

# Fault Detection and Diagnosis in Air Handling Units with a Novel Integrated Decision Tree Algorithm

Vivekanadam Balasubramaniam,

Faculty of Computer Science and Multimedia,

Lincoln University College,

Kota Bharu, Malaysia.

[vivekanadam@lincoln.edu.my](mailto:vivekanadam@lincoln.edu.my)

**Abstract:** In air handling units (AHUs), wide attention has been attracted by data-driven fault detection and diagnosis techniques as the need for high-level expert knowledge of the concerned system is eliminated. In AHUs, decision tree induction is performed by means of classification and regression tree algorithm which is a data-driven diagnostic strategy based on decision tree. Expert knowledge as well as testing data may be used for validation of fault diagnosis reliability with easy interpretation and understanding ability offered by the decision tree. The diagnostic strategy established and its interpretability are increased by incorporating a regression model and steady-state detector with the model. ASHRAE, Oak Ridge National Lab (ORNL), National Renewable Energy Lab (NREL), Pacific Northwest National Lab (PNNL) and Lawrence Berkeley National Lab (LBNL) datasets are used for validation of the proposed strategy. High average F-measure and improved diagnostic performance may be achieved with this strategy. There is a compliance between the expert knowledge and certain diagnostic rules generated in the decision tree as seen from the expert knowledge implemented diagnostic decision tree interpretation. Based on the interpretation, it is evident that certain diagnostic rules are valid only under specific operating conditions and some of the generated diagnostic rules are not reliable. Data driven models are used for emphasizing the significance of interpretability of fault diagnostic models.

**Keywords:** Interpretation, Fault Diagnosis, Air Handling Unit, Decision Tree, Feature selection, Fault Detection

## 1. Introduction

Indoor thermal discomfort and increased energy consumption are some of the issues caused by faults that many occur in Heating, Ventilation, and Air-Conditioning (HVAC) systems used in buildings due to their vulnerability during operation [1]. Reduction and carbon footprint and improving the energy efficiency of the buildings are the benefits of successful and timely fault detection and isolation in HVAC systems. In HVAC systems for building, several advancements in fault detection and diagnosis (FDD) has been made and efficient strategies have been developed over the past few decades. Fault identification is performed by fault diagnosis while healthy operating conditions are determined in a more than that of fault detection [2]. Data-driven, rule-based and model-based are the three major categories of FDD models. Due to the manageable computational cost and elimination of the need for high-level expert knowledge of the system, wide attention is gained by the data driven schemes [3]. In HVAC systems, the air handling unit (AHU) is a significant component. Several literatures are available on the AHU fault detection and diagnosis schemes. Expert knowledge is used for formulation of a rule based FDD strategy using a set of if-then rules for AHU temperature sensors [4]. Under various operating modes, the sensor faults at AHU are isolated in an efficient manner according to the performance evaluation results from a small scale AHU simulator and green building data.

In order to perform efficient fault diagnosis in AHUs, a data-driven decision tree based scheme is proposed in this paper. A satisfactory accuracy is obtained in prediction of energy utilization in buildings on implementation of the decision tree classifier. A set of if-then rules are generated by the decision tree for interpretation by the proposed strategy. When compared to the existing data-driven diagnostic strategies, this technique offers improved performance. Fault isolation of various AHUs may be performed by understanding the diagnostic strategy with the interpretability feature of the proposed model. In data-driven FDD techniques as in the proposed model, defining and analysis of key features need to be performed for automatic feature selection.

