

Fast and Accurate Classification of Time Series Data Using Extended ELM: Application in Fault Diagnosis of Air Handling Units

Ke Yan, Zhiwei Ji, Huijuan Lu, Jing Huang, Wen Shen, and Yu Xue, *Member, IEEE*

Abstract—The extreme learning machine (ELM) is famous for its single hidden-layer feed-forward neural network which results in much faster learning speed comparing with traditional machine learning techniques. Moreover, extensions of ELM achieve stable classification performances for imbalanced data. In this paper, we introduce a hybrid method combining the extended Kalman filter (EKF) with cost-sensitive dissimilar ELM (CS-D-ELM). The raw data are preprocessed by EKF to produce inputs for the CS-D-ELM classifier. Experimental results show that the proposed method is more suitable for real-time fault diagnosis of air handling units than traditional approaches.

Index Terms—Air handling unit (AHU), cost-sensitive dissimilar ELM (CS-D-ELM), extended Kalman filter (EKF), extreme learning machine (ELM), fault diagnosis.

I. INTRODUCTION

THE extreme learning machine (ELM), which was proposed by Huang *et al.* [1] in 2004, is reputed as a next-generation machine learning method utilizing a single hidden-layer feed-forward neural network. It generates a suitable number of nodes for the hidden layer, assigns random input weight, and forms the feature mappings in a one-step manner, which results in a generally faster algorithm comparing with traditional machine learning techniques, such as the support vector machine (SVM) and multilayer neural networks. Due to its fast learning speed, many extensions of the ELM are proposed to solve various classification problems.

Manuscript received September 21, 2016; revised January 10, 2017 and April 2, 2017; accepted April 2, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61602431 and Grant 61402417, in part by the Zhejiang Provincial Natural Science Foundation under Grant LY15F020037, in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions, and in part by the Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology. This paper was recommended by Associate Editor G.-B. Huang. (Ke Yan and Zhiwei Ji contributed equally to this work.) (Corresponding author: Huijuan Lu.)

K. Yan and H. Lu are with the College of Information Engineering, China Jiliang University, Hangzhou 310018, China (e-mail: hjlu@cjljlu.edu.cn).

Z. Ji is with the School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018, China.

J. Huang is with the CNRS/LTCI, Telecom ParisTech, 75013 Paris, France.

W. Shen is with the Department of Informatics, University of California Irvine, Irvine, CA 92697 USA.

Y. Xue is with the Nanjing University of Information Science and Technology, Nanjing 210016, China.

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

Digital Object Identifier 10.1109/TSMC.2017.2691774

Time series remote sensor data of HVAC systems, such as temperature, humidity, air flow rates, and electricity readings, are usually recorded by building management systems (BMSs) in constant time interval. The purposes of recording such data include analysis, proper maintenance, scheduling, fault detection, and diagnosis for the HVAC system. The energy consumed by HVAC systems across the world approximately occupies over 40% of the total building energy consumption [2], [3]. If the HVAC systems are under improper maintenance, the overall wasted energy can be huge. On one hand, immediate and fast fault diagnosis approaches is highly demanded for HVAC systems. On the other hand, the stored data in BMS usually contains a much larger chunk of normal conditional data compared to faulty conditional data, which leads to highly imbalanced multiclass datasets [4]. Moreover, the remote sensor recording usually involves noises [5]. Conventional machine learning techniques are hard to achieve stable classification accuracy for imbalanced noisy datasets [6].

The air handling unit (AHU) is one of the most extensively equipped components in an HVAC system. It absorbs the indoor air, mixes the indoor air with the outdoor air, conditions the mixed air with coiling or heating coils, and supplies it back to the indoor environment again with predefined temperature and humidity. Unfortunately, the AHU is also the most frequently faulty component in a HVAC system, provided that improper maintenance is applied.

In this paper, we propose a hybrid classification method to solve the fault diagnosis problem for AHUs, which is called EKF-CS-D-ELM. First, we introduce an extended Kalman filter (EKF) model to preprocess the measurement data. The purpose of the preprocessing is to filter out the noises, as well as stationarizing the time series data. Second, we introduce the cost-sensitive dissimilar ELM (CS-D-ELM) [7], to deal with the noisy and imbalanced data. The state vectors obtained from the first step are separated into training and testing datasets for multiclass classification, i.e., the diagnosis. The contributions of this paper are summarized as follows.

