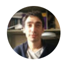
Machine Learning-Based Prognostics for Central Heating and Cooling Plant Equipment Health Monitoring

Article *in* IEEE Transactions on Automation Science and Engineering · June 2020

DOI: 10.1109/TASE.2020.2998586

4 authors:

Chunsheng Yang
National Research Council Canada



Burak Gunay
Carleton University

Zixiao Shi
Shopify inc.

Weiming Shen
Huazhong University of Science and Technology

Machine Learning-Based Prognostics for Central Heating and Cooling Plant Equipment Health Monitoring

Chunsheng Yang, Burak Gunay, Zixiao Shi, and Weiming Shen, *Fellow, IEEE*

***Abstract*— Fault detection, diagnostics, and prognostics (FDD&P) ensure the operation efficiency and safety of engineering systems. In the building domain, they can help significantly reduce energy consumption and improve occupant comfort. Specifically, prognostics are becoming increasingly important as a pro-active fault prevention strategy through continuously monitoring the health of energy systems. In this article, we develop a machine learning-based method for building systems. The proposed method can help develop predictive models from historical operation and maintenance data. After the detailed description of the proposed machine learning-based prognostic method, a case study involving prognostics on central heating and cooling plant (CHCP) equipment is provided. To this end, a year's worth of sensor and actuator data from four boilers and five chillers of a CHCP in Ottawa, Canada are collected. The plant operators are interviewed to understand how they handle failure events, and their logbooks are reviewed to extract the date and time of the recorded failure events. The sensor and actuator data up to two weeks prior to each of these failure events are used to develop regression tree models that predict time to failure (TTF). The results indicate that about half of the modeled failure events could be accurately predicted by looking at the data available in the distributed control system. Finally, the future work is outlined.**

***Note to Practitioners*—This article was motivated by the problem of fault detection, diagnostics, and prognostics (FDD&P) of the building systems. We contemplated to develop an advanced technology for heating, ventilation and air conditioning (HVAC) prognostics, in particular, for central heating and cooling plant (CHCP) health monitoring, aiming to save energy consumption and the operational cost. The developed machine learning-enabled predictive modeling technique, which can help build predictive models from historic operational and maintenance data, can be applied to other application domains such as oil pipeline system monitoring and high-speed train prognostics.**

***Index Terms*— Central heating and cooling plant (CHCP), data-driven models, heating ventilation and air conditioning (HVAC), machine learning, predictive modeling, prognostics.**

I. Introduction

HEATING, ventilation, and air conditioning (HVAC) systems inside commercial buildings often experience suboptimal operation due to unintended faults. The presence of such faults can lead to energy waste, thermal discomfort, or even safety hazards [1]–[4]. An example of these faults include stuck reheat valve inside variable air volume (VAV) terminals, failed damper motors inside air handling units (AHUs), and lower supply water temperature that leads to tripping events inside boilers [5], [6]. Although early detection algorithms for most zone- and system-level faults are usually sufficient, faults in critical central equipment should be prevented to avoid large-scale implications. For example, a central boiler failure can lead to loss of heating and hot water supply for multiple buildings. To this end, prognostics should be applied and a way to do so is to predict the time to failure (TTF) [2].

However, lack of prognostics and predictive maintenance in building systems has been rarely studied. Of the few research articles in this field, Yang *et al.* [7] employed text mining on operator logbooks, and produced multiple high-level metrics such as failure probabilities and mean-TTF. Nirjon *et al.* [8] investigated the usage of audio recordings for the development of HVAC prognostics. Goldman [9] used vibration data to predict potential HVAC equipment failures. Ahmad and Atta [10] used electric current to predict failures in the HVAC motors. Wang *et al.* [11] developed an algorithm based on particle filters for the prognosis of heat exchanger degradation.

To address these challenges and advance HVAC prognostics, this article proposes machine learning-based modeling methods for HVAC prognostics specifically for central heating and cooling plant (CHCP) equipment health monitoring. Using the machine-learning algorithms, we developed predictive models from historic building operation data; and such data-driven models are intended to predict CHCP failures/faults before they occur and estimate TTF for a given failure or a monitored component or subsystem. Consequently, the building operators can perform proactive actions to fix failures or faults before they interrupt the routine operation; in turn, it will

Manuscript received January 30, 2020; revised March 28, 2020; accepted May 25, 2020. This article was recommended for publication by Editor B. Vogel-Heuser upon evaluation of the reviewers' comments. This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada, in part by the Natural Science Foundation under Grant 61963026, and in part by the National Research Council Canada. *(Corresponding author: Chunsheng Yang.)*

Chunsheng Yang is with the National Research Council Canada, Ottawa, ON K1A 0R6, Canada, and also with the Department of Education Technology, Shanghai Normal University, Shanghai 200234, China (e-mail: chunsheng.yang@nrc-cnrc.gc.ca).

Burak Gunay is with the Department of Civil Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: burak.gunay@carleton.ca).

Zixiao Shi is with the National Research Council Canada, Ottawa, ON K1A 0R6, Canada (e-mail: zixiao.shi@nrc-cnrc.gc.ca).

Weiming Shen is with the School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: wshen@ieee.org).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2020.2998586

reduce maintenance costs and CHCP energy consumption, and improve occupant comfort.

This article presents the proposed prognostic methods for CHCP health monitoring. We demonstrate it by applying the proposed methods to boiler and chiller prognostics in a CHCP as a case study. In this application, we used the sensor and actuator data from four boilers and five chillers. Furthermore, the date and time of the failure events were extracted from the operator logbooks by consulting the building operators. First, the plant operators were interviewed to understand their workflow and how they handle abrupt failure events. Subsequently, regression algorithms were examined to predict the TTF by looking at the sensor and actuator data records leading to these failure events. Operating conditions associated with failure instances were identified for CHCP health monitoring.

This article is organized as follows. Following the introduction is a related work section on HVAC fault detection, diagnostics, and prognostics (FDD&P). Section III presents a machine learning-based method to develop predictive models from historic operational and maintenance data. Section IV presents a case study to demonstrate the application of the proposed modeling methods to boiler and chiller prognostics. Section V discusses the preliminary results obtained from the case study, limitations and remaining challenges. Section VI provides the conclusion.