## 2. Related Works

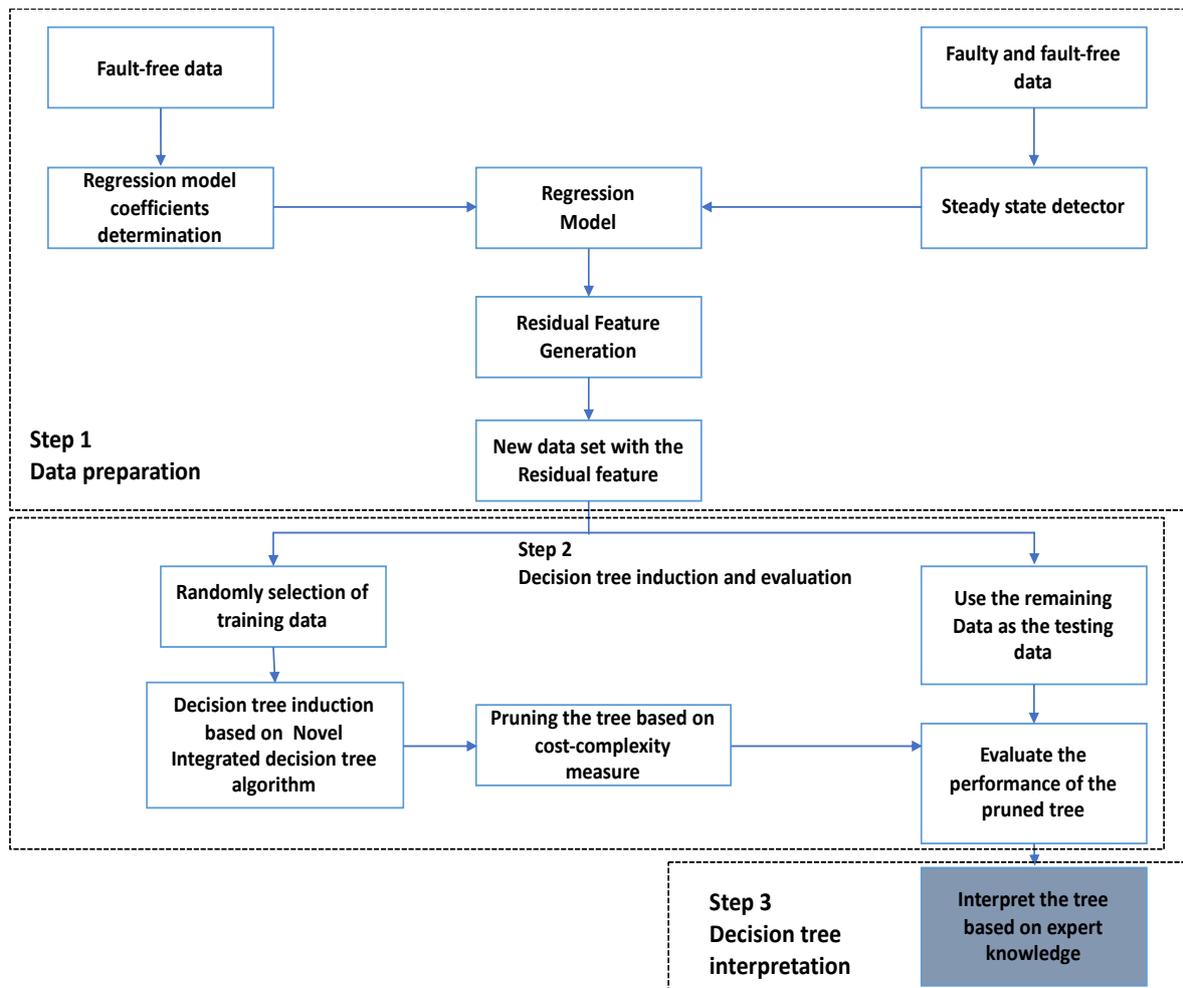
Initially, pattern classification schemes were used for developing data-driven fault diagnosis techniques [5]. Machine learning models and static based models are the common categories of data driven FDD schemes for AHUs [6]. Common statistic based techniques such as Fisher Discriminant Analysis (FDA) and Principal Component Analysis (PCA) are used. A new set of uncorrelated variables are used for transferring the dataset with the help of a multivariate analysis technique termed as PCA. Most variations are retained by the first few principal components in this model [7]. A Q-contribution plot is used for fault diagnosis by comparing the PCA model with a pre-defined threshold based prediction error in AHU sensors [8]. Data is presented in lower dimensions after extraction of critical information from high dimension data with the help of PCA. For maximization of separation among various classes, the linear dimensionality reduction scheme based FDA classifiers are used [9].

