], Zhiwei Ji[b], Jing Huang[b,*]

[a] *College of Information Engineering, China Jiliang University, 258 Xueyuan Street, Hangzhou, China, 310018*

[b] *School of Information & Electronic Engineering, Zhejiang Gongshang University, 18 Xuezheng Road, Hangzhou, China, 311300*

Abstract

Modern data-driven fault detection and diagnosis (FDD) techniques show impressive high diagnostic accuracy in recognizing various air handling units (AHUs) faults. Most existing data-driven FDD approaches simply adopt supervised machine learning techniques that presume the availability of sufficient number of faulty training data samples. However, in real-world AHU FDD scenarios, the number of faulty training samples is not enough to support supervised learning methods, since faults are usually fixed within short periods of time. In this study, a semi-supervised learning FDD framework is proposed to deal with the above problem. By using the proposed framework, the training pool can be enriched by iteratively inserting confidently labeled testing samples, which mimics the scenario of detecting faults the earliest possible. Furthermore, the proposed framework can be easily extended with various kinds of state-of-art classifiers. Three important tradeoffs are observed through a series of experiments. With a reasonably small number of faulty training data samples available, the performance of the proposed semi-supervised learning technique is comparable to the classic supervised FDD methods.

Keywords: Air Handling Unit, Fault Detection and Diagnosis, Semi-supervised Learning, Support Vector Machine

*Corresponding author.
Email address: gabriel.jing.huang@gmail.com (Jing Huang)

1. Introduction

Fault detection and diagnosis (FDD) at early stage is mandatory for critical components of heating ventilation air-conditioning (HVAC) systems, such as air handling units (AHUs). An effective FDD method of HVAC subsystems offers proper maintenance of those important components, prevents further damage, maintain indoor air quality and saves 15% to 30% of total energy consumed by buildings [1]. As one of the most extensive and important components for a HVAC system, AHU supplies indoor environment with conditioned fresh air with a supply fan and exhausts the return air using a return air fan. The main components of AHU include fans, one cooling or heating coil, ducts and dampers. In general, AHU maintains the indoor air quality, temperature, humidity and indoor comfort. Regular monitoring with automatic FDD technique is highly recommended for AHUs.

Traditional FDD (such as statistical analysis) can be extremely tedious for large-scale HVAC sub-systems, such as AHUs and chillers [2]. Artificial intelligence (AI) enhanced early detection of various AHU faults based on existing historical data represents an important next generation fault detection and diagnosis (FDD) technique for HVAC maintenance, energy and labour savings, which attracts enormous attentions throughout the world over the past few decades. Among all AI enhanced data-driven FDD methods, the supervised machine learning technique provides impressive and promising FDD results for HVAC components, such as chillers and AHUs. The detection and diagnosis accuracy rate can be as high as over 93%, as reported by existing works published in the recent five years [3, 4, 5, 6].

However, there is still a large gap between theoretical data-driven FDD methods and practical industrial FDD applications. Supervised machine learning techniques require a substantial amount of the faulty data samples available in the training dataset, which can hardly be the case in real-world situations. In real-world scenarios, faults are usually fixed within a short period of time. A practical FDD method is demanded to detect and diagnose various AHU faults with smallest possible number of faulty samples available.

In this study, a semi-supervised FDD framework is proposed which intends to use the minimal number of faulty samples for training, detects and diagnose possible faults as early as possible for AHUs. Moreover, through a series of experiments, three tradeoffs are observed, which include:

- The tradeoff between the initial number of faulty samples selected in the training dataset and the final classification accuracy.

- The tradeoff between the initial number of faulty samples selected in the training dataset and the number of iterations (computational cost).
- The tradeoff between the threshold of confidently levels and the final classification accuracy.

The experimental results of this study can be useful under two scenarios:

- **Scenario 1** There are a number of faults experienced by a particular AHU; however, the number of training samples for each fault is small, i.e., the faults are always fixed within a short period of time. When another fault occurs, it is unclear that whether a supervised FDD method is able to identify the fault correctly. The experiments of this study suggest the minimal number of faulty training samples for an indicated data-driven FDD method to correctly identify the fault types.
- **Scenario 2** There are absolutely not enough faulty training samples available in the database. When another fault occurs, the building management system (BMS) requires some response time to recursively identify the fault type using the proposed semi-supervised learning framework. New faulty samples can be absorbed into the training data pool with high confidence level. The experiments of this study suggest the minimal response time for the proposed FDD frmework to correctly identify the fault types while there are not enough faulty training samples available in the training pool.

The proposed semi-supervised FDD framework requires a base classifier, which is supposed to be able to diagnose existing fault types efficiently with relatively small quantity training samples available. As a modern machine learning technique and data-driven method, support vector machine (SVM) is employed in many recently published works, and has demonstrated its classification ability in the field of FDD for HVAC subsystems [4, 7, 8]. As a result, we use SVM as a base classifier for the proposed semi-supervised classification framework. The source code of our program is freely available at our project webpage at: http://keddiyan.com/files/SemiSupervisedAHU.html. We also show the ease of extending our program by replacing the SVM with other modern machine learning techniques in Section 4. The simulation performance shows an over 80% accuracy using a training set consisting of 8,000 normal samples and only around 30 samples for each fault type in summer, and close to 90% accuracy in winter. It is noted that the AHUs discussed in this study contain no energy recovery systems.

2. Related Works

Fault detection and diagnosis methods for various HVAC subsystems faults can be generally categorized into model-based approaches and data-driven approaches [9]. Model-based FDD methods define rules and thresholds to distinguish normal/abnormal situations. Sterling *et al.* [10] studies and compared both qualitative and quantitative models for AHU FDD. Simulation results showed similar FDD results from both model-based methods. Yan *et al.* [5] developed a rule-based system using decision tree for AHU FDD. The F-measure for fault diagnosis of AHU achieved 0.97 with nine different faults. Dey and Dong [11] built Bayesian belief network on top of the expert rules to diagnose faults in AHU. Wang and Chen [12] applied a residual-based exponentially weighted moving average (EWMA) control chart method to diagnose faults for variable-air-volume (VAV) AHUs. The experimental results showed that the EWMA chart effectively improved the performance of rule-based system for diagnosing multiple faults of VAV AHUs.

