Lawrence Berkeley National Laboratory

Design of Machine Learning Models with Domain Experts for Automated Sensor Selection for Energy Fault Detection

R.L. Hu[1]
J. Granderson[3]
D. M Auslander[2]
A. Agogino[2]

[1]Salesforce Research
[2]University of California Berkeley
[3]Lawrence Berkeley National Laboratory

Energy Technologies Area
November, 2018

Please cite as:
Hu, R., Granderson, J., Auslander, D. and Agogino, A. (2019). Design of machine learning models with domain experts for automated sensor selection for energy fault detection. *Applied Energy*, 235, pp.117-128.

Disclaimer:

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Design of Machine Learning Models with Domain Experts for Automated Sensor Selection for Energy Fault Detection

R.L. Hu[a,1,*], J. Granderson[c], D.M. Auslander[b], A. Agogino[b]

[a]*Salesforce Research, Palo Alto, California, 94301*
[b]*Department of Mechanical Engineering, University of California, Berkeley, Berkeley, California 94720*
[c]*Building Technology and Urban Systems, Lawrence Berkeley National Laboratory, Berkeley, California 94720*

Abstract

Data-driven techniques that extract insights from sensor data reduce the cost of improving system energy performance through fault detection and system health monitoring. To lower cost barriers to widespread deployment, a methodology is proposed that takes advantage of existing sensor data, encodes expert knowledge about the application system to create 'virtual sensors', and applies statistical and mathematical methods to reduce the time required for manual configurations. The approach combines sensor data points with encoded expert knowledge that is generic to the application system but independent of a particular deployment, thereby reducing the need to tailor to individual deployments. This paper not only presents a method that detects faults from measured energy data, but also 1) describes an engagement method with experts in the energy system domain to identify data, 2) integrates domain knowledge with the data, 3) automatically selects from among the large pool of potential input data, and 4) uses machine learning to automatically build a data-driven fault detection model. Demonstration on a commercial building chiller plant shows that only a small number of virtual sensors is necessary for fault detection with high accuracy rates. This

*Corresponding author
Email address: lhu@salesforce.com (R.L. Hu)

[1]Work done while at University of California, Berkeley and at Lawrence Berkeley National Laboratory

corresponds to the use of only five out of 52 original sensor data points features. With as few as four features, classification F1 scores exceed 90% on the training set and 80% on the testing set. The results are implementable and realizable using off-the-shelf tools. The goal is to design with domain experts an energy monitoring system that can be configured once and then widely deployed with little additional cost or effort.

Keywords: machine learning, time series, fault detection, anomaly detection, energy savings, energy efficiency

---

1. Introduction

Renewable energy technologies as well as building energy management systems have upwards of hundreds of existing sensor data points used for control and monitoring. Furthermore, innovations in "Internet of Things" (IoT) devices have led to connected power meters, lights, occupancy sensors, and appliances that are capable of data collection and communication. This data presents a valuable opportunity to extract meaningful information and take data-driven action.

The motivation for transforming data from these devices into actionable information is to improve operations, monitor system health, increase energy generation, and decrease energy waste. The development and widespread use of energy conservation and renewable energy technologies are critical to minimizing negative environmental consequences. To that end, increasing profitability for users and decreasing costs of these technologies enables market penetration and widespread adoption. On the energy demand side, commercial buildings consume 19% of US primary energy [1]. Of this, an estimated 15% to 30% of energy used in commercial buildings is wasted by poorly maintained, degraded, and improperly controlled equipment [2].

However, one cannot achieve scalable deployments of analytics and applications across systems if deploying solutions requires vendors and domain experts to install sensors and information technology infrastructures that require tailoring each solution for each deployment. Today, even well-established commercial offerings are not deployed at scale because costs are prohibitive. Thus, a major challenge to scalability is reducing hardware and software installation costs, manual configuration requirements, and manual monitoring.

To address this challenge, this paper describes a novel framework for fault detection that combines existing data from control systems with expert knowledge that is general to the type of energy system to create new virtual sensors. These are 'virtual' in the sense that they do not correspond to values taken directly off a physical sensor but are calculated from values taken from physical sensor(s). The use of expert knowledge that is generic to the energy system application but independent of individual instances reduces the need for configuration to individual deployments. Techniques from statistics and machine learning automate the process of selecting inputs to the fault detection system. The input data is converted into a simple form suitable for off-the-shelf machine learning algorithms.

This work's novelty is the process to obtain relevant domain knowledge, create and select virtual sensors, and build a data-driven fault detection system. This method is not tied to a particular type of energy system. Furthermore, the designer of this fault detection system does not require knowledge about the energy system. Deployment at a specific instance of an energy system also does not require detailed knowledge about the specification and configuration of a specific instance.

Energy systems are complex with many potential data points that can be tracked. These include signals, inputs, responses, and states related to thermodynamic, meterological, control, mechanical, and electrical properties of the system. We present a method for selection of the most useful sensors for fault detection. The desirability of fault detection and the problem of a large quantity of potential sensors to select from are common across energy systems, be it HVAC, energy generators, distribution systems, or energy end uses. Our data driven methodology can be extended to other energy systems. People who are not experts in the energy system nor a specific instance can follow our proposed scheme to create a fault detection system.

Energy applications for machine learning include assessing solar and wind energy. Solar irradiance was predicted by applying logistic regression to important meteorological variables found through analysis with Boosted Regression Trees [3]. To estimate wind turbine power output, five models were compared based on measure-correlate-predict methods using artificial neural networks, Support Vector Machines for regression and random forests [4].

For energy consumption in buildings, various machine learning models were regressed on five commonly accessible building and climate features to estimate annual commericial building energy consumption across the US [5]. Deep learning was applied to forecast 24 hr cooling loads in both a su-

pervised manner and an unsupervised manner to extract features as input for cooling load prediction [6]. Real time building energy consumption data was predicted by combining time-series sliding window analysis with meta-heuristic optimization-based machine learning [7]. Machine learning has also been used to create building control strategies by constructing approximate model predictive control laws by using multivariate regression and dimensionality reduction algorithms [8]. By learning occupant behaviour through supervised and unsupervised learning, rules were created to infer real-time room setpoints for office cooling [9]. For fault detection, machine learning frameworks have been proposed to identify and classify islanding and grid disturbance [10] and to detect faults and in wind turbines [11]. This paper generalizes and adapts [11] for fault detection in buildings.

In fault detection literature, a paper in a three part review series of fault detection and diagnosistics for processes [12] describes a common set of criteria with which to compare and evaluate fault detection and diagnostic methods and provides and describes a taxonomy for the methods. There also exist review papers that survey specific categories of fault detection, such as supervisory methods, model based techniques, and trends in applications of model based techniques [13]. Another review paper surveys fault diagnostics with multivariate statistical models [14]. A further paper comments on the era of "big data" in the context of analyzing process history data [15]. An overview of analytical methods and features of methods used in commercial fault detection and diagnostics offerings for buildings is provided in [16]. It documents that rule based approaches are still common, but there is growing use of process history based approaches.

Fault detection approaches using machine learning has also been applied to HVAC systems. For fault detection in buildings, machine learning has been applied to fault diagnostics in air handling units modeled as a Bayesian probablistic model [17] and for fault detection in chiller plants modeled as a gaussian mixture model [18]. By using physical models, state estimation techniques have been applied to identify faults by comparing the deviation of actual measurements from their values as predicted by the physical model [19]. HVAC sub-systems have also been modelled as agents in a directed graphical model that is trained on HVAC data under normal running conditions [20]. Neural networks have been employed to predict physical properties and calculated weighting factors for the neural networks to get thresholds above which faults are detected with subtractive clustering analysis used for fault diagnosis [21]. In terms of black box modeling, three different data

mining techniques have been compared for outlier detection to detect faults: density-based spatial clustering of applications with noise, K-means, and classification and regression tree [22]. Grey box methods include use of Statistical Process Control (SPC) to measuring and analyze variations, Kalman filtering to provide predictions and determine SPC control limits, and system analysis for fault propagation across subsystems [23]. Rule based techniques include using expert rules to create a cloud-based expert-rule based fault detection system [24].