AHU sensor faults are detected using PCA and diagnosed using FDA. Fault detection using PCA is the first step in this fault detection and isolation approach for AHU sensor [10]. At the second stage, based on active functional testing, fault isolation is performed. In case of AHU FDDs, the application of artificial neural networks (ANNs) and other machine-learning based schemes are studied extensively. An AHU FDD approach based on ANN has been developed by certain researchers [11]. Faulty as well as normal data are used for training a back-propagation based two-layer feed-forward ANN in which the measured and predicted values are used for selection of certain user-defined residual variables. Mechanical faults and specific number of typical AHU sensors are considered [12]. The root cause for AHU faults may be identified by the FDD techniques based on ANN. Wavelet analysis in combination with ANN is used for AHU FDD. Fault diagnosis and AHU fault detection is performed by coupling wavelet analysis with Elman neural network using Back-propagation neural network (BPNN).

AHU FDD is also performed using machine learning techniques like Support vector machine (SVM). SVM based diagnostic strategy for AHU is proposed by certain researchers [13]. Different fault classes may be isolated using this strategy by identifying the boundaries that enable maximization of margins between the classes. The SVM classifier is fed with input

variables that are the residuals between actual and predicted values [14]. Four physical models are articulated for this purpose. Selection of the SVM kernel function decides the diagnostic accuracy of the SVM based schemes. AHU FDD is performed with an autoregressive model with exogenous variables (ARX) and SVM classification which is a model-based FDD strategy [15]. Feature selection based on relief F is used by the ARX model and the exogeneous variables are selected from the mixed and supply air temperature while the dependent variable is obtained from the supply air humidity. Naïve Bayes, random forest and ANN classifiers are compared with the SVM based scheme for comparison and evaluation of performance [16]. In terms of diagnostic accuracy, the other techniques were outperformed by the SVM-based scheme.

### 3. Proposed Work



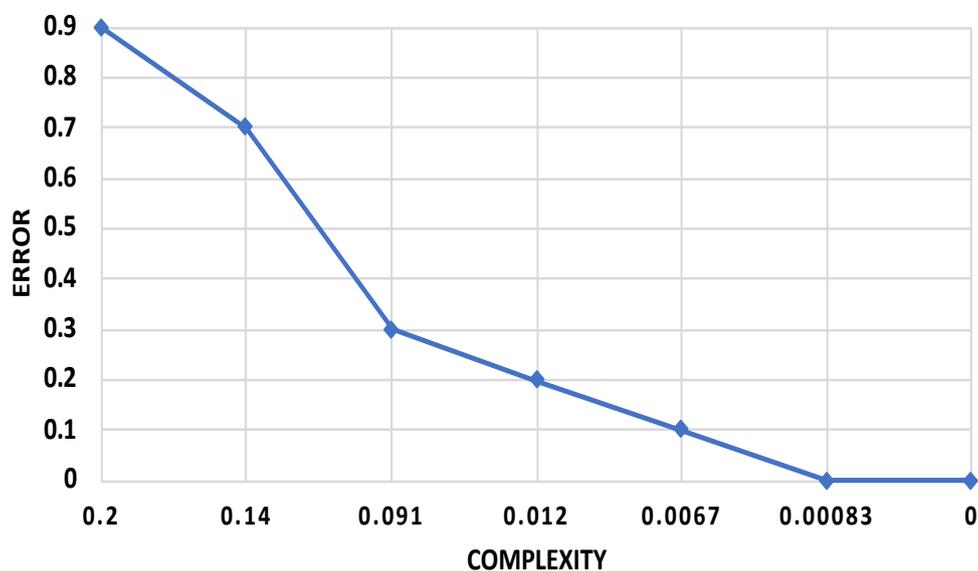
**Fig. 1.** Proposed AHU based diagnostic framework