II. Related Work

Two types of faults can occur within building systems: hard faults and soft faults. Hard faults are physical faults in sensors, meters, actuators, and other equipment. An example of such a fault is an AHU damper getting stuck open, which leads to higher outdoor air intake, increasing energy consumption. If the heating/cooling capacity of the AHU is not sized to compensate for this fault, indoor temperature setpoints will not be met, further leading to comfort or even safety problems. Soft faults, on the other hand, are caused by inappropriate control programs. Soft faults can also waste energy and cause discomfort. Furthermore, they usually go undetected for long periods of time since stakeholders could simply assume the building systems are operating "as intended." For instance, if the minimum fresh air intake is not calculated correctly within the control program, spaces within the building may not receive sufficient fresh air, which can lead to poor indoor air quality or even health hazards. So far, the related work has been focusing on HVAC FDD. These works can be categorized into performance indexing-based methods and inverse modeling methods.

Performance indexing-based FDD largely relies on thresholds to detect and diagnose issues. They generate alarms when sensor readings exceed thresholds. In contrast, inverse model-based FDD looks at the statistical characteristics of prediction residuals to identify abnormal performance. Two main types of modeling techniques exist: gray box modeling [14] and black box modeling [15], [16]. Gray box models are simplified models derived from analytical solutions. The idea is to capture the dominant contributors (such as outdoor air temperature) to performance changes, and use data-driven approaches to estimate the model parameters. This avoids pitfalls of overfitting a complicated first principle physical model with a high degree of freedom. Black box models [17], as their name suggests, are statistical or machine learning models that do not necessarily capture the physical dynamics of a system. These data-driven models have been used to predict the heating or cooling loads through machine learning algorithms [18], [19], generalized linear models [20], and others. Most black box models are used to predict the energy consumption of buildings or provide an evidence for setting control policies [21], [22].

Instead of comparing a single performance metric with the previous approach, inverse modeling tries to estimate the model parameters and use them for FDD purposes. For instance, parameters of a black box model which represent the effects of ambient temperature to heating load can be an indicator of performance deviations [24]. Both gray box and black box models can be used for this purpose in FDD applications. It is desirable that FDD&P is able to provide a solution for building operators to detect, identify, or predict failures at the system level or the component level [7].

Most of the previous research studies are focused on FDD, while little work has been done on developing prognostics for building systems. However, the prognostics, practically, has been widely applied in other disciplines, such as aerospace, railway, and automobile industries [24], [25]. In general, prognostics involve developing models to predict failures before they interrupt the normal operation given a target time window and to estimate the TTF or so-called remaining useful life. With accurate TTF estimation, proactive action [27] can be performed in time and at the right place.

III. Machine Learning-Based Methodology

Building on techniques from machine learning, data mining and statistics, we developed a novel method for CHCP prognostic. The proposed method is intended to help develop predictive models from historic building operation and maintenance data. The models are expected to predict faults in CHCP equipment well before they occur, and sometimes they may be designed to estimate the TTF. Table I presents the pseudocode of the proposed methods. The methods consist of four main processes: data preparation (*dataGatheringAndProcess()*), data engineering (*dataEngineering()*), model building (*ModelBuilding()*), and model evaluation (*ModelEvaluation()*). Following sections describe each process in detail.

A. Data Preparation

In general, prognostics are defined as a process to develop models to predict failures well before they occur and to estimate the remaining useful life or TTF. To this end, we cast CHCP prognostic problem as a binary classification task with two class values: positive ("*P*") and negative ("*N*") and a regression task, developing a regression model to estimate the TTF. Many supervised learning techniques can be used to address this problem. Therefore, the task is to develop binary classifiers or regression models from the historic data. To create a comprehensive data set for prognostics modeling, one would need to retrieve CHCP failure events, extract the

TABLE I
PROPOSED MACHINE LEARNING-BASED METHODS

```
Input:   Operation database (DB) and
         Maintenance database (MDB)
Output: Predictive Model (m)
------------------------------------------------
Process:
{
   Mdataset = dataGatheringandProcess(DB, MDB);
                          /* create a dataset from  DB and MDB */
   For all x_ij in Mdataset {
        c_ij = Label( x_ij , w);        /* labeling data */
   }
   For selected attribute a_i from a⃗ {
          fa = dataEngineering(a_i);
                     /* data engineering for feature generation */
   }
   (S, T) = split(Mdataset);
          /* split Mdataset into training dataset (S) and
                                   testing dataset (T) */
  For each A^i in A{
     For adopted algorithm j {
        m_ij = ModelBuilding(A^i, algorithm j, S );
              /* building models. A^i is a subset of the A*/
     }
  }
  For all m_ij{
        m = ModelEvaluation(m_ij, T);  /* evaluating models */
  }
}                              /* end of the process */
```

operation data corresponding to each failure, and label each instance in the data set.

In Table I, *DB* and *MDB* are data from historic routine operation and maintenance. x_{ij} is an instance from time series j at ith position. c_{ij} is a class value for instance x_{ij}. w is the time window for labeling (discussed below). $\vec{a}$ is a vector of attributes, and a_i is a given attribute for feature engineering. *fa* is a new feature from feature engineering algorithms and it is added to $\vec{a}$ to create a feature set A, a full set of features, and A^i is a subset from A. S and T are the data set for training and testing. While conducting modeling experiments later A is divided into different subsets. For each subset, we apply different algorithms to building models, noted as $m_{ij.}$ Finally, one suitable model noted as m is selected from a set of models.

1) Retrieving Failure Events: In building operation, HVAC failure events are logged into a maintenance (such as work-order records) database. These records provide useful information such as dates, fault types, problem descriptions, actions taken to fix the problems, fault locations (subsystems or components), and causes. Currently, there are two kinds of records: operators' logbooks and event reports from building energy management systems (BEMSs) [22]. From such a data set, it is possible to retrieve a set of failure events which contain date, time, location, HVAC component identifier, and so on. Furthermore, these faults can be linked to the collected operation data [7], [22].

2) Gathering Operation Data: Once a set of failure events is extracted from the work-order database, the relevant instances from the operational database can be retrieved based on the selected sensor data points. For each failure, the data are obtained m days prior to each failure. The number m depends on the data set and the component; however, m is generally selected so that there are enough instances prior to a failure. The real instance number depends on the application and sample rate. Usually, we need to consider the balance between normal operation data samples and near-failure data samples.