In contrast to these studies, this paper not only presents a method that detects faults from measured energy data, but also 1) describes an engagement method with experts in the energy system domain to identify data, 2) integrates domain knowledge with the data, 3) automatically selects from among the large pool of potential input data, and 4) uses machine learning to automatically build a data-driven fault detection model. This scheme is not specific to HVAC systems but has also been demonstrated on wind turbines [11]. Most previous works rely on experts to directly select data input and design the fault detection system, for example, by specifying rules or deciding how to model the system being monitored. In contrast, we present a methodology that automates the data selection and fault detection design.

One challenge with data-intensive energy monitoring systems is the implementation and processing costs associated with the large number of sensors required. This paper describes a framework that combines expert knowledge about an application with data readily available from an existing control system. Automated tools are used to identify data points relevant for health monitoring, and then machine learning to identify the most effective model parameters for configuring the system design. This procedure is demonstrated on developing a health monitoring system for fault detection in a commercial building HVAC system.

The additional expert knowledge features are based on 1) an understanding of the measured physical values and performance metrics, 2) the time series behaviour of the sensor measurements, and 3) statistical features. Feature selection methods are then applied to this expanded feature set to select the most important overall features and to validate whether the new features selected are more useful for prediction than the original ones. In the health monitoring field and in application domains, variable and feature selection is often obtained from expert judgement. In the machine learning field, variable and feature selection commonly involves using mathematical and statistical techniques, as summarized by Guyon and Elisseeff [25] and Li et al [26],

which will be the process employed in this paper.

Next, machine learning models are trained to predict faults using both the expanded feature set and the original data set from the control systems to predict faults. The data from the control system is collected from the supervisory control and data acquisition (SCADA) system, the industrial computer system used for automation and control.

This paper is organised as follows: Section 2 describes the proposed methodology. Section 3 describes a case study to design a fault detection system for a commercial building chiller plant. Section 4 applies the methodology to the case study. Section 5 reports the experimental results. Section 6 summarises the conclusions.

2. Method

2.1. Overview

A data-driven scheme is presented for generalized fault detection using a combination of encoded expert knowledge and statistically derived features from sensor data. The proposed method, illustrated in Figure 1, uses machine learning and statistical techniques to optimally chose a subset of sensor data from existing data in the system. This sensor data is supplemented with expert engineering knowledge that is generic to the class of equipment. The method is applied to a chiller plant that provides centralized chilled water to cool large buildings and multi-building campuses.

Using the data from the control system, new features are created that incorporate knowledge about the application. The creation of new features derived from existing data variables in the control system reduces the cost of implementation by avoiding the purchase, installation, and setup of new sensors. New features are also included to improve fault detection performance.

To create application domain knowledge features, experts were interviewed to determine metrics that were useful to monitor the condition of equipment and energy performance.

2.2. Collection of Domain Knowledge

2.2.1. Selection of Experts

Individuals with hands on experience with maintenance, operations, and working with the existing sensor data were targeted for interviews. For chiller plants, these experts include building, chiller plant, maintenance, and facilities managers, operators, and specialists. Similar roles exist for other types

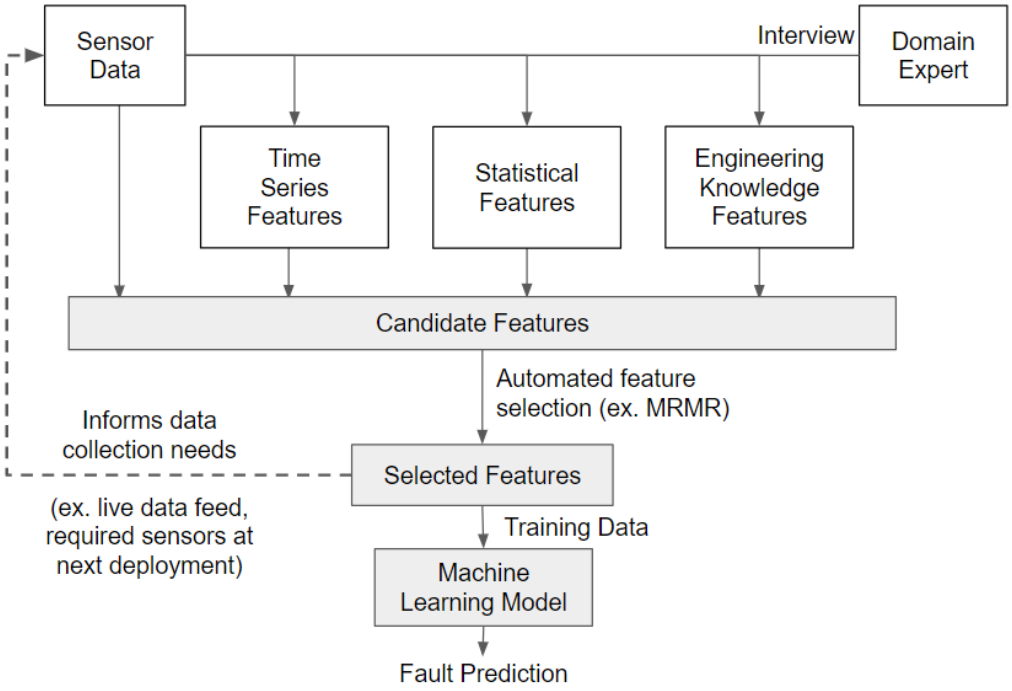

Figure 1: Schematic diagram illustrating the incorporation of domain knowledge from experts with sensor data for automated feature and sensor selection and a data driven machine learning model to detect faults

of energy systems. For wind turbines, these experts have titles such as performance specialists and wind turbine operators.

Experts also include scientists, researchers, technical leads at control companies, and operators. Other sources of domain knowledge include operations and design training manuals, previous work experience in the industry, and basic science and engineering knowledge. This reference material can be used to generate a list of proposed features that can then be presented to experts for comments.

2.2.2. Interview Questions

Once identified, experts were individually interviewed in sessions lasting between 30 minutes to one hour. During interviews, the following questions were asked. These questions were asked at the level of the whole system (i.e., the chiller plant), components (e.g., primary chiller water loop), and equipment (e.g., a chiller, a pump, a valve).

- What sensor values do you check? How?
- What metrics do you calculate and track?
- What are common and/or standard metrics?
- What sensors are associated with those metrics?
- How are these metrics calculated?
- What data do you wish was readily available?
- If you had a dashboard for your data, what would you like to see on it?
- What types of plots do you create? Can you please describe them? What are on the axes?
- What values do you plot against one another?
- What time scales do you look at?
- What time scales would be too noisy or transient?

Experts do not need to be interviewed for each deployment. Experts should be interviewed when there are differences in the type of equipment or components. For example, interviews should be conducted for fixed versus variable speed chillers, and for plants with secondary chilled water loops versus those with only primary chilled water loops. However, additional interviews are not necessary for differences in, for example, equipment size, model number, or differing numbers of cooling towers or chillers.