Data-driven FDD of AHUs and other components of the HVAC system has been studied for decades. As early as 1990s, Lee *et al.* [13] started to use artificial neural networks (ANNs) to diagnose faults in AHUs. Years later, the same group of researchers concluded the strengths and weaknesses of various classification methods for AHU FDD, which include the ANN model, K-nearest-neighbor algorithm, rule-based systems and Bayesian algorithms [14]. The classification accuracy is low and the variety of the classification methods is limited at that time. Lee *et al.* [15] improved their DD-FDD method using general regression neural networks and applied the upgraded method to subsystem level AHU faults. The classification performance starts to become acceptable. Besides neural networks, rule-based expert system is another commonly used tool that can provide high classification accuracy. House *et al.* [2] discussed and concluded a set of rules for AHU FDD. Schein *et al.* [16] introduced an expert system based on air handling unit performance assessment rules to detect faults in AHUs. Other than machine learning technologies, there are statistical data-driven methods for AHU FDD. Yoshida *et al.* [17] proposed an online variable air volume (VAV) AHU FDD method using recursive auto-regressive with exogenous variable (RARX) algorithm. The auto-regressive methods can be considered as gray box model, which are also built on historical data [18]. Du *et al.* [3] used combined neural networks to improve the robustness of AHU FDD. Recently, Mulumba *et al.*

[4] combined the auto-regressive methods with machine learning methods to diagnose faults in AHUs, and achieved diagnostic accuracy over 90%. Zhao *et al.* [6, 19] proposed a diagnostic bayesian network to diagnose faults from various sources of an AHU. Yan *et al.* [20] proposed a hybrid method combining extended Kalman filter and cost-sensitive SVM to diagnose various faults under different seasons for AHU. Experimental results show that the pre-processing of the historical data using statistical model can potentially ease the classification process.

The semi-supervised leaning represents one of the three main components of machine learning technology. The other two components are well-known as supervised learning and unsupervised learning [21, 22]. Semi-supervised learning learns information from unlabelled data and enriched the minority of the training dataset using highly confident testing data samples. Traditional semi-supervised learning methods include discriminative learning [23], co-training [24], transductive learning [25] and etc. Readers may note that the semi-supervised learning approach is not often adopted in the area of FDD, since the supervised learning methods generally provide more promising results; and the costs of false negatives are usually high [26]. Lemos *et al.* [27] developed a semi-supervised fault detection and diagnosis method with fuzzy rules. The developed approach was applied to FDD of a industrial actuator. All failure modes can be well identified without any prior knowledge given. Monroy *et al.* [28] combined Gaussian mixture model (GMM) and SVM to form a semi-supervised learning machine for fault diagnosis of chemical processes. Experimental results show that the semi-supervised learning enhanced the traditional classifiers' classification ability and is suitably applicable to the complex industrial environment. Li and Zhou [29] applied a semi-supervised learning method called Co-Forest, which combined the concepts of co-training and random forest, to computer-aided medical diagnosis applications using benchmark data.

3. Methodology

A semi-supervised data-driven FDD algorithm is proposed to detect and diagnose the HVAC faults acquiring only a few faulty training samples. By pre-processing the raw data, including normalization and a feature selection process, the original dataset is converted into a subset containing less feature variables in order to improve the training efficiency. The training pool only a few faulty samples for each fault, i.e., from 5 samples to 55 samples of each faulty dataset, along with a large number of normal data samples, mimicking the real-world early detection scenarios for AHU faults. In the testing phase, we purposely use equal size of normal testing samples (N) and testing samples of various fault types ($F_1, F_2, ..., F_n$). All classified tested samples will be assigned a confidence level after a label is assigned by the classifier. Only those classified samples with strong confidence levels will be assigned a label and inserted into the training pool. The rest testing samples will be classified again using the enlarged training pool. The experimental results simulate the diagnosis accuracy rates under various scenarios, while the number of faulty training data samples differ. The resulting statistics can be significantly useful for real-world applications measuring the number of faulty training samples required to detection specific faults. The overall flowchart of our algorithm is shown in Figure 1.

Preserved space for Figure 1.

3.1. Data Description

The AHU normal/faulty operational data is collected by Li *et al.* from 2007 to 2008, through a series experiments performed in the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) project No. RP-1312, titled 'Tools for Evaluating Fault Detection and Diagnostic Methods for Air-Handling Units' and published in 2010 [30, 31, 32]. In the experimental phase of ASHRAE project No. RP-1312, two identical AHUs are utilized to generate sensor data, such as temperature, humidity, fan speed and etc. One AHU, named AHU-A, ran normally and generated normal conditional data; the other AHU, named AHU-B, ran with artificially added faults, such as damper stuck, valve stuck and duct leaking, to generate data under various faulty conditions. Three months data were collected for summer, spring and winter, respectively. Since, the HVAC load is usually maximized during the summer and winter seasons, the data from these two seasons is taken to analyze the three tradeoffs stated at the end of Section 1.

There are in total thirteen faults simulated in dataset collected by ASHRAE project No. RP-1312 during the summer season, where we select six typical faults to perform multi-class classification. We list the six selected fault types and enumerate them from F1 to F6:

1. Exhausted air (EA) damper stuck (fully open);

2. Return fan at fixed speed;
3. Cooling coil valve control unstable;
4. Cooling coil valve partially closed (15% open);
5. Outdoor air damper leak;
6. AHU duct leaking (after supply fan (SF)).

Similarly, six typical faults from the winter season dataset are selected and enumerated from F7 to F12:

1. Outdoor air damper stuck (fully close);
2. Outdoor air damper leaking (62% open);
3. Exhaust air damper stuck (fully open);
4. Exhaust air damper stuck (fully close);
5. Heating coil fouling;
6. Heating coil reduced capacity.

Together with the normal operational data type, the multi-class fault diagnosis becomes a 7-class classification problem for both summer and winter seasons, respectively (Section 3.3). There are in total 21,600 normal operational data, accompanied with 1440 faulty samples for each fault. In the preprocessing phase, we reorganize the original data in order to learn the minimal number of faulty samples needed for SVM to correctly diagnose the AHU faults. The initial training pool contains a large number of normal operational AHU data together with only a small number of samples for each fault. For each experimental simulation, the initial number of normal data samples in the training pool is around 8,000; and the initial number of faulty data samples in the training pool is from 5 to 55. In the testing phase, each testing set contains equal numbers of normal and faulty samples to give a fair evaluation of the proposed semi-supervised SVM performance.