3) Labeling: In order to use classification algorithms to build/train classifiers, we need to add a class value to each instance in the obtained data set (*Mdataset* in Table I). The number of possible class values depends on the classification task. For binary classification tasks, the class value is either "P" (for potential of a failure) or "N" (for no failure) given a target time window (w). The algorithm, *label(),* is designed to implement this labeling task. Using (1), we add a class attribute (c_{ij}) to each instance (x_{ij}) based on the timestamp of each instance. For regression task, we have to add the ground truth of TTF to each instance. The target time window (w) can be used in TTF estimations. The beginning and end of a target time window need to be determined empirically based on the requirements of an application. For HVAC prognostics, we assigned w as 14 days before a failure happens

$$c_{ij} = \begin{cases} p, & \text{if } t_w \in w \\ N, & \text{if } t_w \notin w. \end{cases} \tag{1}$$

B. Data Engineering

As in other machine-learning applications, the quality of the representation of the data input is a key factor for building high-performance predictive models. Therefore, data engineering is to generate or extract the attributes from the original data set to enhance the data representation. Accordingly, we rely on constructive induction to create new powerful features using the feature extracting algorithms available in machine learning, including principal component analysis (PCA), and wavelet of time series. We also perform time series analysis to extract potentially relevant time series characteristics. For instance, to smooth the noise influence we can compute the moving average in a given window for a specified attribute to add this new feature to feature set (A). Finally, we apply feature selection algorithms on the augmented data representation to automatically remove correlated or irrelevant features. Data engineering is a time-consuming and empirical task, and it is also a data-dependent task.

C. Model Building

Our goal is to build a predictive model for the classification task or regression task—i.e., classifying HVAC operation data as "positive" (fault status) or "negative" (normal status). Therefore, many classification algorithms are available from machine learning research, including instance-based learning (IBk), Naïve Bayes, support vector machine (SVM), decision trees, and artificial neural networks (ANNs), even deep-learning algorithms such as recurrent neural network (RNN) and convolutional neural network (CNN). For regression task (TTF estimation), the algorithms available are logistic regression, random forest decision tree, deep nets, and so on. In this article, we tend to prefer simpler algorithms such as decision trees, random forest decision tree, and Naïve Bayes over more complex ones because they can be applied quickly and generate outputs that are suitable for human interpretation. We systematically apply the same algorithm several times with varying attribute subsets to obtain a set of heterogeneous models for the model evaluation step.

D. Model Evaluation

After predictive models are built from the historical data, it is necessary to evaluate the model performance, aiming to select a suitable model for CHCP prognostics. To this end, we have to adapt an adequate criterion. The simplest way is to compute model performance by either error-rate or accuracy. The error rate is defined as the expected probability of misclassification: the number of classification errors (i.e., the false positives and the false negatives) over the total number of test instances. The accuracy is 1minus the error-rate. Because some errors can be costlier than others, it is sometimes desirable to minimize the misclassification cost rather than the error-rate. For example, false positives can be more acceptable than false negatives for critical CHCP equipment such as chillers and boilers. The received operation curve (ROC) method, area under the ROC curve (AUC) [24], and cost curve are popular in practical applications since they allow the user to evaluate models based on the proportions of positive instances. In prognostic, the main task is to predict the failures for monitoring systems and to estimate its TTF. It requires us to take TTF into consideration. Unfortunately, all existing metrics do not take TTF into consideration; they are not available for evaluating the performance of prognostic models. Therefore, we proposed a score-based method to rank models. The score is computed with (2) based on the samples on testing data set during the model evaluation

$$\text{score} = \left(\frac{\text{NbrOfPositive}}{\text{NbrofCases}}\right)^{\text{Sign}} \sum_{i=1}^{p} rc_i \tag{2}$$

where p is the number of positive predictions in a given g data set, NbrofPositive the number of HVAC failure events which contain at least one positive prediction in the target time window (w), NbrofCase the total number of HVAC failure events in a given test data set, and Sign the sign of $\sum_{i=1}^{p} rc_i$.

When Sign < 0 and NbrofPositive $= 0$, the score is set to 0; and rc_i is calculated as follows:

$$rc_i = \begin{cases} -1, & t < T_3 \quad \text{and } t \nexists w \\ \alpha t, & T_3 \leq t <= T_2 \quad \text{and } t \in w \\ 1, & T_2 < t <= T_1 \quad \text{and } t \in w \\ 1 - \alpha t, & T_1 < t <= T_0 \quad \text{and } t \in w \end{cases} \tag{3}$$

where rc_i is the award value for the instance I based on the predictive timing, w is the given target time window for failure prediction. T_0 is the failure time, T_1, T_2, and T_3 are constants determined by w. This definition reflects different reward values for the different phases of failure prediction. In prognostics, the prediction can be divided into four phases: normal stage ($t < T_3$), early failure stage ($T_3 \leq t < T_2$), effective prediction stage of failure ($T_2 \leq t < T_1$), and later stage of the failure prediction ($T_1 \leq t < T_0$). For normal stage of the operation, the model should predict the states as negative. If it predicts the state as positive, this will be a false positive. Therefore, the reward score will be -1 as computed from (3). For other three stages, the reward score value is computed differently. The parameter α is a coefficient for CHCP prognostics.

To evaluate the performance of machine learning-based regression models (TTF estimation), the traditional metrics available from statistics are deployed. These metrics are the mean absolute error (MAE), the sum of squared error (SSE), the mean squared error (MSE), the root mean squared error (RMSE), and the mean absolute percentage error (MAPE). They are defined as follows:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |e_i| \tag{4}$$

$$\text{SSE} = \sum_{i=1}^{N} e_i^2 \tag{5}$$

$$\text{MSE} = \sum_{i=1}^{N} \sqrt{e_i} \tag{6}$$

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{7}$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N} \left|\frac{e_i}{y_i}\right| \times 100 \tag{8}$$

where e_i is the individual prediction error; y_i is the actual value; and N is the number of samples in the testing data set.

Usually, we run the model by applying the test set or the so-called unseen data and compute the metrics. Each metric has advantages and limitations given application requirements. In this article, we compute three metrics, namely MAE, MSE, and RMSE for evaluating the performance of TTF estimation regression models.

IV. Case Study: Prognostics of Chiller and Boiler

We have demonstrated the usefulness of the proposed modeling methods to develop predictive models for CHCP failure prediction, as a classification task by using simulated building operation data in [27]. In this article, we merely focus on the demonstration of development of regression models for estimating TTF for CHCP. Therefore, we built predictive models to estimate the TTF for chillers and boilers monitoring using the collected operational data and maintenance records.