Figure 1 provides the illustration of the proposed AHU based diagnostic framework. The major levels of the strategy includes data formulation, induction and evaluation of decision tree and interpretation of the decision tree. The regression model coefficients are determined with the help of fault-free data at the initial level. The developed decision tree complexity is reduced by generating a residual feature with the help of the regression model. Further, the complexity of the decision tree is reduced and the fault diagnosis accuracy is improved by filtering out data during the transient period using a steady state detector while processing the data labelled with various faults in AHU as well as the fault-free data recorded. It is possible to generate a residual between the predicted and measured fan speed control signal as well as a new dataset with the residual feature with the help of the filtered data and the regression model. Fault diagnosis may be performed by preparing the filtered data.

The core diagnostic strategy is the next level where induction and evaluation of the decision tree is performed. The data is initially divided into three portions and two of them are randomly chosen for training the decision tree based on the classification and regression tree (CART) algorithm. All possible diagnostic rules are captured by making the decision tree fully grown in the initial level. Over-fitting results may be obtained and difficulties in model interpretation may appear due to the large number of branches in a fully grown tree. Utilization of appropriate schemes for pruning the tree is essential. For the purpose of tree pruning, the cost-complexity measure is used in CART. F-measure is used for developing the diagnostic decision tree and its accuracy is tested and validated using the new test data set that is prepared after pruning the decision tree. A prediction model is fit into each recursive partition of the dataspace in the decision tree during decision tree induction schemes with the help of CART. Binary decision tree represents the partitions graphically. Regression as well as classification may be performed using CART unlike Chi-square Automatic Interaction Detector (CHAID) and other tree induction techniques that can only perform classification. A diagnostic model may be constructed by using a classification tree induction scheme like CART.