2.3. Application Domain Knowledge Features

The collected application domain knowledge supplements control system data as shown in Figure 1 to create candidate features. These candidate features can be thought of as virtual sensors because they do not correspond to values taken directly off a physical sensor but are calculated from values taken from physical sensor(s). The original data is combined with this new knowledge in a simple form that is appropriate as input into various machine learning algorithms. The new features created using application domain knowledge include: engineering knowledge features, time series features, and statistical features.

2.3.1. Engineering Knowledge Features

Engineering knowledge features are created using knowledge of the quantities that the original features correspond to, the location of where data for those features is collected, and an understanding of the operations of the application. For example, the cooling tower range temperature (the temperature drop across the cooling tower) is an important metric to track the performance of a cooling tower. Thus, a new feature is the difference between the temperature of the water approaching a cooling tower and the temperature of the water leaving a cooling tower.

2.3.2. Time Series Features

Since physical systems are the focus of these fault detection applications, it is known that the original features and derived features are physical quantities that form time series. Timestamped data samples are not independent from one another. It is desirable to include time related data. To that end, time series features are created from the data samples, the selection of which would be driven by the targeted application and faults.

In the example presented here, the data samples are converted to sliding time series representations of one hour using lagged features. To represent the data as time series, lagged features — also called delays — for each original feature are created to include the data from time $t - 60min$ to t. For example, if the original feature "Chiller 1 Power" is at time t and the sampling resolution is 5 minutes, one new feature is "Chiller 1 Power at time $t - 5min$", another new feature is "Chiller 1 Power at time $t - 10min$", etc. all the way to "Chiller 1 Power at time $t - (n * 5)min$".

The n, which is called the order of the lag, is selected using Akaike's Information Criterion (AIC) [27] and Bayesian Information Criterion (BIC) [28].

In this study, for each feature, orders recommended by both AIC and BIC are calculated and the max order between AIC and BIC is selected.

2.3.3. Statistical Features

Statistical features are created from the original features and can be selected based on the application. The first statistical feature is the sliding average. For example, the sliding average of "Outside Air Temp" at sample time t is the average of "Outside Air Temp" between $t - n$ to t. The sliding average is also a finite impulse response filter commonly called a boxcar filter. This has the benefit of suppressing high frequency noise. The second

statistical feature is the sliding standard deviation. The size of the sliding window is the size of order selected for delay variables.

2.4. Feature Selection and Data Reduction

2.4.1. Overview

Feature selection methods are conducted on the original variables and new features to find subsets of features useful for prediction of faults. Reducing the number of features is also desirable as it reduces the data requirements and computation time. To validate the usefulness of the new features to detect faults, we can observe if and how many of the new features are among the top selected features.

2.4.2. Univariate Feature Selection

First, variance is calculated for each feature. Then, features with zero variance are removed as their values are constant and provide no diagnostic information. Features with low variance can also be filtered out and removed from the list of candidate features.

There are numerous univariate feature selection methods such as univariate statistical tests of significance and similarity scores. These methods identify individual variables. However, individual variables may become redundant when used collectively.

2.4.3. Minimal-Redundancy-Maximal-Relevance criterion

Next, the mutual information based minimal-redundancy-maximal-relevance criterion (MRMR) is run on the derived features and the original features. MRMR is a feature selection method that is used to find a subset of features useful for prediction of the faults. MRMR reduces redundancy in the features and selects those most relevant to prediction [29]. MRMR can operate in one of two schemes: Mutual Information Difference (MID) or Mutual Information Quotient (MIQ).

First, the feature set S with the maximum relevance to the target class c is chosen based on the average mutual information between each individual feature x_i and class c. However, features selected according to maximum relevance could also have high redundancy, i.e., many highly dependent features. If two features are highly dependent, then removing one of the features would not have much of an effect on distinguishing between classes. Thus, a minimum redundancy condition is added [30]. MRMR combines the above two constraints.

2.4.4. Other Methods and Automated Tools

An overview of feature selection is provided in [26] and implementations of some automated feature selection methods are provided in the Python packages scikit-learn and sklearn. The authors of MRMR also released their own implementation of their MRMR algorithm in C/C++ and Matlab [30].

2.5. Machine Learning Model Creation

To detect a fault, the fault detection problem is framed as a classification problem. The classes can be "no fault", "fault type A", "fault type B", etc. The different types of faults can also be grouped together into a single "fault" class.

2.5.1. Creation of Training and Testing Sets

When testing on time series, care must be taken to not use future data to predict past data. To this end, training of the fault detection model is conducted on 'past' data and testing is conducted on 'future' data. For example, data from January is used as the training data set and data from February is used as the testing data set.

2.6. Performance Metrics

The metrics of recall, precision, and F1-score, metrics commonly used in classification problems, are used to quantify performance. The formulae for calculating recall, precision, and the F1-score are:

$$Recall = \frac{tp}{tp + fn} \tag{1}$$

$$Precision = \frac{tp}{tp + fp} \tag{2}$$

$$F1 = \frac{2tp}{2tp + fp + fn} \tag{3}$$

where tp is the number of true positives, i.e., correctly predicted fault samples, fp is false positives, fn is false negatives, i.e., fault samples incorrectly labelled as no-fault, and tn is true negatives.

The fault detection performance of the models that use the new features is compared against the fault detection performance of the models that use the raw variables for different numbers of features.

3. Sensor Data

3.1. Sensor Data Collection

The methodology described in Section 2 is demonstrated on the design of a fault detection system for a chiller plant. Data for normal operations and for operations under fault conditions was generated using a physics-based model of a commercial building chiller plant. It is undesirable to introduce faults in the real plant to collect measured data because introducing faults would interfere with normal operations of the facility and damage the expensive equipment. The physics-based model was created and calibrated by a research partner using the Modelica modeling language [31] within the Dymola modeling and simulation environment. The specific facility is the chiller plant at the Molecular Foundry at the Lawrence Berkeley National Laboratory. The Molecular Foundry is a six-floor nanoscience research facility cooled by a chiller plant that is operational 24 hours a day, seven days a week. The chiller plant has three chillers, two cooling towers, three primary chilled water pumps, three condenser water pumps, and one secondary chilled water pump.

Simulations were run for two months under no fault conditions and fault conditions for each fault that is investigated. The chiller plant faults included in this data set are: *cooling tower fan failure*, *abnormal chiller cycling*, and *secondary pump degradation*. While the types of faults present in the data affect the features that are selected in the feature selection process, the methodology presented here can be applied to other faults. The number of data points for each of the faults is summarized in Table 1.

Table 1: Included Faults and Corresponding Frequencies

Fault	**Number of Data Points**
Cooling Tower Fan Failure	775
Abnormal Chiller Cycling	16672
Secondary Pump Degradation	17518
No Fault	68538

From the simulation, measurements that correspond to available measurements common in real world chiller plant control systems were extracted at a commonly available sampling rate. This results in 52 features with a sampling resolution of every 5 minutes. These features include control signals, temperatures, power, and On/Off status of equipment in the chiller

plant, such as the cooling tower, chillers, and pumps. Table A1 lists the sensor points.

3.2. Data Cleaning and Pre-Processing

To clean the data, samples that are in between the five minute sampling times are removed. Samples with the same timestamp are averaged.

While the simulation provides continuous values, in the building control system, some variables are only binary On/Off signals. To ensure that the simulation reflect the available binary data in the building control system, for those variables, continuous values are converted to binary values.

3.3. Creation of Classification Labels

The fault/no-fault labels for the samples are determined based on the times for which the equipment with a fault is operating. Samples obtained from running the simulation with no faults is labelled with the 'no faults' label. When the simulation is run with a fault, the resulting samples are labelled with the corresponding 'fault' class label when the equipment with the fault is On. Otherwise, samples are labelled as 'no fault'. For example, when a simulation is run with a mechanical failure fault in Pump 1, samples during which Pump 1 is On is labelled with the 'mechanical fault' label.