This case study was conducted on data from a CHCP with four boilers and five chillers serving ten buildings in Ottawa, Canada. First, the plant operators were interviewed to better understand how they undertake operational decisions and handle emergency maintenance events. Subsequently, a year's worth of operations data from the boilers and

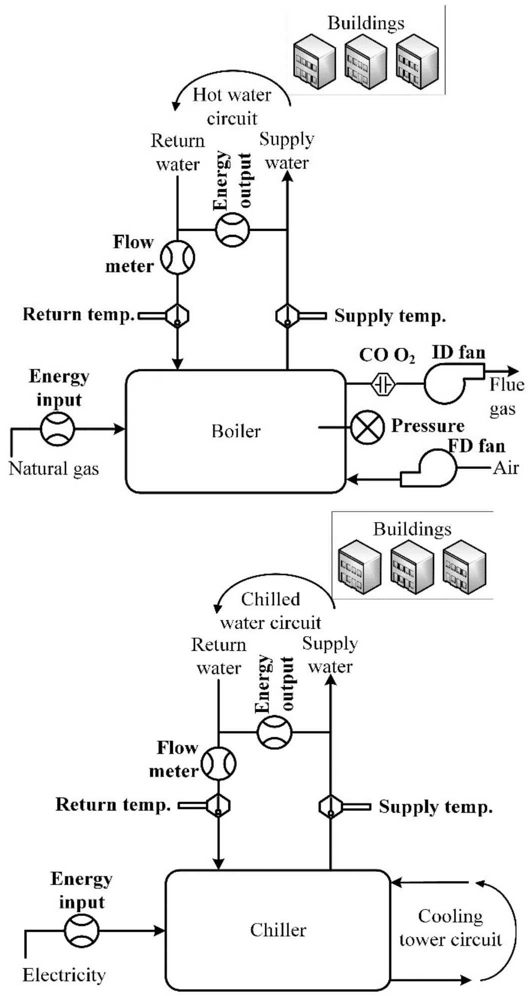


Fig. 1. Control points of (top) boilers and (bottom) chillers of this article.

chillers were extracted and analyzed for developing machine learning-based predictive models following the proposed modeling methodology; finally, we conducted the experiments for modeling and model evaluation. The results will be discussed in the next section.

A. Data Description

The sensor and actuator data from these boilers and chillers were sampled at 10-min intervals from the distributed control system of the CHCP and stored in a central database. The data from the boilers include the input energy from the natural gas (Q_{in}), output thermal energy (Q_{out}), drum pressure (P), induced and forced draft fan states (fan_{ID} and fan_{FD}), hot water flow rate (Q_{flow}), and return and supply water temperatures (T_{ret} and T_{sup}). In addition, the flue gas CO and O_2 concentrations (CO and O_2) were monitored from three of the four boilers. The data from the chillers include the input energy from the electricity, output thermal energy, chilled water flow rate (Q_{flow}), and return and supply water temperatures (T_{ret} and T_{sup}). Fig. 1 presents a schematic of the sensor points on the boilers and chillers. Note that for the chillers, the cooling tower and refrigerant circuits were automated separately, and the data were not available within the distributed control system of the CHCP.

TABLE II

Average of the Data From the Boilers. R_{pl}, E, and F_{fail} Are the Average Part-Load Ratio, Efficiency, and Failures in 1000 h of Operation, Respectively

Boiler	T_{ret} (°C)	T_{sup} (°C)	Q_{flow} (L/s)	fan_{FD} (%)	fan_{ID} (%)	P (Pa)	CO (ppm)	O_2 (%)	R_{pl}	E	F_{fail}
1	116	144	35	22	8	-29	5.8	6.1	0.5	0.86	0.3
2	113	144	34	20	12	-26	12.4	5.5	0.5	0.83	0.4
3	114	144	35	21	9	-26	10.5	5.6	0.5	0.86	0.0
4	133	143	22	17	53	-1	-	-	0.3	0.65	2.5

TABLE III

Average of the Data From the Chillers. R_{pl}, COP, and F_{fail} Are the Average Part-Load Ratio, COP, and Failures in 1000 h of Operation, Respectively

Chiller	T_{ret} (°C)	T_{sup} (°C)	Q_{flow} (L/s)	R_{pl}	COP	F_{fail}
1	8.5	5.3	83	0.4	3.7	0.0
2	8.0	4.0	116	0.7	6.4	1.6
3	10.0	5.0	100	0.7	7.2	1.1
4	9.3	4.6	120	0.7	5.7	0.0
5	9.6	5.2	138	0.5	5.6	0.0

Tables II and III present the descriptive statistics for the data records collected from individual boilers and chillers, respectively. Note that each boiler operates under different conditions. For example, boiler 4 is intended to meet the hot water needs during the cooling season. Boiler 1 operates during the shoulder seasons and the winter months. And, boilers 2 and 3 are used to meet the space heating needs during the winter months. Boiler 4 tends to have lower efficiency, lower water flow rates, higher return temperatures, and higher induced draft fan actuation rates than the other three boilers. Despite higher levels of induced draft fan actuation, the depressurization of boiler 4 was significantly less than the other boilers. Average heating output from the boilers was 30%–40% of their capacity. Another important metric in a boiler's performance is the composition of the flue gas. When there is not enough O_2 supplied to the burner, CO will form. And, if there is excess O_2 supplied to the burner, useful heat will leave the exhaust flue with the hot gases. The flue gas of boilers 1–3 contained 3%–9% O_2 during operation. Similar to the boiler data set, the conditions when the chillers were operational were vastly different. For example, the mean water flow rates in different chillers ranged from 83 to 138 L/s. The mean outdoor temperatures when these chillers were operational ranged from 15 °C to 23 °C. On average, the cooling energy output from these chillers was about 40%–70% of their capacity. The average coefficient of performance (COP) of these chillers ranged from 3.7 to 7.2. It is important to note that the authors did not upgrade or recalibrate the sensors used in this article; we were interested in developing prognostics models from commercial grade, typical in use, CHCP data from chilled/hot water circuits.

The failure instances were extracted by searching for the term tripping in the operator logbooks. We identified over 20 recorded tripping events. About 11 of the recorded tripping events resulted in the equipment ceasing to operate—that is, the equipment could not be reset. In this article, our focus is on developing models predicting these 11 failure events. Nine of these 11 events were associated with the boilers, and two were associated with the chillers. Seven of the nine failure events in boilers were generated by boiler 4.