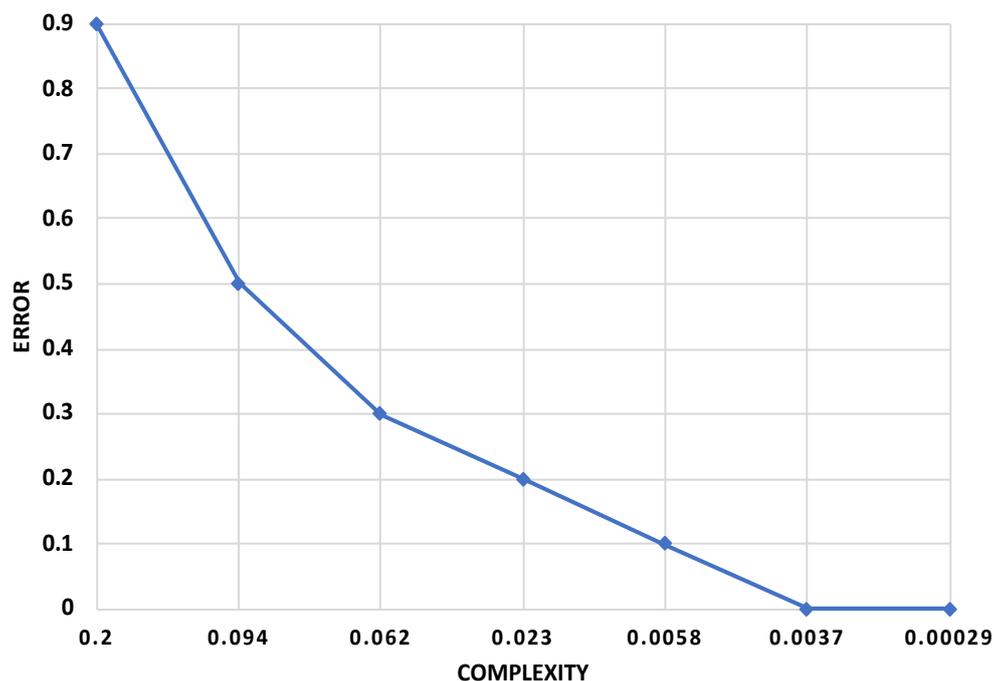
#### 4. Results and Discussion

ASHRAE, Oak Ridge National Lab (ORNL), National Renewable Energy Lab (NREL), Pacific Northwest National Lab (PNNL) and Lawrence Berkeley National Lab (LBNL) datasets are used for testing and validation of the proposed AHU fault diagnosis strategy. An experimental AHU facility is used for testing various AHU faults. The fault-free test is conducted over a period of five days while the faults are tested for a duration of 6 hours with a one minute sampling rate. This study considers a wide variety of AHU faults such as duct leakage, stuck air damper, return fan failure, stuck cooling coil valve and leakage in heating coil valve. These are some of the typical mechanical faults found in AHU FDD literature. During summer, the AHU operation under mechanical cooling mode is used for testing and validation of the diagnostic strategy with the help of the collected data. 22°C of indoor temperature and 13°C of supply air temperature are maintained during the mechanical cooling mode. The operating speed of the supply fan and the supply air static pressure is maintained at a fixed ratio for controlling the air supply.

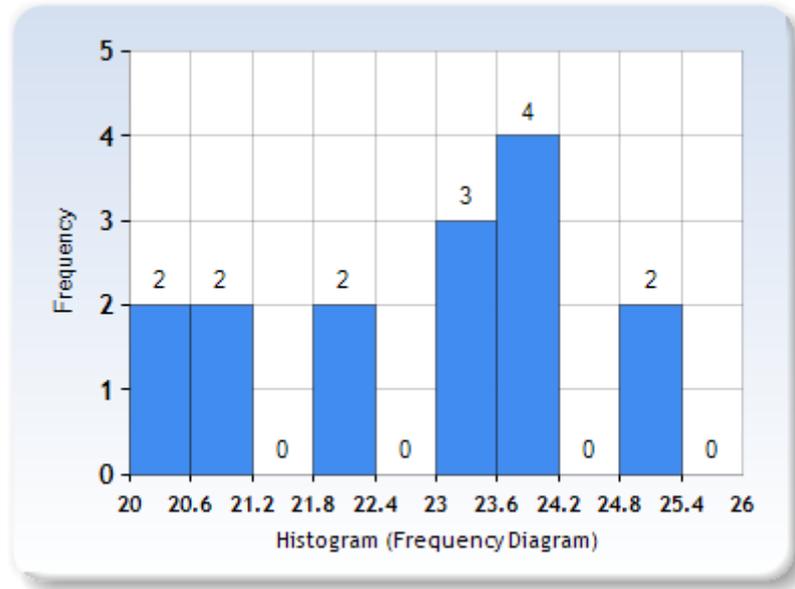


**Fig. 2.** Complexity parameters with respect to the relative error and total splits considering the residual feature

Figure 2 provides the comparison of complexity parameters with respect to the relative error and total splits considering the residual feature. Leaf node with minimum observations are generated in the fully grown decision tree based on the training data that is randomly selected. Comparison of the observations from the training data in terms of fault free and faulty conditions is performed. Visualization is made easier and generation of large decision tree is avoided by constraining the total observations in a leaf node. Around 26 splits may be available in total in a fully grown decision tree. The classification results are used for labelling the leaf nodes while the test conditions are used for labelling the internal nodes and root for splitting the observations with various characteristics. The right branches did not meet the node splitting criteria while the internal nodes and root in the left branches could meet them. Out of the total observations classified in a leaf node, the ones that are correctly classified represents the number of leaf node. Figure 3 provides the comparison of the complexity parameters with respect to the relative error and total splits without considering the residual feature.



**Fig. 3.** Complexity parameters with respect to the relative error and total splits without considering the residual feature



**Fig. 4.** Histogram analysis of air temperature and observations

On proper isolation of the fan based faults, the corresponding splitting feature is selected. Figure 4 represents the fault-free and faulty condition based density distribution of the fan power return. The return fan completely failed fault is overcome by as observed in the density distribution due to the 0 return fan power under this fault condition. It is possible to understand the isolation of various faults by means of the fault diagnostic strategy interpretability. During actual operation, identification of target fault may be challenging if only randomly selected test data is used for validation of the fault.