4. Methodology Applied to the Case Study

4.1. Engineering Knowledge Features

The original features in the Building Automation System (BAS) are supplemented with new virtual sensor features created from the original features using knowledge of chiller plants. This knowledge of chiller plants includes knowledge of the quantities that the original features correspond to; the location of where data for those features is collected; and an understanding of the operation of the chiller plant. To collect this list of virtual sensor features, experts were interviewed in the process described in Section 2.2.

A summary of the new features derived from the original features using domain knowledge is shown in Table 2. This results in the addition of 67 features.

Table 2: Example Features from Knowledge of Chiller Plants

Cooling Output	Total Plant Tonnage
	Cooling Tonnage for Each Chiller
Efficiencies	Kilowatt per Ton for Each Chiller
	Efficiency for Each Cooling Tower
Differences	Temperature Range for Each Cooling Tower
	Chilled Water Delta T for Each Chiller
	Condensed Water Delta T for Each Chiller

4.2. Time Series Features

Time series features are created using the process described in Section 2.3.2. The inclusion of delay features results in 706 time-delay features. The order of the lag is selected using AIC and BIC. For each feature, the max order between AIC and BIC is selected. AIC consistently selected equal or higher orders than BIC. For most of the features, the maximum lag is selected because the AIC and BIC orders exceed the maximum lag. When the maximum lag is not selected, the resulting orders are close to the max. The variables for which the selected lags are not equal to the max lag are shown in Table 3, along with the selected lags by AIC and BIC.

4.3. Statistical Features

For each feature, the sliding mean and standard deviation is calculated. The window size for the sliding mean and standard deviation is the order as selected by AIC and BIC in the previous step. Where the AIC and BIC selected an order of 0, a window size of 12 is used.

4.4. Data Cleaning and Standardization

Data cleaning and standardization do not result in the creation of new features. Features that are constant are removed. Next, the data is standardized. The data is centered by removing the mean value of each feature, then scaled by dividing features by their standard deviation. This results in the data having a mean of zero and unit variance.

4.5. Machine Learning Model Training

4.5.1. Support Vector Machines

For this case study, support vector machines (SVMs) are used as the machine learning model. Support vector machines are supervised learning

Table 3: Selected lag orders for some features. For features not shown, the selected lag order is 12

Feature	**AIC Order**	**BIC Order**
Total Tonage	11	11
Chiller 1 Temp Setpoint	10	2
Chiller 1 Condenser Water Flow	12	11
Chiller 1 Leaving Condenser Temp	12	11
Chiller 2 Temp Setpoint	10	2
Chiller 3 Temp Setpoint	10	2
Chiller 3 Power	10	8
Chiller 3 On Off Status	12	7
Chiller 3 Condenser Water Delta Temp	6	12
Chiller 3 Leaving Chilled Water Temp	12	8
Cooling Tower 1 Fan Power	11	11
Primary Loop Chilled Water Supply Temp	12	6
Secondary Loop Pressure Difference	12	11
Secondary Loop Pump On Off Status	0	0
Secondary Loop Pump Control	12	11
Chiller 3 Condenser Water Pump Control	12	1
Chiller 3 Chilled Water Pump On Off	12	1
Relative Humidity	12	12

models with associated learning algorithms that solve for separating hyperplanes. SVMs are primarily used for classification and regression. Given training data that is labelled, SVMs solve for an optimal hyperplane that can classify new samples, depending on which side of the hyperplane the new samples are located.

4.5.2. Model Tuning and Cross Validation

A set of hyperparameters are used to specify an SVM model. The hyperparameters that are searched over are C, γ, and the kernel. These hyperparameters are tuned using 10-fold cross validation on the training set to find the hyperparameters which yielded the highest f1 classification score. The C parameter is a penalty on misclassification of training samples, which must be traded off against simplicity of the decision surface. A low value for C makes the decision surface smooth, while a high value for C aims to

correctly classify all training samples. The γ parameter defines the size of the influence of any single training sample. The larger the γ, the closer other samples must be to be affected. The kernel is a similarity function that calculates the similarity between two inputs to the function. The purpose of the kernel is to take the input data and transforms it in feature space to simplify the learning problem. For example, data that is not linearly separable may become linearly separable in polynomial feature space.

5. Results and Discussion

5.1. Selected Features

Table 4 lists the top ten features as selected by the MIQ and MID schemes to classify no-fault operations and fault operations. Because the two feature selection schemes use different measures of distance, some variation in rankings between the MIQ and MID schemes is expected. There is some overlap between the features selected under the MIQ and MID schemes.

Table 4: Top 10 features selected for fault VS no fault by MRMR under the MIQ and MID scheme with feature rankings

Feature	**MIQ**	**MID**
Chiller 3 kilowatt per ton at t-40min	1	NA
Secondary loop pump power at t-10min	2	2
Condenser water pump 1 OnOff status, t-15 min	3	NA
Chiller 3 kilowatt per ton	4	1
Secondary loop pump power standard deviation	5	4
Chiller 3 kilowatt per ton at t-60min	6	5
Chiller 3 kilowatt per ton at t-10min	7	NA
Secondary loop pump power	8	NA
Chiller 3 kilowatt per ton at t-55min	9	NA
Chiller 3 kilowatt per ton at t-5min	10	NA
Chiller 1 OnOff status at t-60min	NA	3
Relative humidity t-20min	NA	6
Cooling tower 1 control	NA	7
Chiller 3 kilowatt per ton at t-30min	NA	8
Chiller 1 OnOff status at t-5min	NA	9
Secondary loop pump power at t-50min	NA	10

In the feature selection process, the new derived features are favoured more than the original features. Under the MIQ scheme, all of the top ten features are new features, with one exception — Secondary loop pump power, which is ranked #8. All the top ten features under the MID scheme are all new features except for one feature — Cooling tower 1 control. The first ten features selected by the MIQ scheme represent six distinct data points from the control system. The top ten features contain representatives from all the different types of new features: application domain knowledge features, time series features, and statistical features. The type of new feature is marked in Table 5 in the same order as Table 4.

Table 5: Type of feature in top features selected by MRMR where Eng = Engineering Knowledge Features, Stat = Statistical Features, Lag = Time Series Features, and Orig = Original Features

Feature	**Eng**	**Stat**	**Lag**	**Orig**
Chiller 3 kilowatt per ton at t-40min	X		X	
Secondary loop pump power at t-10min			X	
Condenser water pump 1 OnOff status, t-15 min			X	
Chiller 3 kilowatt per ton	X			
Secondary loop pump power standard deviation		X		
Chiller 3 kilowatt per ton at t-60min	X		X	
Chiller 3 kilowatt per ton at t-10min	X		X	
Secondary loop pump power				X
Chiller 3 kilowatt per ton at t-55min	X		X	
Chiller 3 kilowatt per ton at t-5min	X		X	
Chiller 1 OnOff status at t-60min			X	
Relative humidity t-20min			X	
Cooling tower 1 control				X
Chiller 3 kilowatt per ton at t-30min	X		X	
Chiller 1 OnOff status at t-5min			X	
Secondary loop pump power at t-50min			X	

The number of new derived features among the features selected by MRMR is then compared against the total number of selected features. Comparisons are made between the number of selected new features and the expected value of selected new features if features were randomly selected.