B. Failure Event Extraction and Failure Time Series Creation

An interview was conducted with the operators to better understand their role in the operation of the plant and their maintenance-related decision-making process. We used a questionnaire approach to get operators' insight for CHCP maintenance and routine operation procedure. We designed six questions as follows.

1) What type of day-to-day operational decisions do you make?
2) Which boilers and chillers do you use in which season and on which days of the week? How do you make this decision?
3) Do you have any information regarding the efficiencies of the chillers or boilers in your plant?
4) How do you respond to chiller and boiler tripping events? Is there a standard protocol in-place to handle such events?
5) Do the plant operators use a common logbook to capture important events such as routine maintenance events or abrupt failures? If yes, is this a paper-based or computer-based logbook?
6) Would you have benefited by having information regarding operating conditions leading to tripping events?

Table IV provides the operators' responses for these interview questions. Two senior plant operators were present during the interview. It was identified that the operators heuristically decide which of the four boilers and five chillers should be turned ON/OFF each day. There were no written protocols to guide them in making these sequencing decisions. The operators reported that they prefer to use boiler 4 and chiller 1 during the cooling season, and boiler 1 and chiller 4 during the heating season. They indicated that their preference was primarily based on the convenience to start these boilers and chillers and their perceived reliability. Furthermore, they did not have access to any information regarding the equipment efficiencies and failure rates. Note that the sequencing of the equipment plays an important role on the operating conditions. If conservatively many boilers/chillers are kept operational, the equipment will operate at undesirable part load conditions—which may adversely affect the plant energy efficiency and failure rates.

About 14 days of operations data prior to each of the 11 failure events were extracted, forming 11 distinct data sets corresponding to 11 failures. As we described above, each failure is associated with a time series (with time stamps). With time series data, we can generate some new features such as moving average if necessary. These data sets were used to train prognostics models predicting the TTF. The regressors were the sensor and actuator data up to two weeks before each failure event, and the response variable was the natural logarithm of the remaining TTF (i.e., the response variable was log-transformed). In this preliminary investigation, we decided to use regression trees to develop the prognostics models because of their simplicity and suitability to derive rules to avoid operational conditions leading to failure events.

TABLE IV

PLANT OPERATORS' RESPONSES

Q #	**Responses from the operators**
Q1	• *Decide the daily on-off sequencing of the boilers and chillers* • *Monitor operation of the plant equipment* • *Troubleshoot prompted by alarms* • *Schedule routine maintenance activities*
Q2	• *Use boiler 4 during the cooling season and boiler 1 during the shoulder and heating seasons* • *Use chiller 4 during the heating season and chiller 1 during the shoulder and cooling seasons* • *Use other chillers and boilers occasionally when needed* • *Decision is based on the convenience to start, reliability, habits*
Q3	• *Do not have any information about the equipment efficiencies*
Q4	• *Try to resolve the problems internally first* • *Often need to engage external contractors*
Q5	• *A paper-based logbook is used to capture sequencing, abrupt failures, and routine maintenance events* • *A computerized maintenance management system is not available*
Q6	• *Unsatisfied about frequent false positive alarms with the existing automation system* • *Expressed interest in better AFDD&P capabilities*

C. Prognostic Modeling

Following the proposed modeling methods, we develop regression models to estimate TTF for each failure using the data generated above. The regression tree models for each of the 11 failure events were developed by employing a binary recursive partitioning procedure—the CART algorithm as implemented in MATLAB's *fitrtree* function. The CART algorithm automatically makes decision splits in recursion and keeps branching out until a stopping criterion is satisfied. The decision splits at each branch node are made with respect to the regressors which yield the greatest information gain, whereby the information gain is defined as the difference between the entropy (a metric to quantify level of uncertainty) before and after a decision split. Thus, the regressors of the greatest importance are expected to appear at the top of the tree—and likely to appear multiple times. Insignificant regressors may be used infrequently at lower tier decision splits—or they may not be used at all. For example, if return water temperature is an important variable to predict the remaining time to a failure event, it would appear at the top of the tree. If it is an insignificant predictor, it may not appear in the trained model at all.

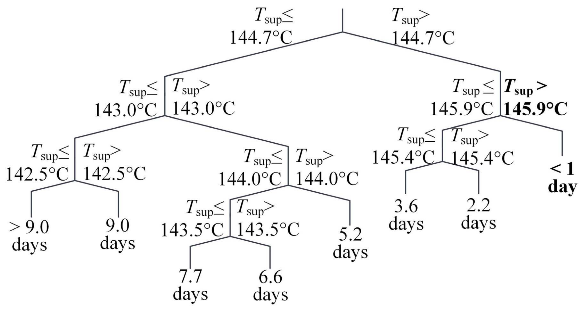


Fig. 2. Illustrative example for a regression tree model predicting the remaining TTF. It is developed by using the operations data from boiler 1 failure event 1. The critical conditions prior to the failure event are highlighted in bold font.

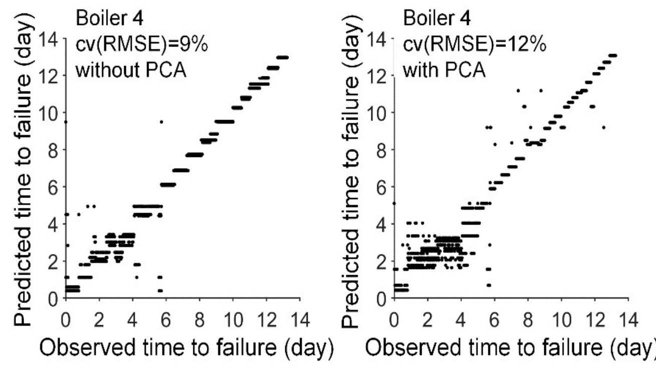


Fig. 3. Example demonstrating the predictive accuracy of the TTF models. The example is for boiler 4 failure event 5. Each scatter point corresponds to a predicted and an observed TTF event.