#### 4. Conclusion

Classification and Regression Tree (CART) based fault diagnostic scheme for data driven AHU is presented in this paper. This model overcomes the need for defining and analyzing the key features while performing feature selection automatically. Interpretable fault diagnostic results may be obtained from the proposed strategy when compared to the data-driven fault diagnostic schemes based on black-box model which does not have precise physical significance. The expert knowledge is matched well with the CART algorithm based fault diagnostic rules that are generated based on the interpretation of the diagnostic tree developed in this paper. Various

AHU faults may also be isolated successfully by means of the proposed fault diagnostic technique using CART algorithm. Future work is directed towards development of more efficient fault diagnostic strategies by combining expert knowledge and data-driven techniques.

## References

- [1] Fan, C., Liu, X., Xue, P., & Wang, J. (2021). Statistical characterization of semi-supervised neural networks for fault detection and diagnosis of air handling units. *Energy and Buildings*, 234, 110733.
- [2] Yun, W. S., Hong, W. H., & Seo, H. (2021). A data-driven fault detection and diagnosis scheme for air handling units in building HVAC systems considering undefined states. *Journal of Building Engineering*, 35, 102111.
- [3] Piscitelli, M. S., Mazzarelli, D. M., & Capozzoli, A. (2020). Enhancing operational performance of AHUs through an advanced fault detection and diagnosis process based on temporal association and decision rules. *Energy and Buildings*, 226, 110369.
- [4] Yan, K., Chong, A., & Mo, Y. (2020). Generative adversarial network for fault detection diagnosis of chillers. *Building and Environment*, 172, 106698.
- [5] Zhao, Y., Zhang, C., Zhang, Y., Wang, Z., & Li, J. (2020). A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis. *Energy and Built Environment*, 1(2), 149-164.
- [6] Fan, C., Liu, Y., Liu, X., Sun, Y., & Wang, J. (2021). A study on semi-supervised learning in enhancing performance of AHU unseen fault detection with limited labeled data. *Sustainable Cities and Society*, 70, 102874.
- [7] Zhao, Y., Li, T., Zhang, X., & Zhang, C. (2019). Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. *Renewable and Sustainable Energy Reviews*, 109, 85-101.
- [8] Zhong, C., Yan, K., Dai, Y., Jin, N., & Lou, B. (2019). Energy efficiency solutions for buildings: Automated fault diagnosis of air handling units using generative adversarial networks. *Energies*, 12(3), 527.

- [9] Zhang, S., Zhu, X., Anduv, B., Jin, X., & Du, Z. (2021). Fault detection and diagnosis for the screw chillers using multi-region XGBoost model. *Science and Technology for the Built Environment*, 1-24.
- [10] Gao, T. (2020). Integrated building fault detection and diagnosis using data modeling and Bayesian networks (Doctoral dissertation, Ecole nationale supérieure Mines-Télécom Lille Douai).
- [11] Karami, M., & Wang, L. (2018). Fault detection and diagnosis for nonlinear systems: A new adaptive Gaussian mixture modeling approach. *Energy and Buildings*, 166, 477-488.
- [12] 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.
- [13] Yan, K., Ji, Z., Lu, H., Huang, J., Shen, W., & Xue, Y. (2017). 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*, 49(7), 1349-1356.
- [14] Shakya, S. (2020). Performance analysis of wind turbine monitoring mechanism using integrated classification and optimization techniques. *Journal of Artificial Intelligence*, 2(01), 31-41.
- [15] Joseph, S. I. T., & Thanakumar, I. (2019). Survey of data mining algorithm's for intelligent computing system. *Journal of trends in Computer Science and Smart technology (TCSST)*, 1(01), 14-24.
- [16] Kumar, T. S. (2020). Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning Algorithm. *Journal of Artificial Intelligence*, 2(03), 185-193.