3.2. Feature Selection using Cost-sensitive Sequential Forward Feature Selection

Feature selection is an important pre-processing step for data-driven fault detection and diagnosis of HVAC subsystems. It reduces the data dimension for classification and consequently improves the efficiency as well as possibly increases the resulting classification accuracy. In the original dataset collected by ASHRAE project No. RP-1312, there are in total 102 features measured, which introduce a heavy load for SVM to build a hyper-plane in such a high dimension [33]. Recently, Yan *et al.* [34] proposed a cost-sensitive sequential forward feature selection (CS-SFS) algorithm that is suitable to select features from imbalanced datasets using SVM as a base classifier. In this study, we employ the CS-SFS algorithm to select the most effective features for AHU diagnosis from the original 102 features.

The CS-SFS extends the classic sequential forward feature selection algorithm and emphasizes those falsely classified faulty samples rather than falsely classified normal samples, while calculating the precision values [35]. The CS-SFS algorithm starts from a small feature subset containing one or two features and gradually expand the feature subset to have more important features. In this study, we select T_{sa} (supply air temperature) and T_{ra} (return air temperature) as the initial feature set and use SVM as the measuring classifier. The top most important features selected for the summer and winter datasets are listed in Tables 1 and 2, respectively.

Preserved space for Table 1.

Preserved space for Table 2.

The semantics of the selected features and the basic structure of the tested AHU are shown in Figure 2.

Preserved space for Figure 2.

3.3. Support Vector Machine and Its Extensions

Support vector machine (SVM) is one of the most commonly used machine learning techniques for small to medium size data classification problems and has been applied to FDD approaches for HVAC subsystems in existing state-of-art works [4, 7, 8]. Given a set of multi-dimensional data with only two types, i.e., true and false, SVM finds a hyperplane in the high-dimensional feature space to separate the two types of data. And the hyperplane can be further used to label new testing samples. In the process of finding the hyperplane, or in the training process, three parameters have to be optimized, which are C, γ and the type of kernel [36]. The parameters have to be optimized every time when the training data varies and the parameter optimization processes can be time consuming for semi-supervised SVM.

Multi-class SVM extends the traditional binary SVM to deal with datasets containing more than two types. There are in general two approaches available to build the multi-class SVM, namely, one-against-one and one-against-all, where we adopted the one-against-all approach in this study, because of its robustness for imbalanced datasets [37]. Figure 3 shows the confusion matrix size difference between the binary and a 7-class classification. The red slots represent all falsely classified testing samples. For each testing sample, the chance of incorrect labelling for multi-class SVM is much higher than binary SVM [38].

Preserved space for Figure 3.

Semi-supervised multi-class SVM technique is employed in this study. Besides C and γ, the semi-supervised SVM requires one more parameter to be optimized, which is the confidence level threshold (refer to Figure 1). In the testing phase, the traditional SVM assigns a predicted label to each testing sample with a confidence level that indicates how confident the sample belongs to the labeled type. The confidence level is also called 'classification probability', which is a proportion value between 0 and 1. After all confidence levels are generated for available testing samples, a threshold is tuned to separate the labeled testing samples into two groups, namely, high confident testing sample group and low confident testing sample group. At the end of each iteration (all testing samples been tested), the high confident testing sample group will be inserted into the training pool to enrich the minority training groups; and the whole semi-supervised SVM is also revised by calculating new C and γ values.

3.4. A Semi-supervised FDD Approach for AHU Faults

Algorithm 1 formally describes the proposed semi-supervised FDD approach for AHU faults. After the preprocessing phase (as stated in Figure 1), the processed original dataset can be divided into training and testing sets, where the training set contains 5, 10, 15, 20, ... or 55 faulty training samples, accompanied with 8,000 normal training samples. The testing set contains around 1,400 faulty testing samples with over 8,000 normal testing samples (the number of samples ratio between faulty testing samples and normal testing samples is 1:1) for both summer and winter datasets. The entire semi-supervised FDD algorithm is bounded by a while-loop. In each iteration, the SVM model is first trained using the revised training set. Then, each testing sample is assigned a predicted label with a confidence level by the trained SVM model. All testing samples with confidence levels greater than a pre-defined threshold will be inserted into the training set (marked as 'labeled' and will not be tested again); and the predicted labels are treated as actual labels in next round SVM model training. All testing samples with confidence levels less than the pre-defined threshold will be tested again in the next iteration. The whole process terminates when all testing samples are marked 'labeled' or the maximum number of iteration is reached.

Algorithm 1 Semi-supervised FDD Approach for AHU Faults

Input: 1. The training set **TR** contains only a few faulty training samples with a large number of normal training samples.
2. The testing set **TE** contains equal numbers of faulty and normal samples.
Output: Prediction label set **PL** for all testing samples.

```
Initialization: mark all testing samples as 'unlabeled'.
while There are 'unlabeled' testing samples & the number of iteration
  ≤ the maximum number allowed do
  Train the multi-class SVM model with TR.
  for each 'unlabeled' testing sample x in TE do
    Collect the prediction label l and confidence level δ for x from the
      trained SVM model.
    if δ ≥ pre-defined confidence level threshold then
      Mark x as 'labeled'.
      Insert tuple (x, l) into TR.
    end if
  end for
end while
```

The proposed semi-supervised FDD approach differs with the traditional FDD approaches for AHUs, e.g., [4], from three perspectives. First, a semi-supervised SVM is designed to deal with the imbalanced training dataset that contains a large number of normal data samples with only a few faulty samples. Highly confident testing samples with fault types been labeled can be absorbed into the training pool to enrich the faulty training dataset. Second, once the highly confident testing samples been inserted into the training pool, the parameters of the SVM have to be revised; and the time complexity increases. The number of iteration has to be minimized with reasonable parameter settings, e.g., a suitable confidence level threshold value. Third, most existing works, such as [4] and [20], separate the detection and diagnosis phases and use binary SVM in the detection phase. In this study, we employ the multi-class SVM to classify both normal and faulty classes simultaneously. The confusion matrix size has been increased, which introduces extra difficulties for our approach to achieve competitive diagnostic accuracy.