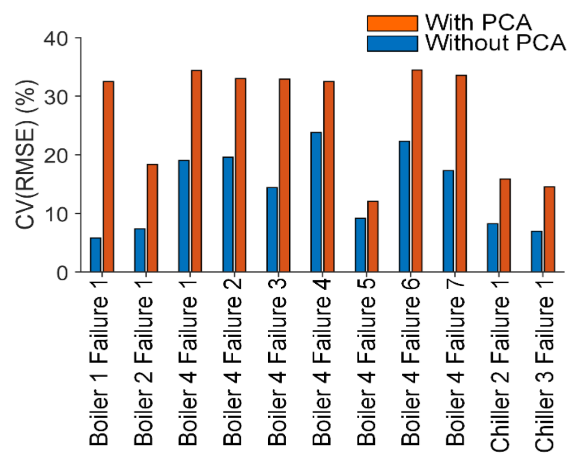


Fig. 4. Accuracy of the regression tree models in predicting the remaining TTF. The accuracy is quantified in terms of the CV (RMSE) metric, which is the RMSE normalized by the mean of the response variable.

D. Regression Trees and Rule Extraction

As described above, the machine-learning algorithm, regression tree used is prognostic modeling. Therefore, the developed prognostic model is a "regression tree." The trees are expected to grow to a maximal size without the use of stopping rules. To create parsimonious trees, three measures were taken in this article. First, the splitting was terminated when there were less than 12 h worth of data remaining at a given leaf node. Second, the sections of a tree that provide no additional information were removed—an action also known as pruning. Third, a twofold cross-validation approach was employed. In addition to these three measures, to understand the effect of using a dimensionality reduction technique, the regression trees were developed both with and without the PCA procedure. When the PCA is used, the regressors are transformed to linearly uncorrelated variables—that is, principal components. Only the top principal components explaining 95% of the variance are used in regression tree model training. Further information on classification and regression trees can be found elsewhere [12], [13].

Fig. 2 presents an example regression tree model predicting the TTF for boiler 1. In this example, all the decision splits were across a single variable: the supply water temperature T_{sup}. About nine days prior to this failure event, T_{sup} of boiler 1 started slowly increasing from less than 142.5 °C to more than 145.9 °C. Therefore, one can argue that this failure event was not an abrupt occurrence, and the sensor data could lend itself for the detection of this event several days in advance.

V. Results and Discussion

Fig. 3 presents an example demonstrating the predictive accuracy of the TTF models for boiler 4 failure event 5. The coefficient of variation (CV) of the RMSE CV was also annotated in Fig. 3. The results indicate that the use of PCA for dimensionality reduction slightly reduced the predictive accuracy of the models in this example. As shown in Fig. 4, the use of PCA more drastically increases the CV (RMSE) in all other failure events. This can be attributed to the fact that PCA is an algorithm that does not consider the response variable. It treats the predictors with large variance as the important predictors; while this process may result in principal components with low (or no) importance weights on some important predictors. Consequently, it is concluded that the PCA technique in predictor selection for prognostics models for boiler and chillers may not be appropriate.

The results shown in Fig. 4 indicate that five 5 of the 11 failure events can be predicted at a CV (RMSE) lower than 10%. The remaining six events were predicted at a CV (RMSE) higher than 15%. Simply put, not all failure events can be accurately detected in advance by looking at trends in the operations data available in this article. Of the five events that could be detected with a CV (RMSE) of less than 10%, three were in boilers and two were in chillers. In boiler 4, only one of the seven failure events could be detected at a CV (RMSE) less than 10%. This can be interpreted that the failure events are less predictable with some equipment than others. This preliminary analysis suggests that developing prognostic-oriented models for chillers and boilers may be possible. However, these models will not be suitable for some failure modes which do not have noticeable early symptoms at least several hours in advance. Furthermore, such models

TABLE V

OPERATING CONDITIONS PRIOR TO FAILURE. THESE CONDITIONS ARE EXTRACTED FROM THE REGRESSION TREE MODELS (WITHOUT USING PCA) BY LOOKING AT THE CONDITIONS WITHIN 24 h TO THE FAILURE EVENT. THE EVENTS THAT WERE PREDICTED WITH A CV (RMSE) OF 10% OR LESS WERE HIGHLIGHTED IN BOLD FONT

Event	*Conditions leading to failure*	*Fraction of operating time conditions met*
Boiler 1 Failure 1	$\boldsymbol{T_{sup} > 146°C}$	**0.09**
Boiler 2 Failure 1	$\boldsymbol{T_{ret} > 103°C \wedge CO < 30}$ **ppm**	**0.01**
Boiler 4 Failure 1	$Q_{flow} > 22$ L/s $\wedge$ $T_{ret} > 136°C$	0.12
Boiler 4 Failure 2	$T_{sup} > 145°C$	0.17
Boiler 4 Failure 3	$Q_{flow} > 24$ L/s $\wedge$ $T_{ret} > 133°C$	0.05
Boiler 4 Failure 4	$T_{sup} < 142°C \wedge Q_{flow} < 22$ L/s	0.09
Boiler 4 Failure 5	$\boldsymbol{Q_{out} > 68}$ **kW** $\boldsymbol{\wedge Q_{flow} < 0.7}$ **L/s**	**0.01**
Boiler 4 Failure 6	$T_{ret} > 137°C$	0.14
Boiler 4 Failure 7	$T_{sup} < 140°C \wedge T_{ret} > 133°C$	0.03
Chiller 2 Failure 1	$\boldsymbol{T_{ret} < 11°C \wedge T_{sup} > 4°C}$	**0.01**
Chiller 3 Failure 1	$\boldsymbol{T_{ret} < 19°C \wedge T_{sup} < 5°C}$	**0.01**

may not perform adequately for some equipment with limited sensing or due to sensor/meter calibration issues.

Table V presents the operating conditions prior to the 11 failure events. These conditions were extracted from the regression tree models (without using the PCA technique) by looking at the conditions within 24 h of the predicted failure event. In Table V, the events that were predicted with a CV (RMSE) of 10% or less are highlighted in bold font. The results indicate that each failure condition appears to be somewhat unique. For example, the failure event in boiler 1 followed supply water temperatures exceeding 146 °C; whereas, the failure event in boiler 2 followed return water temperatures exceeding 103 °C when the flue gas CO concentration was less than 30 ppm. Given that we have operations data associated with a small number of failure events, we did not formulate metarules overarching multiple events from the same equipment type. As more data become available, future research should investigate the possibility of developing metarules for CHCP prognostics.

The five failure conditions defined by the accurate regression tree models (which were highlighted in bold font in Table IV) were infrequent occurrences. Four of them occurred during only 1% of the equipment's operating hours, and one of them occurred during 9% of the equipment's operating hours. Note the use of inaccurate models can lead to many false positives or negatives. Both situations may cause operators to ignore the prognostic alarms and may have negative consequences that outweigh the benefits of true-positive prognostics.