Figure 2 plots the number of new derived features among the features selected by MRMR under the MIQ scheme, along with their respective rank-

ings. The dotted red line traces the expected curve if all the top features are new derived features. The deviations of the curves from the dotted reference line are the number of original features among the top selected features. Up to approximately the first 20 features of MIQ, the actual curves follow this reference line closely. This indicates that the top 20 selected features are almost all new features. In the top 50 selected features, there are 42 derived features and eight original measurement features. This suggests that among the top features selected to maximize relevance while minimizing redundancy, most of the selected features are new features. Under the MID scheme, more original measurement features are selected than under the MIQ scheme.

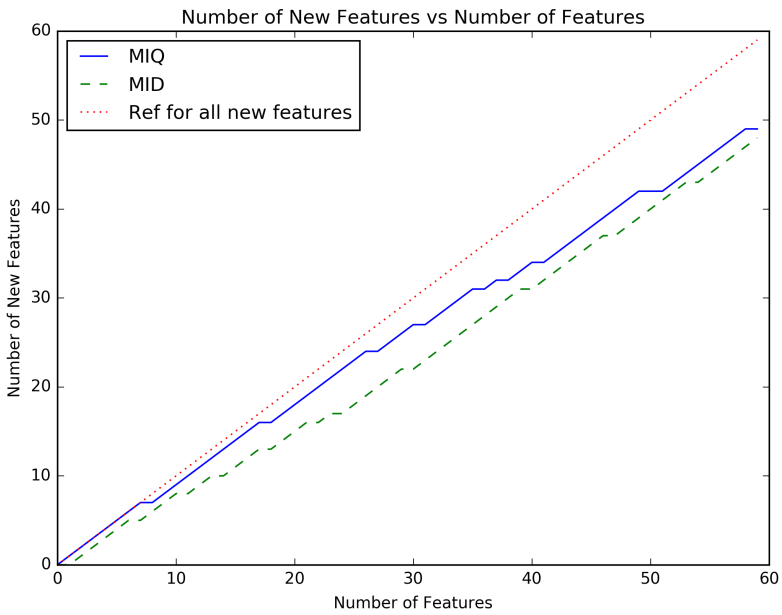

Figure 2: Number of new features among the top features with a dotted reference line indicating the expected line if all the features were new features

The top selected features are probabilistically related to the faults. However, the top selected features are not all directly at the fault location at time t. Many of the top selected features are delay features. The high ranking of expert engineering features agrees with field knowledge. For example, chiller kilowatt per ton is frequently recommended to monitor chiller performance. Chiller kilowatt per ton likewise ranks highly on this list, given that there

are chiller degradation faults in the data set. However, it is interesting to observe that the delays of chiller kilowatt per ton are selected. This may be due to transients in the system causing a delay in time between the start of the fault and the sensors registering symptoms of the fault. For example, in the case of the chiller, temperatures in the chiller may raise for a few minutes before temperature sensors in the chiller register, the temperature raises above the deadband, the chiller control system responds, compressors start working harder, and power consumption increases.

The top selected features are engineering features that are calculated from several sensor measurements. For example, chiller kilowatt per ton requires three to four input sensors. Thus, the use of engineering features fuses information from multiple sensors.

The selection of features is dependent on the faults in the data set. If a different set of faults is to be detected, the choice of features will very likely be different. Nonetheless, the results are promising in that nine faults representing four types of faults show fault detection performance reaching high asymptotic values at four features, as will be shown in the next section.

The feature selection method MRMR is based on information theory. Four main groups of feature selection methods have been identified: similarity-based, information-theoretical-based, sparse-learning-based, and statistical-based methods [26]. Table 6 lists the top ten features selected by common methods from the similarity-based and statistical-based group: Fisher score and chi-squared respectively. Both methods gave the same features in the same order. The top features are all variations of "Chiller 3 kilowatt", an engineering knowledge feature. The variations include the mean, which is a statistical value, and time lagged values. These feature selection methods consider each feature in relation to the fault prediction but do not consider interactions between the features themselves. Thus, the top features are all related.

5.2. Model Fault Prediction Results

The optimized set of existing and new virtual features resulted in fault detection with high detection performance while only using a surprisingly small number of features.

Figures 3 and 4 plot prediction performance metrics against the number of features. Figure 3 contains the performance metrics for detection of any of the faults in the data set, which is framed as a two class classification problem,

Table 6: Top features and corresponding feature type selected by chi-squared and Fisher score, where Eng = Engineering Knowledge Features, Stat = Statistical Features, Lag = Time Series Features, and Orig= Original Features

Rank	**Features Selected by Fisher Score and Chi-Squared**	**Eng**	**Stat**	**Lag**	**Orig**
1	Chiller 3 kilowatt mean	X	X	X	
2	Chiller 3 kilowatt per ton	X			
3	Chiller 3 kilowatt per ton at t-5min	X	X	X	
4	Chiller 3 kilowatt per ton at t-10min	X	X	X	
5	Chiller 3 kilowatt per ton at t-15min	X	X	X	
6	Chiller 3 kilowatt per ton at t-20min	X	X	X	
7	Chiller 3 kilowatt per ton at t-25min	X	X	X	
8	Chiller 3 kilowatt per ton at t-30min	X	X	X	
9	Chiller 3 kilowatt per ton at t-35min	X	X	X	
10	Chiller 3 kilowatt per ton at t-40min	X	X	X	

where the two classes are "fault" and "no fault". Figure 4 considers the the fault identification problem, where each fault is a different class.

Within Figures 3 and 4, the first column of graphs plots the F1 score, the second column the precision, and the third column the recall. The first row of graphs within each figure are the scores on the training set and the second row of graphs are the scores on the testing set. The x-axis is the number of features. Based on the higher scores under the MIQ scheme compared with the MID scheme, experiments were only run under the MIQ scheme. The metrics for fault detection are for the "fault" class and the metrics for fault identification are the weighted average across all classes.

In the fault detection setup using MRMR with expert knowledge features, with as few as four features, classification scores exceed 90% on the training set and 80% on the testing set. The original measurement feature set has 52 features, which is significantly more than the number of features required under this methodology. Recall scores are consistently above 90% starting at one feature. The precision increases with increasing number of features, which in turn improves the F1 score. This is because the F1 score is a weighted sum of precision and recall.

For fault identification using MRMR with expert knowledge features, classification scores are slightly lower by 5% F1 score on the testing set while remaining the same on the training set compared to the fault detection set-

ting. F1 scores in the fault identification setting do not drastically change between one and two features. At three or more features, performance plateaus. Performance also varies between the types of faults.

All scores asymptote after four features. The asymptotes on the training data are approximately 95% for all the performance metrics using MRMR with expert knowledge. The asymptotes on the testing data are approximately 80% for precision, 90% for recall, and 85% for F1 score. The 10% difference between performance on the training set versus the testing set suggests that the fault detection model is slightly over-fitting to the training data.

We reran the experiments with features selected by MRMR from among features that did not use knowledge gained from experts during the interview process. That is, we trained machine learning models without the engineering knowledge features and without the statistical features, which required information about appropriate time windows obtained during the interview process. Experts were asked the appropriate window size. The choice of window would not have been obvious to individuals who are non-HVAC experts. The appropriate window size was found during the interview process, particularly from the questions "What time scales do you look at?" and "What time scales would be too noisy?" When the features with expert knowledge are removed, fault detection performance drops by more than 40% on the training set and more than 30% on the testing set in F1-score, as shown in Figure 3. The selected features are listed in Table A2 in Appendix B. All data points are classified as faults by the resulting fault detection system, resulting in poor recall and poor F1 scores. Fault identification performance drops by almost 20% in F1-score, as shown in Figure 4. The results suggest that gathering knowledge from experts to create new features improves fault detection and fault identification performance.