4. Experimental Analysis

In this section, we answer the three questions raised in Section 1, which are the three tradeoffs between the number of initial faulty training samples, FDD accuracy and confidence level threshold selection. Each experiment is performed twice for both summer and winter seasons to justify our conclusions. For each season, the original dataset is re-organized into eleven training subsets, respectively. Each training subset consists of 8,000 normal data samples with 5, 10, 15, 20, ..., or 55 faulty data samples for each fault type. For example, if there are 5 faulty data samples for each fault type, the training subset contains 5×6 (faults) = 30 faulty data samples. All training faulty data samples are randomly selected from the original summer or winter dataset from ASHRAE project No. RP-1312. Each testing subset contains equal numbers of normal and faulty samples. The proposed semi-supervised SVM is compared with existing modern semi-supervised methods, including semi-supervised extreme learning machine (ELM) [39], decision tree (CART) [40], k-nearest-neighbor (KNN) [41] and random forest (RF) [42]. The fairness of experimental result comparison is guaranteed by 30 times repetition of each experiment.

4.1. The Number of Initial Faulty Training Samples v.s. FDD Accuracy

We implement four state-of-art semi-supervised methods and compare the FDD accuracy with the proposed semi-supervised SVM method. The details of the implementations and the project page can be found on our website at: http://keddiyan.com/files/SemiSupervisedAHU.html. The semi-supervised SVM implementation can be easily extended to other semi-supervised methods by replacing the SVM in Figure 1 with other classic machine learning models, such as KNN, CART, RF and ELM. The classification accuracy rates for various numbers of initial faulty samples in the summer and winter seasons are listed in Tables 3 and 4, respectively. It is noted that each accuracy rate is an average rate over 30 times repetitive simulations.

Preserved space for Table 3.

Preserved space for Table 4.

Figures 4 and 5 show the FDD accuracy rates of the five semi-supervised machine learning methods in the summer and winter seasons, respectively. For both summer and winter datasets, the proposed semi-supervised SVM outperforms the existing state-of-art methods, when the number of initial faulty training samples is greater than 10. For summer dataset, the semi-supervised ELM outperforms our method for 5 initial faulty training samples, but goes down when the number of faulty samples increases. The proposed method is able to achieve over 80% accuracy rate with only 30 initial faulty training samples for each fault type in summer and close to 90% in winter. Moreover, while the number of initial faulty samples keeps increasing, the FDD accuracy tends to be stabilized and approaches to the classification accuracy of supervised learning. From Figures 4 and 5, suggestion can be made that the minimal number of initial faulty training samples for an effective semi-supervised AHU FDD is around 30 for each fault type.

Preserved space for Figure 4.

Preserved space for Figure 5.

4.2. The Number of Initial Faulty Training Samples v.s. Computational Complexity

The second tradeoff (as shown in Section 1) concerns about the computational complexity of the semi-supervised algorithm. The actual time consumed by each method depends on the hardware configuration of the machine. For modern machine learning tools, such as SVM, ELM, KNN, CARt and RF, the training time for a medium size dataset with less than 10,000 data samples is less than 5 seconds. Therefore, in this subsection, we only show the number of iterations (training processes) for each compared semi-supervised method (Tables 5 and 6).

Preserved space for Table 5.

Preserved space for Table 6.

It can be seen that, in order to achieve competitive FDD accuracy, the number of iterations increases. For 30 initial faulty samples for each fault, the semi-supervised SVM takes 27.1 iterations in average for the summer dataset and 10.9 for the winter dataset. In a standard lab machine with i7-7700 and 8G RAM, the running time is less than 2 minutes, which is still in the acceptable range for real-world industrial applications.

4.3. Confidence Level Threshold v.s. FDD Accuracy

An appropriate confidence level threshold maximizes the semi-supervised SVM classification accuracy and also increases the overall FDD efficiency. In this study, we select the confidence level thresholds by investigating the highest classification accuracy versus various threshold values with different initial faulty training sample numbers in the training subset. For training subsets with 10, 20, 30, 40 and 50 initial faulty samples for each fault type, the optimized confidence level threshold is always from 0.9 to 0.95, which suggests that only highly confident testing samples can be inserted into the training pool in order to maximize the overall FDD accuracy. Figure 6 depicts a general relationship between confidence level thresholds and averaged diagnostic accuracy for AHU faults using 10, 20, 30, 40 and 50 faulty training samples in the training dataset for the summer season dataset. It can be seen that the confidence level threshold tuning is important to obtain competitive diagnostic performance for the proposed semi-supervised FDD approach.

Preserved space for Figure 6.

5. Conclusion and Future Works

Traditional supervised machine learning techniques showed an impressive high diagnosis accuracy for various HVAC faults [3, 4, 5, 6]. However, research gaps exist while there are not enough training data from the fault types in real-world industrial applications. Two questions arise: 1) How do we apply existing supervised machine learning methods to cases when there are not enough faulty training samples? 2) How many faulty training samples are called 'enough' to make accurate FDD predictions?

This study answers the above two questions and fills an important gap between traditional theoretical supervised FDD methods and practical FDD applications for various AHU faults. The proposed semi-supervised SVM solves the above problem and demonstrates the classification performance with between 80% and 89% accuracy rates using a training set consisting of 8,000 normal samples and only 30 samples for each fault type in both summer and winter seasons. Furthermore, three tradeoffs (as shown in Section 1) are addressed. Experiments are performed with different numbers of initial faulty training samples to show the minimal number of faulty samples required to diagnose a particular fault. There are also suggestions provided about the most appropriate confidence level threshold balancing the tradeoff between classification accuracy and efficiency. The proposed semi-supervised FDD framework can be easily extended by replacing the SVM with other modern machine learning techniques. This is one reason that we make our source code freely available to the public. In Section 4, we show other semi-supervised methods with modern classifiers, such as KNN, decision trees, random forest and ELM.

The future work of this study is to extend the current work to AHUs with energy recovery systems. Another research direction is to perform unsupervised learning for HVAC subsystem FDD, which detects and diagnoses faults without any training samples. The state-of-art clustering methods provide hints to cluster unlabeled samples according to their feature space centroids. The specific way of labeling those clustered groups is left to be explored in our future studies.