Aside from making predictions for the failure events, the regression tree models can be used to reconfigure the control sequences of equipment during recommissioning. For example, a control technician can adjust the setpoints and the control loop parameters to avoid possible failure conditions identified through regression tree modeling and rule extraction. Building on the findings of this article future research should demonstrate these application cases by using models developed upon more comprehensive data sets.

It is important to note that this article is one of the first efforts toward prognostics-oriented modeling of CHCP systems. It was conducted on a limited data set with only a small number of events. Thus, many fundamental aspects, such as the ability to derive metarules overarching multiple failure modes in different equipment, could not be explored. Building on the findings of this preliminary study, future work should develop prognostic-oriented models based on more comprehensive data sets and demonstrate their use in the operational decision-making process. At the same time, the following two challenges are still remaining in predictive modeling for CHCP prognostics: model adaptation for operational environment changing and model transparency for reasoning rationales.

A. Model Adaptation Ability

In this article, the predictive models are developed using supervised learning techniques. All models were built using one-shot training data sets with labeled (ground truth) samples. When the models are used to predict unseen data or the so-called test data, it is assumed that the independent and identical distribution (i.i.d.) from test data set is the same as that in training data set. Due to this strong constraint, the current predictive models exist with the following deficiencies: 1) once the operational environments or conditions change, the model may not work well or even become useless in the new environment and 2) a large number of samples with labels is required. In particular, while deploying the models into prognostic health management (PHM) systems to monitor the system operation and forecast the states of a system, it is still difficult to answer the following questions in practice.

1) How long can such machine learning-based predictive models function well as expected for a given application domain?
2) If the operational environments change from time to time, are the models still trustable?
3) If the model is built with the data from chiller A in a CHCP, does it fit a similar chiller B in other CHCP?

For instance, if one has access to raw data from thousands, or even millions, of buildings, how the model built from one building can be deployed effectively for all buildings?

Transfer learning (or called domain adaptation, continuous learning, lifetime learning, concept learning, multitask learning, or robust learning) [28]–[30] can transfer the trained models cross domains by reducing or minimizing the distribution difference, such that the previous trained models could still work well. Transfer learning has been widely applied to many real-world applications such as text categorization, sentiment analysis, image classification, and video summarization. In order to transfer the trained predictive models to new operation environments or adapt the models to the new domains, we are working on developing a transfer learning-based methodology to address this challenge as our on-going task

B. Model Transparency of Reasoning Rationales

In developing machine learning-based models, one remaining challenge is how we can explain the behavior of model reasoning. So far, all models developed using machine learning techniques are a "black box." There is rarely a proper solution to help understand the rationales of the model output. In machine learning community, there are few of research focusing on transformation of a model into a decision table using if–then rules by losing some performance; and using a polynomial math model to express the "black box" insight by finding a convergence formula as close as possible. However, both approaches cannot be easy to find a right solution since e performance of model will degrade largely. Therefore, this is a major barrier to extend the application of machine learning-based modeling technologies, in particular, in HVAC prognostics. This will be our future work.

VI. Conclusion

This article proposed a machine learning-based predictive modeling methodology for developing data-driven models for CHCP/HVAC prognostics. The proposed modeling approach was applied to develop regression models for estimating TTF for chiller and boilers monitoring in a CHCP. An analysis was carried out upon operations data gathered from four boilers and five chillers of a CHCP to develop prognostic-oriented models.

In this article, we hypothesized that operating conditions should relate to failure events. Thus, we argued that we may be able to understand unfavorable operating conditions and predict the TTF by looking at conditions prior to a failure event. To this end, the dates of equipment failure events were extracted from the plant operators' logbooks. The operations data up to two weeks prior to each of these failure events were used to develop regression tree models that predict the remaining TTF. In addition, the influence of a dimensionality reduction algorithm PCA was investigated in reducing the number of predictors used in the regression tree model development process.

The results indicate that about half of the modeled failure events could be predicted accurately by looking at the data available in the distributed control system of the plant. It was identified that the use of PCA drastically reduced the predictive accuracy of the models. Finally, rules that define unfavorable conditions were extracted by using regression tree models. These rules can be used by a control technician to reconfigure the setpoints and control loop parameters.

Acknowledgment

The authors would like to thank the reviewer's valuable comments and recommendations from CASE 2019 to extend the conference paper into this special issues paper.

References

[1] S. Katipamula and M. Brambley, "Review article: Methods for fault detection, diagnostics, and prognostics for building systems—A review, Part II," *HVAC R Res.*, vol. 11, no. 2, pp. 169–187, Apr. 2005.

[2] S. Katipamula and M. R. Brambley, "*Review Article*: Methods for fault detection, diagnostics, and prognostics for building systems—A review, Part I," *HVAC R Res.*, vol. 11, no. 1, pp. 3–25, 2005.

[3] W. Kim and S. Katipamula, "A review of fault detection and diagnostics methods for building systems," *Sci. Technol. Built Environ.*, vol. 24, no. 1, pp. 3–21, 2017.

[4] B. Gunay, W. Shen, B. Huchuk, C. Yang, S. Bucking, and W. O'Brien, "Energy and comfort performance benefits of early detection of building sensor and actuator faults," *Building Services Eng. Res. Technol.*, vol. 39, no. 6, pp. 652–666, Nov. 2018.

[5] Y. Yu, D. Woradechjumroen, and D. Yu, "A review of fault detection and diagnosis methodologies on air-handling units," *Energy Buildings*, vol. 82, pp. 550–562, Oct. 2014.

[6] D. Dey and B. Dong, "A probabilistic approach to diagnose faults of air handling units in buildings," *Energy Buildings*, vol. 130, pp. 177–187, Oct. 2016.

[7] C. Yang, W. Shen, Q. Chen, and B. Gunay, "A practical solution for HVAC prognostics: Failure mode and effects analysis in building maintenance," *J. Building Eng.*, vol. 15, pp. 26–32, Jan. 2018.

[8] S. Nirjon, R. S. Srinivasan, and T. Sookoor, "Smart audio sensing-based HVAC monitoring," in *Smart Cities: Foundations, Principles, and Applications*. Hoboken, NJ, USA: Wiley, 2017, pp. 669–695.

[9] S. Goldman, "Vibration-based predictive maintenance for HVAC systems in buildings," *ASHRAE J.*, vol. 37, no. 1, pp. 24–28, 1995.