For features selected by the Fisher score or chi-squared with expert knowledge features, prediction performance is worse than using features selected by MRMR with expert knowledge features by a large margin. This difference is more than 40% in the F1 score on the training set and more than 30% on the F1 score on the testing set compared to using features selected by MRMR with expert knowledge. The fault detection and identification performance using Fisher score or chi-squared to select from features that include features derived with expert knowledge is about the same as performance with MRMR without using features derived with expert knowledge.

The prediction performance using Fisher score or chi-squared is also in-

dependent from the number of features, holding constant within 2%, except at four features in the fault detection setup. The consistency is because the selected features are all related and correlated with one another, being the mean and time lags of the same value. Thus, adding them does not benefit predictions.

These results show that feature selection methods that consider redundancy between features give better fault detection performance than feature selection methods that do not. Time lagged features and sliding means are highly correlated for measured physical properties in energy systems. Thus, feature selection methods that take into account correlation between features should be used instead of univariate methods that consider features individually.

Misclassification error rate on the training set can be traded-off with misclassification error rate on the testing set. That is, the model can be trained to fit the training data less and a higher misclassification rate on the training data can be tolerated in exchange for hopefully lower misclassification rate on the test set. This is known as the bias-variance trade-off. The training set was chosen to be historical data in relation to the testing set. Thus, the testing set likely deviates from the training set due to changing weather conditions impacting cooling load demand and chiller plant performance.

To detect a larger number of different faults, however, the number of features needed for fault detection may be higher than as shown in these results. Different faults at different locations within the chiller plant will likely require features at the different locations. Different faults within the same location may also need additional features to distinguish between the faults. Nonetheless, the proposed methodology allows for user choice in the types of faults to focus on. Also, this methodology enables a trade off between detection accuracy and number of fault types with data collection and computation costs. Real sensor data points can be supplemented with virtual sensor data points and can be processed in workflows used to create fault detection models and machine learning models.

5.3. Selected Sensors

From Figure 3, most of the improvement in fault detection performance occurs with the addition of the first three features. Fault detection performance plateaus after the third feature. This suggests that with the first three features, we can already achieve most of the predictive performance for this machine learning model.

The most important sensors are the sensors associated with these three features. The sensors associated with these three features are shown in Table 7. There are only five sensors associated with these features. In comparison, this HVAC system had 52 original sensor points.

Table 7: Top three features (virtual sensors) and their associated real sensors

Feature	**Associated Sensor Values**
Chiller 3 kilowatt per ton at t-40min	Chiller 3 Power Chiller 3 Evaporator Entering Temperature Chiller 3 Evaporator Leaving Temperature
Secondary loop pump power at t-10min	Secondary loop pump power
Condenser water pump 1 OnOff status at t-15 min	Condenser water pump 1 On Off status

5.4. Limitations

The effectiveness of this proposed framework to benefit from expert derived features is limited by the cooperation of experts to share their knowledge. We have found that experts are more open to discussions when the interview intent is framed as identifying their pain points and designing a useful data analysis and information tool for them. That is, the discussion should focus on their needs and desires in an automated fault detection, instead of how to automate their jobs away.

Because this is a data-driven approach to fault detection, this method requires data under both fault and fault-free operating conditions. This data can be either real or simulated. Data may also be normalized for differences in, for example, plant sizes and equipment sizes. Generalizability of fault predictions may degrade over time due to sensor drift and changes to the plant such as retrofits and changes to the control strategy.

Another limitation of this method is the reliance on extracting sensor data from the control system, the ease of which may be highly variable. The data selection method does not consider the cost of the sensor data. It is possible to remove expensive sensors from the ranked list of selected sensors and then re-rank the remaining candidates.

The input data into this framework requires a consistent sampling rate across samples. The time series and statistical features requires a span of

data across several time steps. Missing data truncates continuous time spans of data, but missing data can be interpolated or imputed.

6. Conclusion

When trying to extract meaningful information from existing data and deployed sensors, system designers may be faced with an overwhelming number of data points. This paper presents a data-driven approach to designing and configuring energy monitoring systems. The proposed approach automatically selects data to input into energy monitoring systems using statistical methods. A method based on information theory is used to recommend features that are strong predictors of a fault, are highly relevant, and minimally redundant. The results show that only a few sensors may be needed to detect faults with high accuracy rates and that the sensors can be identified using automated statistical tools.

Application domain knowledge can be encoded to create virtual sensors. These virtual sensors do not correspond to values taken directly off a physical sensor and can be more predictive of a fault than direct sensor data. The top selected features have a probabilistic relationship to the physical location of a fault, but may not be the sensor directly at the fault location at the time of the fault. The proposed methodology allows for user choice in the types of faults to focus on and the ability to trade off detection accuracy and number of types of faults with data collection and computation costs.

The ability to detect faults with high accuracy with few sensors is encouraging for widespread deployment of such a fault detection system. The few numbers of features shows that across a type of system, for data collection purposes, individual instances of a system do not need to share a large number of sensors. To configure the data collection part of such a fault detection system, only a subset of all the sensors need to be identified.

The flexibility in the proposed design methodology eases adoption of the methodology for different types of systems. This methodology is demonstrated on a chiller plant in a commercial building and has also been applied to wind turbines [32] [11]. It can be generalized to a wide range of other instrumented energy systems, such as HVAC systems, solar panels and natural gas turbines. The expert knowledge collection process outlined here can be applied to experts of other energy systems. The methodology can be followed to select the most useful sensors for fault detection. Thus, those who are not

experts in the energy system nor a specific instance can follow the proposed scheme to create a fault detection system.

To apply this methodology to design a fault detection system for another application only requires 1) the identification of common existing data points, and 2) basic application knowledge in the form of a list of metrics. The application knowledge can be gathered, for example, through interviews of application experts or through analysis of basic engineering relationships in the system. The proposed methodology enables increased fault detection accuracy with fewer features and fewer data points with fewer installation and configuration requirements from vendors and experts. The resulting software can also be readily deployed to other instances of the same type of system, due to the lack of reliance on individual specific information. Deployment costs may also be lowered due to this approach's demonstrated effectiveness when using basic application knowledge and off-the-shelf, open source tools and algorithms.

6.1. Future Work

The proposed fault detection methodology can be extended to include faults that evolve or worsen and also can be used to account for when faults are fixed, the physical system is changed or upgraded, or when maintenance occurs, as happens in real world systems. The fault detection capabilities can be expanded to include fault diagnosis, identification, and prognosis. This can be done, for example, by incorporating knowledge bases, simulations of fault development, or other representations of expert knowledge of fault evolution. The modularity of the proposed methodology allows for future research on more effective choices of parameters. For example, further analysis can be conducted on the choice of feature selection process, machine learning model, and time series characteristics.

Additional research is needed to collect more sensor data and fault labels under normal operations and different faults. Labelled data is difficult to find. Sensor measurements under faulty operations are even more difficult to obtain, due to the infrequency of faults and the undesirability of introducing faults to expensive systems for the purposes of data collection.

New methods to combine data from different sources and across different systems are also needed. A new deployment may have limited, if any, historical data, and it is desirable to use data from other deployments and other systems, both real measurements and simulated data. Given differences between individual instances, for example between individual buildings or sim-

ulated and real systems, procedures are needed to fuse data from different sources to take advantage of available data of comparable systems.

Future research can also improve the ease of installation and the user experience. User interfaces should be easy to use, accessible, and encourage consistent response to faults and energy performance improvement opportunities. This may be accomplished through research in integration of fault detection systems with operations and maintenance, design of the control automation system, and design of user interfaces. Energy monitoring could be as simple as verbally speaking or assisted by digital assistants. More advanced decision-analytic queries, such as where to invest a quantity of money for energy improvement, may also be possible.