Acknowledgment
This work was supported by the National Natural Science Foundation of China (grant numbers: 61850410531 and 61602431). Authors would like to express their appreciation to the ASHRAE project number 1312-rp for providing the AHU FDD datasets.

References

[1] S. Katipamula, M. R. Brambley, Methods for fault detection, diagnostics, and prognostics for building systemsa review, part I, Hvac&R Research 11 (1) (2005) 3–25.

[2] J. M. House, H. Vaezi-Nejad, J. M. Whitcomb, An expert rule set for fault detection in air-handling units/Discussion, Ashrae Transactions 107 (2001) 858.

[3] Z. Du, B. Fan, J. Chi, X. Jin, Sensor fault detection and its efficiency analysis in air handling unit using the combined neural networks, Energy and Buildings 72 (2014) 157–166.

[4] T. Mulumba, A. Afshari, K. Yan, W. Shen, L. K. Norford, Robust model-based fault diagnosis for air handling units, Energy and Buildings 86 (2015) 698–707.

[5] R. Yan, Z. Ma, Y. Zhao, G. Kokogiannakis, A decision tree based data-driven diagnostic strategy for air handling units, Energy and Buildings 133 (2016) 37–45.

[6] Y. Zhao, J. Wen, F. Xiao, X. Yang, S. Wang, Diagnostic Bayesian networks for diagnosing air handling units faults–part I: Faults in dampers, fans, filters and sensors, Applied Thermal Engineering 111 (2017) 1272–1286.

[7] W. Kim, S. Katipamula, A review of fault detection and diagnostics methods for building systems, Science and Technology for the Built Environment 24 (1) (2018) 3–21.

[8] P. M. Van Every, M. Rodriguez, C. B. Jones, A. A. Mammoli, M. Martínez-Ramón, Advanced detection of HVAC faults using unsupervised SVM novelty detection and Gaussian process models, Energy and Buildings 149 (2017) 216–224.

[9] Y. Yu, D. Woradechjumroen, D. Yu, A review of fault detection and diagnosis methodologies on air-handling units, Energy and Buildings 82 (2014) 550–562.

[10] R. Sterling, G. Provan, J. Febres, D. O'Sullivan, P. Struss, M. M. Keane, Model-based fault detection and diagnosis of air handling units: A comparison of methodologies, Energy Procedia 62 (2014) 686–693.

[11] D. Dey, B. Dong, A probabilistic approach to diagnose faults of air handling units in buildings, Energy and Buildings 130 (2016) 177–187.

[12] H. Wang, Y. Chen, A robust fault detection and diagnosis strategy for multiple faults of VAV air handling units, Energy and Buildings 127 (2016) 442–451.

[13] W.-Y. Lee, J. M. House, C. Park, G. E. Kelly, Fault diagnosis of an air-handling unit using artificial neural networks, ASHRAE Transactions 102 (1996) 540–549.

[14] J. M. House, W. Y. Lee, D. R. Shin, Classification techniques for fault detection and diagnosis of an air-handling unit, ASHRAE Transactions 105 (1999) 1087.

[15] W.-Y. Lee, J. M. House, N.-H. Kyong, Subsystem level fault diagnosis of a building's air-handling unit using general regression neural networks, Applied Energy 77 (2) (2004) 153–170.

[16] J. Schein, S. T. Bushby, N. S. Castro, J. M. House, A rule-based fault detection method for air handling units, Energy and buildings 38 (12) (2006) 1485–1492.

[17] H. Yoshida, S. Kumar, Y. Morita, Online fault detection and diagnosis in VAV air handling unit by RARX modeling, Energy and Buildings 33 (4) (2001) 391–401.

[18] A. Afshari, N. Liu, Inverse modeling of the urban energy system using hourly electricity demand and weather measurements, Part 2: Gray-box model, Energy and Buildings 157 (2017) 139–156.

[19] Y. Zhao, J. Wen, S. Wang, Diagnostic Bayesian networks for diagnosing air handling units faults–Part II: Faults in coils and sensors, Applied Thermal Engineering 90 (2015) 145–157.

[20] K. Yan, Z. Ji, H. Lu, J. Huang, W. Shen, Y. Xue, Fast and Accurate Classification of Time Series Data Using Extended ELM: Application in Fault Diagnosis of Air Handling Units, IEEE Transactions on Systems, Man, and Cybernetics: Systems PP (2017) 1–8, doi: 10.1109/TSMC.2017.2691774.

[21] S. B. Kotsiantis, I. Zaharakis, P. Pintelas, Supervised machine learning: A review of classification techniques, in: Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word Ai

Systems with Applications in Ehealth, Hci, Information Retrieval and Pervasive Technologies, 3–24, 2007.

[22] G. M. Y. Dauphin, X. Glorot, S. Rifai, Y. Bengio, I. Goodfellow, E. Lavoie, X. Muller, G. Desjardins, D. Warde-Farley, P. Vincent, et al., Unsupervised and transfer learning challenge: a deep learning approach, in: Proceedings of ICML Workshop on Unsupervised and Transfer Learning, 97–110, 2012.

[23] S. Bickel, M. Brückner, T. Scheffer, Discriminative learning for differing training and test distributions, in: Proceedings of the 24th international conference on Machine learning, ACM, 81–88, 2007.

[24] A. Blum, T. Mitchell, Combining labeled and unlabeled data with co-training, in: Proceedings of the eleventh annual conference on Computational learning theory, ACM, 92–100, 1998.

[25] Y. Chen, G. Wang, S. Dong, Learning with progressive transductive support vector machine, Pattern Recognition Letters 24 (12) (2003) 1845–1855.

[26] A. P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern recognition 30 (7) (1997) 1145–1159.

[27] A. Lemos, W. Caminhas, F. Gomide, Adaptive fault detection and diagnosis using an evolving fuzzy classifier, Information Sciences 220 (2013) 64–85.

[28] I. Monroy, R. Benitez, G. Escudero, M. Graells, A semi-supervised approach to fault diagnosis for chemical processes, Computers & Chemical Engineering 34 (5) (2010) 631–642.

[29] M. Li, Z.-H. Zhou, Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 37 (6) (2007) 1088–1098.

[30] S. Li, J. Wen, Development and Validation of a Dynamic Air Handling Unit Model - Part I, ASHRAE Transactions 116 (1) (2010) 45–56.