[10] N. Ahmad and R. Atta, "Cost-effective wireless-controlled motor failure prediction for HVAC system in large buildings using demodulated current signature analysis," *Life Sci. J.*, vol. 11, no. 10s, pp. 33–39, 2014.

[11] P. Wang, R. X. Gao, and Z. Fan, "Switching local search particle filtering for heat exchanger degradation prognosis," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2015, pp. 539–544.

[12] T. A. Reddy, *Applied Data Analysis and Modeling for Energy Engineers and Scientists*. Springer, 2011.

[13] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2016.

[14] S. Wang, Q. Zhoy, and F. Xiao, "A system-level fault detection and diagnosis strategy for HVAC systems involving sensor faults," *Energy Buildings*, vol. 42, pp. 477–480, Apr. 2010.

[15] H. B. Gunay, W. Shen, and C. Yang, "Blackbox modeling of central heating and cooling plant equipment performance," *Sci. Technol. Built Environ.*, vol. 24, no. 4, pp. 396–409, 2007.

[16] B. Gunay, W. O'Brien, and I. Beausoleil-Morrison, "Control-oriented inverse modeling of the thermal characteristics in an office," *Sci. Technol. Built Environ.*, vol. 22, no. 5, pp. 586–605, 2017.

[17] B. Gunay, W. Shen, and C. Yang, "Characterization of a building's operation using automation data: A review and case study," *Building Environ.*, vol. 118, pp. 196–210, Jun. 2017.

[18] G. K. F. Tso and K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy*, vol. 32, no. 9, pp. 1761–1768, Sep. 2007.

[19] M. Yalcintas and S. Akkurt, "Artificial neural networks applications in building energy predictions and a case study for tropical climates," *Int. J. Energy Res.*, vol. 29, no. 10, pp. 891–901, 2005.

[20] T. Catalina, J. Virgone, and E. Blanco, "Development and validation of regression models to predict monthly heating demand for residential buildings," *Energy Buildings*, vol. 40, no. 10, pp. 1825–1832, Jan. 2008.

[21] M. Najafi, "Fault detection and diagnosis in building HVAC systems," Ph.D. dissertation, Dept. Mech. Eng., Univ. California, Berkeley, CA, USA, 2010.

[22] C. Yang, S. Létourneau, and H. Guo, "Developing data-driven models to predict BEMS energy consumption for demand response systems," in *Proc. 27th Int. Conf. Ind. Eng. Appl. Artif. Intell. Expert Syst. (IEA/AIE)*, Taipei, Taiwan, 2014, pp. 188–197.

[23] C. Yang and S. Létourneau, "Learning to predict train wheel failures," in *Proc. 11th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Chicago, IL, USA, Aug. 2005, pp. 516–525.

[24] C. Yang and S. Letourneau, "Model evaluation for prognostics: Estimating cost saving for the end users," in *Proc. 6th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Cincinnati, OH, USA, Dec. 2007, pp. 304–309.

[25] C. Yang and S. Létourneau, "Two-stage classifications for improving time-to-failure estimates: A case study in prognostic of train wheels," *Int. J. Speech Technol.*, vol. 31, no. 3, pp. 255–266, Dec. 2009.

[26] T. Fawcett, "ROC graphs: Notes and practical considerations for data mining researchers," Intelligent Enterprise Technologies Laboratory, HP, Palo Alto, CA, USA, Tech. Rep. HPL-2003-4, 2003.

[27] C. Yang, W. Shen, and B. Gunay, "Toward machine learning-based prognostics for heating ventilation and air-conditioning systems," in *Proc. ASHRAE Winter Conf.*, Atlanta, GA, USA, Jan. 2019, pp. 106–115.
[28] M. Long, J. Wang, G. Ding, S. J. Pan, and P. S. Yu, "Adaptation regularization: A general framework for transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1076–1089, May 2014.
[29] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1410–1417.
[30] S. Pan, Q. Yang, and W. Fan, "Tutorial: Transfer learning with applications," presented at the 25th Int. Joint Conf. Artif. Intell. (IJCAI), 2013.

Chunsheng Yang received the B.Sc. degree (Hons.) in electronic engineering from Harbin Engineering University, Harbin, China, in 1983, the M.Sc. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 1986, and the Ph.D. degree from National Hiroshima University, Higashihiroshima, Japan, in 1995.

He is currently a Senior Research Officer with the National Research Council Canada, Ottawa, ON, Canada. He is an Adjunct Professor with Carleton University, Ottawa, ON, with the Nagaya Institute of Technology, Nagoya, Japan, and also with Shanghai Normal University, Shanghai. He is interested in data mining, machine learning/transfer learning, prognostic health management (PHM), reasoning technologies such as case-based reasoning, rule-based reasoning, and hybrid reasoning.

Burak Gunay received the B.A.Sc. degree in civil engineering from Bogazici University, Istanbul, Turkey, in 2009, and the M.A.Sc. and Ph.D. degrees in civil engineering from Carleton University, Ottawa, ON, Canada, in 2011 and 2016, respectively.

He is currently an Assistant Professor with the Department of Civil and Environmental Engineering, Carleton University. His research examines methods to optimize the operation of commercial buildings for comfort and energy efficiency by using the operational data gathered inside modern automation and control networks.

Zixiao Shi received the B.A.Sc. and M.A.Sc. degrees in civil engineering from Purdue University, West Lafayette, IN, USA, in 2011 and 2012, respectively, and the Ph.D. degree in civil engineering from Carleton University, Ottawa, ON, Canada, in 2018.

He is currently a Researcher with the National Research Council Canada, Ottawa, ON. His research interests include smart building technologies, fault detection and diagnosis for building systems, building energy model reduction, load forecasting, and data normalization and preprocessing.

Weiming Shen (Fellow, IEEE) received the bachelor's and master's degrees from Northern Jiaotong University, Beijing, China, in 1983 and 1986, respectively, and the Ph.D. degree from the University of Technology of Compiegne, Compiègne, France, in 1996.

He is currently a Professor with the Huazhong University of Science and Technology, Wuhan, China, and an Adjunct Professor with the University of Western Ontario, London, ON, Canada. His research interest includes intelligent software agents, wireless sensor networks, Internet of Things (IoT), big data, and their applications in industry.

Dr. Shen is a fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada. He is a Licensed Professional Engineer in Ontario, Canada.