Acknowledgement

The authors would like to thank Rongxin Yin, Thierry Stephane Nouidui, Kevin Leahy, and Ioannis Konstantakopoulos.

Lawrence Berkeley National Laboratory's contribution to this work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, of the U.S. Department of Energy under Contract No. DE- AC02- 05CH11231.

Bibliograpy

[1] U. E. I. Administration, Annual Energy Outlook 2013, U.S. Energy Information Administration, 2013.

[2] S. Katipamula, M. R. Brambley, Review article: Methods for fault detection, diagnostics, and prognostics for building systemsa review, part i, HVACI&R Research 11 (1) (2005) 3–25. arXiv:http://www.tandfonline.com/doi/pdf/10.1080/10789669.2005.10391123, doi:10.1080/10789669.2005.10391123.
URL `http://www.tandfonline.com/doi/abs/10.1080/10789669.2005.10391123`

[3] S. Lou, D. H. Li, J. C. Lam, W. W. Chan, Prediction of diffuse solar irradiance using machine learning and multivariable regression, Applied Energy 181 (2016) 367 – 374. doi:https://doi.org/10.1016/j.apenergy.2016.08.093.
URL `http://www.sciencedirect.com/science/article/pii/S0306261916311916`

[4] S. Daz, J. A. Carta, J. M. Matas, Performance assessment of five mcp models proposed for the estimation of long-term wind turbine power outputs at a target site using three machine learning techniques, Applied Energy 209 (2018) 455 – 477. doi:https://doi.org/10.1016/j.apenergy.2017.11.007.
URL http://www.sciencedirect.com/science/article/pii/S0306261917315866

[5] C. Robinson, B. Dilkina, J. Hubbs, W. Zhang, S. Guhathakurta, M. A. Brown, R. M. Pendyala, Machine learning approaches for estimating commercial building energy consumption, Applied Energy 208 (2017) 889 – 904. doi:https://doi.org/10.1016/j.apenergy.2017.09.060.
URL http://www.sciencedirect.com/science/article/pii/S0306261917313429

[6] C. Fan, F. Xiao, Y. Zhao, A short-term building cooling load prediction method using deep learning algorithms, Applied Energy 195 (2017) 222 – 233. doi:https://doi.org/10.1016/j.apenergy.2017.03.064.
URL http://www.sciencedirect.com/science/article/pii/S0306261917302921

[7] J.-S. Chou, N.-T. Ngo, Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns, Applied Energy 177 (2016) 751 – 770. doi:https://doi.org/10.1016/j.apenergy.2016.05.074.
URL http://www.sciencedirect.com/science/article/pii/S0306261916306717

[8] J. Drgoa, D. Picard, M. Kvasnica, L. Helsen, Approximate model predictive building control via machine learning, Applied Energy 218 (2018) 199 – 216. doi:https://doi.org/10.1016/j.apenergy.2018.02.156.
URL http://www.sciencedirect.com/science/article/pii/S0306261918302903

[9] Y. Peng, A. Rysanek, Z. Nagy, A. Schlter, Using machine learning techniques for occupancy-prediction-based cooling control in office buildings, Applied Energy 211 (2018) 1343 – 1358. doi:https://doi.org/10.1016/j.apenergy.2017.12.002.
URL http://www.sciencedirect.com/science/article/pii/S0306261917317129

[10] X. Kong, X. Xu, Z. Yan, S. Chen, H. Yang, D. Han, Deep learning hybrid method for islanding detection in distributed generation, Applied Energy 210 (2018) 776 – 785. doi:https://doi.org/10.1016/j.apenergy.2017.08.014.
URL http://www.sciencedirect.com/science/article/pii/S0306261917310267

[11] R. L. Hu, K. Leahy, I. C. Konstantakopoulos, D. M. Auslander, C. J. Spanos, A. M. Agogino, Using domain knowledge features for wind turbine diagnostics, IEEE International Conference on Machine Learning and Applications.

[12] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S. N. Kavuri, A review of process fault detection and diagnosis: Part i: Quantitative model-based methods, Computers I& Chemical Engineering 27 (3) (2003) 293–311. doi:http://dx.doi.org/10.1016/S0098-1354(02)00160-6.
URL http://www.sciencedirect.com/science/article/pii/S0098135402001606

[13] R. Isermann, Model-based fault-detection and diagnosis status and applications, Annual Reviews in Control 29 (1) (2005) 71 – 85. doi:http://dx.doi.org/10.1016/j.arcontrol.2004.12.002.
URL http://www.sciencedirect.com/science/article/pii/S1367578805000052

[14] S. Yoon, J. F. MacGregor, Fault diagnosis with multivariate statistical models part i: using steady state fault signatures, Journal of Process Control 11 (4) (2001) 387 – 400. doi:http://dx.doi.org/10.1016/S0959-1524(00)00008-1.
URL http://www.sciencedirect.com/science/article/pii/S0959152400000081

[15] S. J. Qin, Process data analytics in the era of big data, AIChE Journal 60 (9) (2014) 3092–3100. doi:10.1002/aic.14523.
URL http://dx.doi.org/10.1002/aic.14523

[16] J. Granderson, R. Singla, E. Mayhorn, P. Ehrlich, D. Vrabie, S. Frank, Characterization and survey of automated fault detection and diagnostics tools, Tech. Rep. LBNL-2001075, Lawrence Berkeley National Laboratory (November 2017).

[17] M. Najafi, D. M. Auslander, P. L. Bartlett, P. Haves, M. D. Sohn, Application of machine learning in the fault diagnostics of air handling units, Applied Energy 96 (2012) 347 – 358, smart Grids. doi:https://doi.org/10.1016/j.apenergy.2012.02.049.
URL http://www.sciencedirect.com/science/article/pii/S0306261912001481

[18] R. L. Hu, J. Granderson, A. Agogino, Detection of chiller energy efficiency faults using expectation maximization, ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.

[19] M. Bonvini, M. D. Sohn, J. Granderson, M. Wetter, M. A. Piette, Robust on-line fault detection diagnosis for hvac components based on nonlinear state estimation techniques, Applied Energy 124 (2014) 156 – 166. doi:https://doi.org/10.1016/j.apenergy.2014.03.009.
URL http://www.sciencedirect.com/science/article/pii/S0306261914002311

[20] Y. Guo, J. Wall, J. Li, S. West, A machine learning approach for fault detection in multi-variable systems, in: ATES in conjunction with Tenth Conference on Autonomous Agents and Multi-Agent Systems (AAMAS) AAMAS, Vol. 2011, 2011.

[21] Z. Du, B. Fan, X. Jin, J. Chi, Fault detection and diagnosis for buildings and {HVAC} systems using combined neural networks and subtractive clustering analysis, Building and Environment 73 (2014) 1 – 11. doi:http://dx.doi.org/10.1016/j.buildenv.2013.11.021.
URL http://www.sciencedirect.com/science/article/pii/S0360132313003417

[22] I. Khan, A. Capozzoli, S. P. Corgnati, T. Cerquitelli, Fault detection analysis of building energy consumption using data mining techniques, Energy Procedia 42 (2013) 557 – 566. doi:http://dx.doi.org/10.1016/j.egypro.2013.11.057.
URL http://www.sciencedirect.com/science/article/pii/S1876610213017591

[23] B. Sun, P. B. Luh, Q. Jia, Z. O'Neill, F. Song, Building energy doctors: An spc and kalman filter-based method for system-level fault detection in hvac systems, IEEE Transactions on Automation Science and Engineering 11 (1) (2014) 215–229. doi:10.1109/TASE.2012.2226155.