[31] S. Li, J. Wen, X. Zhou, C. J. Klaassen, Development and validation of a dynamic air fandling unit model - part II, ASHRAE Transactions 116 (1) (2010) 57–73.

[32] J. Wen, S. Li, Tools for Evaluating Fault Detection and Diagnostic Methods for Air-Handling Units, ASHRAE 1312-RP (2011) 1–173.

[33] W. S. Noble, What is a support vector machine?, Nature biotechnology 24 (12) (2006) 1565.

[34] K. Yan, L. Ma, Y. Dai, W. Shen, Z. Ji, D. Xie, Cost-sensitive and sequential feature selection for chiller fault detection and diagnosis, International Journal of Refrigeration 86 (2018) 401–409.

[35] M. Tan, J. Pu, B. Zheng, Optimization of breast mass classification using sequential forward floating selection (SFFS) and a support vector machine (SVM) model, International journal of computer assisted radiology and surgery 9 (6) (2014) 1005–1020.

[36] A. Ben-Hur, J. Weston, A users guide to support vector machines, in: Data mining techniques for the life sciences, Springer, 223–239, 2010.

[37] J. Milgram, M. Cheriet, R. Sabourin, One against one or one against all: Which one is better for handwriting recognition with SVMs?, in: tenth international workshop on Frontiers in handwriting recognition, SuviSoft, 2006.

[38] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Information Processing & Management 45 (4) (2009) 427–437.

[39] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1-3) (2006) 489–501.

[40] L. Rutkowski, M. Jaworski, L. Pietruczuk, P. Duda, The CART decision tree for mining data streams, Information Sciences 266 (2014) 1–15.

[41] L. E. Peterson, K-nearest neighbor, Scholarpedia 4 (2) (2009) 1883.

[42] A. Liaw, M. Wiener, et al., Classification and regression by randomForest, R news 2 (3) (2002) 18–22.

Table 1: Top eight important feature variables for AHU FDD in summer season.

Index	**Variable**	**Description**
1	T_{sa}	Supply air temperature
2	T_{ra}	Return air temperature
3	Q_{ccoil}	Cooling coil load
4	T_{oa}	Outside air temperature
5	H_{ra}	Mixed air temperature
6	H_{sa}	Supply air humidity
7	T_{ma}	Return air humidity
8	T_{chwc}	Chilled Water Coil Discharge Air Temperature

Table 2: Top eight important feature variables for AHU FDD in winter season.

Index	**Variable**	**Description**
1	T_{sa}	Supply air temperature
2	T_{ra}	Return air temperature
3	Q_{hcoil}	Heating coil load
4	T_{oa}	Outside air temperature
5	H_{ra}	Mixed air temperature
6	H_{sa}	Supply air humidity
7	T_{ma}	Return air humidity
8	T_{hwc}	Heating Water Coil Discharge Air Temperature

Table 3: The FDD accuracy rates (%) for various semi-supervised learning methods based on different numbers of initial faulty samples in the summer season.

# Init. fault.	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**
Semi-SVM	36.82	57.67	68.07	73.58	78.27	80.99	81.96	85.64	85.70	88.18	88.96
Semi-KNN	27.5	34.26	41.84	50.13	53.54	59.32	65.27	68.62	71.14	74.15	76.43
Semi-CART	30.88	44.38	52.91	60.09	63.37	66.98	69.73	71.48	73.11	73.99	76.29
Semi-RF	20.18	31.1	42.12	49.25	53.68	57.58	64.43	65.20	69.01	70.91	72.95
Semi-ELM	40.57	49.75	51.67	54.48	56.93	59.01	60.78	66.44	69.70	71.20	73.94

Table 4: The FDD accuracy rates (%) for various semi-supervised learning methods based on different numbers of initial faulty samples in the winter season.

# Init. fault.	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**
Semi-SVM	38.17	59.80	70.80	79.03	79.65	88.72	90.08	91.53	92.01	92.28	92.53
Semi-KNN	8.89	25.57	38.72	52.85	56.27	64.98	69.14	73.62	74.83	77.19	79.34
Semi-CART	16.58	28.26	39.56	48.32	56.40	61.11	65.64	67.90	69.91	72.23	73.76
Semi-RF	12.93	20.14	37.41	44.26	54.47	62.2	66.08	69.92	76.06	77.41	79.98
Semi-ELM	30.32	45.67	59.00	69.90	74.10	75.83	79.66	81.20	84.30	85.93	87.29

Table 5: The number of iterations (computational complexity) for various semi-supervised learning methods based on different numbers of initial faulty samples in the summer season.

# Init. fault.	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**
Semi-SVM	12.3	12.7	19.0	16.8	22.7	27.1	28.6	25.4	23.0	26.0	21.4
Semi-KNN	6.7	6.8	7.1	7.3	7.5	8.7	10.3	8.8	10	11.2	11.6
Semi-CART	4.4	4.8	5.1	5.3	5.0	5.2	5.4	6.0	5.0	5.7	7.3
Semi-RF	7.4	9.5	16.0	15.8	19.7	33.8	41.5	40.4	49.6	61.1	68.1
Semi-ELM	16.3	13.7	20.1	38.4	38.2	45.6	52.2	56.8	58.3	65.1	80.2

Table 6: The number of iterations (computational complexity) for various semi-supervised learning methods based on different numbers of initial faulty samples in the winter season.

# Init. fault.	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**
Semi-SVM	2.8	6.2	6.1	6.7	12.4	10.9	11.1	12.9	13.0	11.4	9.1
Semi-KNN	2.0	8.5	7.1	8.0	6.3	8.4	7.3	7.2	7.2	8.1	7.2
Semi-CART	2.3	3.0	4.1	3.8	3.6	4.5	4.2	4.1	4.0	3.9	4.3
Semi-RF	5.9	21.8	27.5	31.2	30.9	32.4	55.3	76.4	82.9	80.1	67.2
Semi-ELM	37.9	38.7	40.0	35.3	37.9	40.7	38.2	39.2	35.5	46	42.7

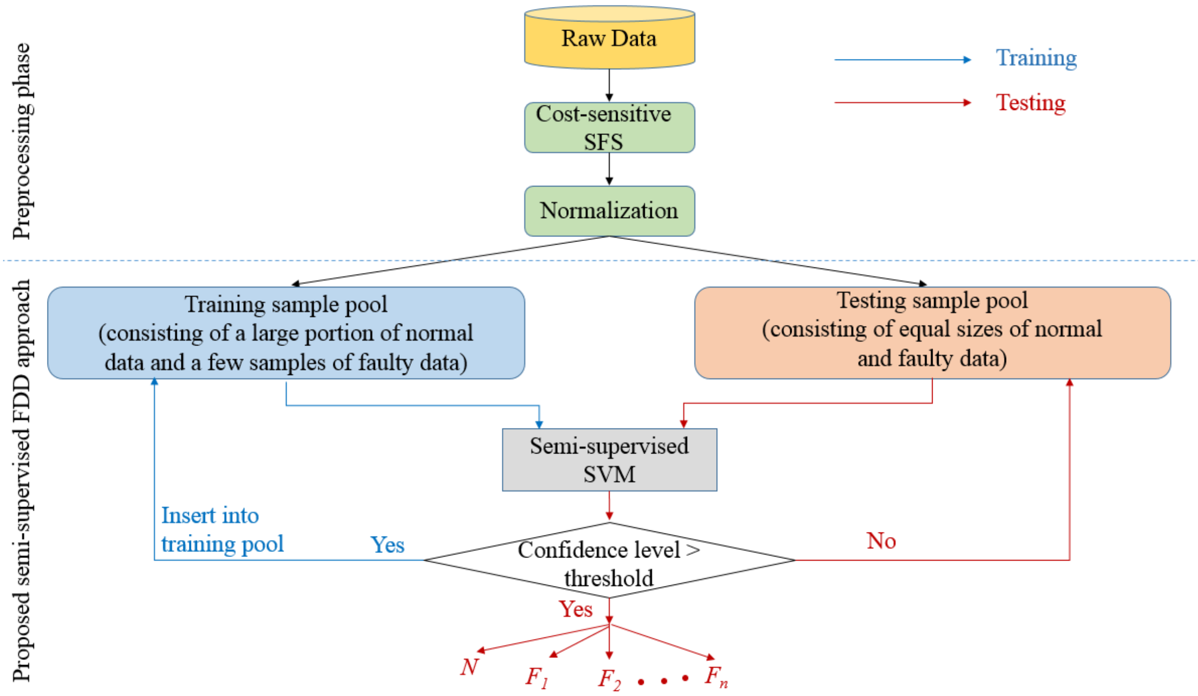


Figure 1: The flowchart of the proposed semi-supervised FDD approach for AHU faults.

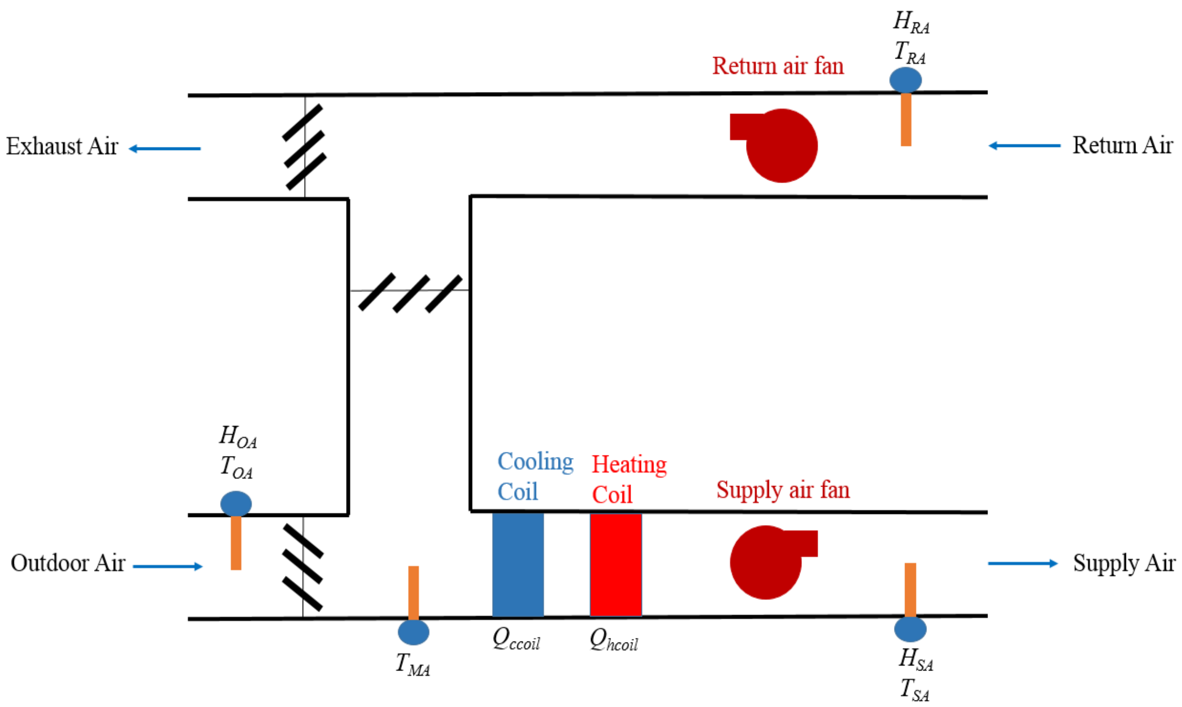


Figure 2: The basic structure of an AHU.

Prediction Results

	Positive	Negative
True	TruePositive	TrueNegative
False	FalsePositive	FalseNegative

Prediction Results

	p1	p2	p3	p4	p5	p6	p7
C1	C1p1	C1p2	C1p3	C1p4	C1p5	C1p6	C1p7
C2	C2p1	C2p2	C2p3	C2p4	C2p5	C2p6	C2p7
C3	C3p1	C3p2	C3p3	C3p4	C3p5	C3p6	C3p7
C4	C4p1	C4p2	C4p3	C4p4	C4p5	C4p6	C4p7
C5	C5p1	C5p2	C5p3	C5p4	C5p5	C5p6	C5p7
C6	C6p1	C6p2	C6p3	C6p4	C6p5	C6p6	C6p7
C7	C7p1	C7p2	C7p3	C7p4	C7p5	C7p6	C7p7

Figure 3: Confusion matrix expansion from binary classification to a 7-class classification. The red slots represent all falsely classified testing samples. For each testing sample, the chance of incorrect labelling for multi-class SVM is much higher than binary SVM.