[24] K. Bruton, P. Raftery, P. O'Donovan, N. Aughney, M. M. Keane, D. O'Sullivan, Development and alpha testing of a cloud based automated fault detection and diagnosis tool for air handling units, Automation in Construction 39 (2014) 70 – 83. doi:https://doi.org/10.1016/j.autcon.2013.12.006.
URL http://www.sciencedirect.com/science/article/pii/S0926580513002276

[25] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
URL http://dl.acm.org/citation.cfm?id=944919.944968

[26] J. Li, K. Cheng, S. Wang, F. Morstatter, T. Robert, J. Tang, H. Liu, Feature selection: A data perspective, arXiv:1601.07996.

[27] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: Selected papers of hirotugu akaike, Springer, 1998, pp. 199–213.

[28] G. Schwarz, et al., Estimating the dimension of a model, The annals of statistics 6 (2) (1978) 461–464.

[29] C. Ding, H. Peng, Minimum redundancy feature selection from microarray gene expression data, Journal of bioinformatics and computational biology 3 (02) (2005) 185–205.

[30] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Transactions on pattern analysis and machine intelligence 27 (8) (2005) 1226–1238.

[31] Modelica and the Modelica Association, `https://www.modelica.org/`, accessed: 2017-02-17.

[32] K. Leahy, R. L. Hu, I. C. Konstantakopoulos, C. J. Spanos, A. M. Agogino, Diagnosing Wind Turbine Faults Using Machine Learning Techniques Applied to Operational Data, IEEE International Conference on Prognostics and Health Management.

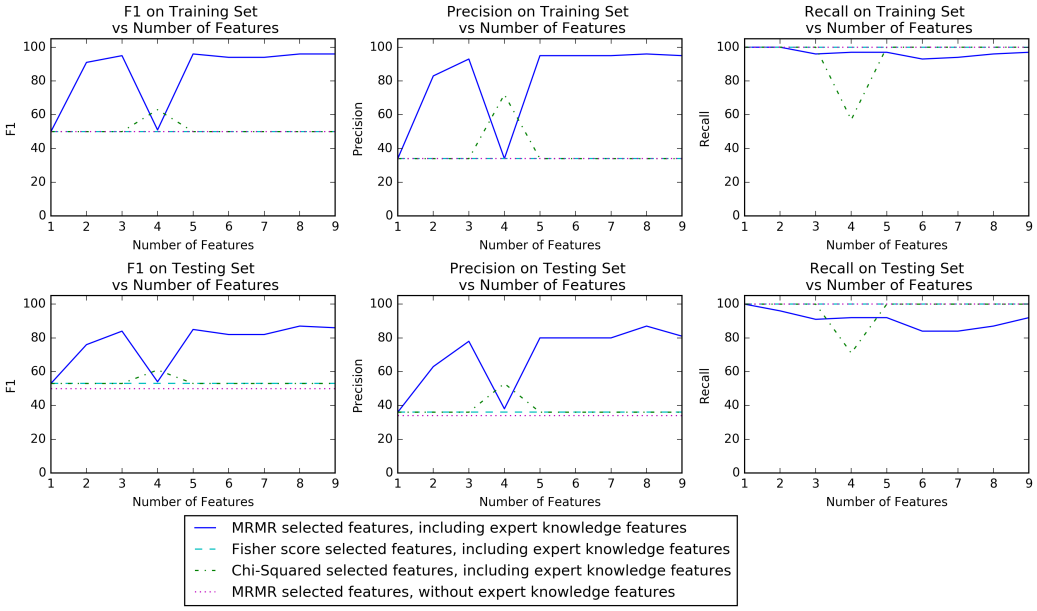

Figure 3: Scores on training and testing sets under the MIQ scheme for fault detection (classification into fault vs. no fault)

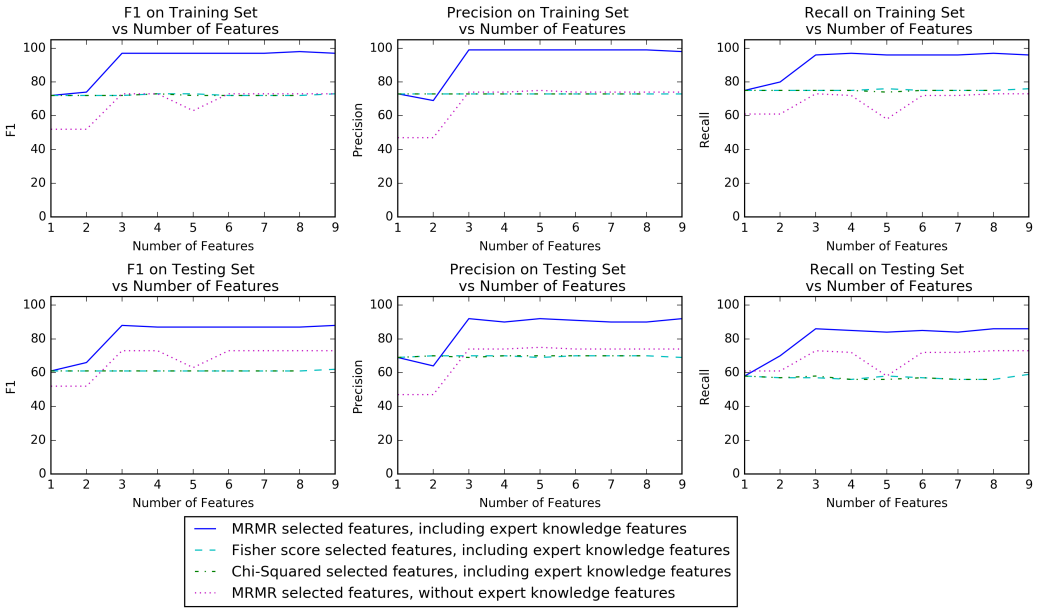

Figure 4: Scores on training and testing sets under the MIQ scheme for fault diagnostics (classification into different types of faults)

Appendix A: Original Sensors

Table A1: Full list of original sensors

For each of the Chillers 1-3
Chiller Power
Chiller Condenser Entering Temperature
Chiller Condenser Leaving Temperature
Chiller Evaporator Entering Temperature
Chiller Evaporator Leaving Temperature
Chiller Setpoint Temperature
Chiller On/Off Status
For each of the Cooling Towers 1-2
Cooling Tower Fan Power
Cooling Tower Approach Temperature
Cooling Tower Leaving Temperature
Cooling Tower Control Signal
For each of the Condenser Water Pumps 1-3
Condenser Water Pump Control Signal
For each of the Primary Chilled Water Pumps 1-3
Primary Chilled Water Pump Control Signal
Secondary Chilled Water Pump
Secondary Chilled Water Pump Control Signal
Secondary Chilled Water Pump Power
Secondary Chilled Water Loop
Secondary Loop Pressure Drop
For each of the Air Refrigeration Units 1-4
Chilled Water Temperature
Weather
Wet Bulb Temperature
Relative Humidity
Time
Second of the Year (represents time in seconds from 12am Jan 1)

Appendix B: MRMR Selected Features without Expert Knowledge

Table A2: Top features and corresponding feature type selected by MRMR from among the features that do not require expert knowledge

Rank	Features Selected by MRMR	Lag	Original
1	Cooling tower 2 approach temperature		X
2	Primary pump 2 On Off Status		X
3	Secondary pump power		X
4	Secondary pump power at t-60min	X	
5	Secondary pump power at t-5min	X	
6	Secondary pump power at t-30min	X	
7	Secondary pump power at t-10min	X	
8	Secondary pump power at t-55min	X	
9	Secondary pump power at t-15min	X	
10	Secondary pump power at t-50min	X	