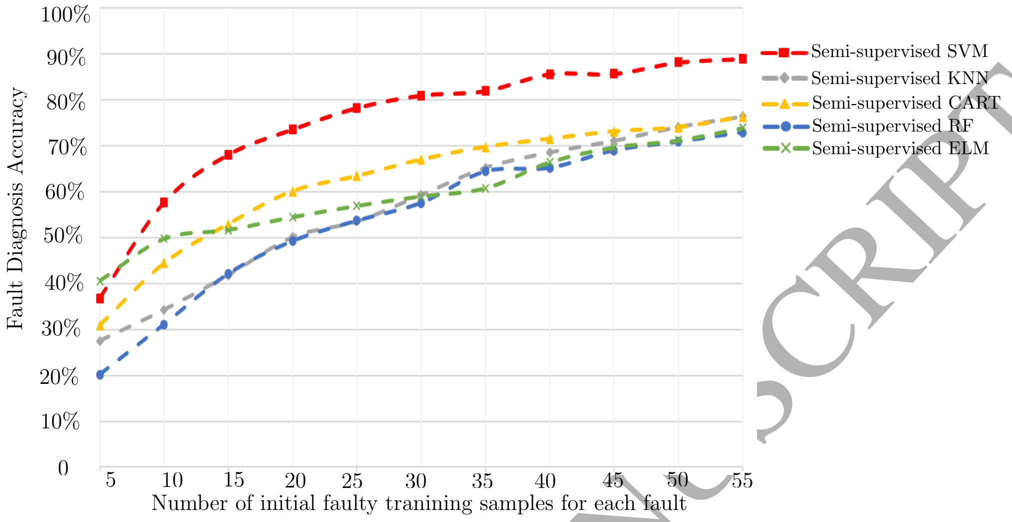


Figure 4: Experimental results comparison between various semi-supervised machine learning methods in summer season.

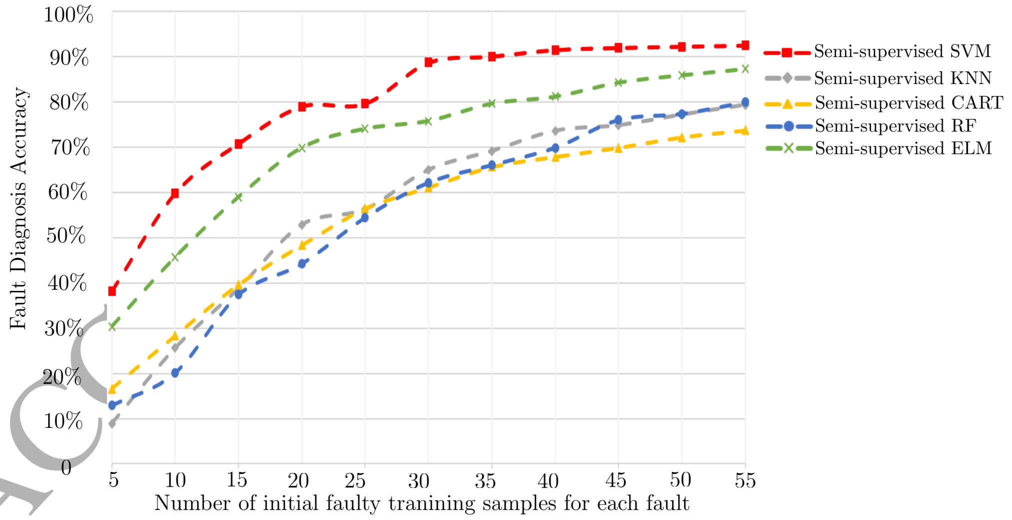


Figure 5: Experimental results comparison between various semi-supervised machine learning methods in winter season.

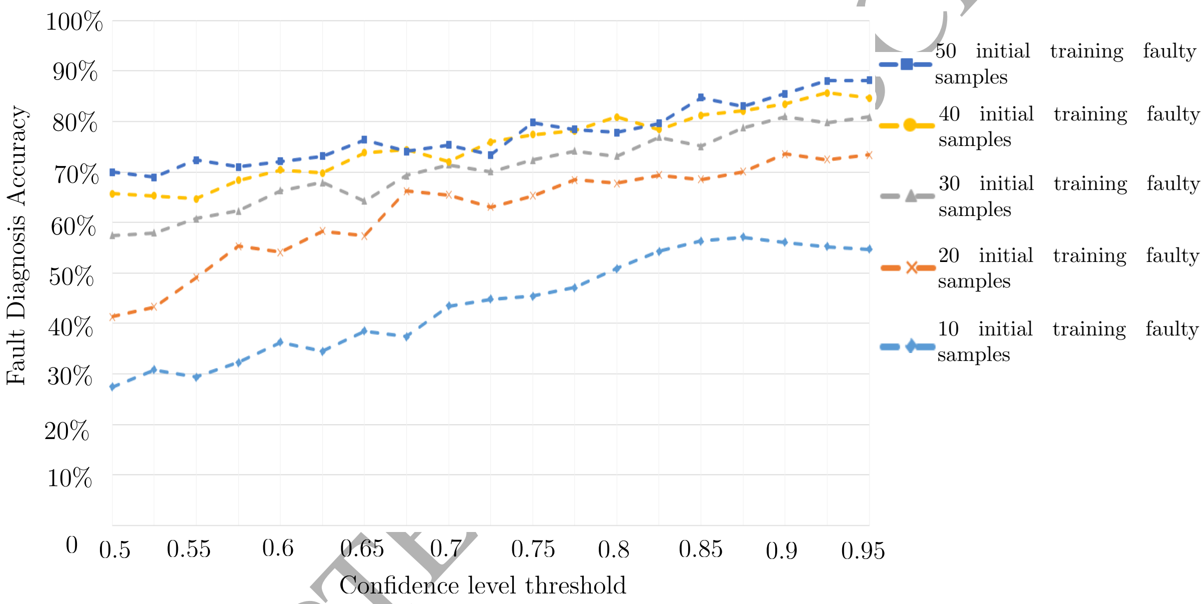


Figure 6: Experimental results showing the averaged diagnostic accuracy rates versus various confidence level thresholds.