- 1) A Novel Hybrid Method for AHU Fault Diagnosis: We claim that the combination of EKF and ELM extensions is novel in the field of AHU fault diagnosis. Our previous work showed that the combination of statistical models and machine learning techniques achieved quality fault diagnosis results for AHUs [8]. In this paper, we replace the SVM with CS-D-ELM. The experimental results show that the CS-D-ELM provides more

accurate, fast, and robust diagnosis results compared to SVM (Section IV).

- 2) *Fast Learning Speed*: The diagnosis speed is a key factor for a large HVAC system to prevent further damage. The CS-D-ELM is an ensemble ELM which utilizes a voting system across a population of sub-ELMs. The overall learning speed depends on the individual learning speed of sub-ELMs and the number of parallelizable cores of the PC. In the experiment, we utilize a 32-core server machine over a voting system consisting of 100 sub-ELMs. The overall learning speed is around 3 to 4 times slower than each sub-ELM. We show that the overall learning speed is still much faster than SVM, which is crucial for real-time diagnosis of AHU faults.
- 3) *Reduced Misclassification Cost*: By embedding the cost-sensitive factors into the dissimilar ELM (D-ELM), we greatly reduced the misclassification costs for the AHU fault diagnosis. The reduction of misclassification cost balances the misclassification of false positive and false negative, which indirectly helps improve the overall classification accuracy.
- 4) *Improved Classification Accuracy*: We show the classification accuracy improvements by replacing the original SVM using CS-D-ELM. The classification accuracy comparison has been done over eight different classifiers based on three datasets. In overall, the proposed method provides the highest accuracy rates for all cases.

II. RELATED WORK

A. Extreme Learning Machine and Its Extensions

ELM first appeared as a single hidden layer feed-forward neural network; and then, many extensions of ELM were proposed to develop it to a more generalized form which may not be neural network alike [9], [10]. The advantages of ELM include global searching ability, one-step learning, fast learning speed, etc. [11]. The disadvantage for a single ELM is that the performance can be instable for large-scale, imbalanced, and noisy datasets [12]. To enhance the stability, efforts were made by integrating several ELMs to form the ensemble ELM methods. Zhai *et al.* [13] introduced a dynamic ensemble ELM method using the idea of Adaboost based on sample entropy. The experimental results showed that the dynamic ensemble ELM overcome the instability problem and was generally robust for various testing datasets. Li *et al.* [47] proposed another ensemble ELM named boosting weighted ELM that again used the Adaboost framework to integrate a number of ELMs. In addition, they considered the initial asymmetric weight assignment for the Adaboost to converge in a rapid speed. Experimental results showed that their method achieved higher classification accuracy rates than weighted ELM. Liu *et al.* [14] proposed a two-stage ELM to solve problems with large-scale data. Two ELMs were integrated with the spectral regression algorithm to reduce the high-dimensional data into low dimension and perform the classification on the reduced dataset. Their results showed that the two-stage ELM increased the scalability as well as the learning speed. Lian *et al.* [15] used ensemble ELM in time series data analysis. The application field for their study is in the prediction of

landslide displacement dynamics. Mirza *et al.* [16] introduced an ensemble method of subset online sequential ELM (ESOS-ELM) for solving the imbalanced drifting data problem. Cao *et al.* [17] proposed a hybrid classifier combining ELM with sparse representation classification. The resulting hybrid classifier showed better performance compared to each individual base classifier. Yang *et al.* [18], [19] introduced a multilayer ELM (ML-ELM) with subnetwork nodes and tested the proposed method on various datasets. Experimental results showed that the ML-ELM with subnetwork nodes outperformed other machine learning methods. Liu *et al.* [20] combined ELM with kernel sparse learning in the application field of tactile object recognition. Cao *et al.* [21] proposed a majority voting-based ensemble ELM (V-ELM) method. The V-ELM generally achieves higher classification accuracy than individual sub-ELM. Lu *et al.* [12] improved [21] by adding a dissimilarity factor to the V-ELM and named the new algorithm as D-ELM. More recently, we improved the D-ELM by adding cost-sensitive factors to increase the stability of the D-ELM [7]. The CS-D-ELM was successfully applied to the gene expression data classification problem which is well-known as a highly noisy, large-scale, and imbalanced data classification problem.

Embedding misclassification cost factors into ensemble ELMs increases both the stability and reliability of the classification result. In almost all application field, the costs for true negative and false negative are different; and there is also no exception in the case of AHU fault diagnosis. Our existing works showed a series of related study on cost-sensitive ELM for various application fields [7], [12], [22], [23]. Experimental results demonstrated that the embedding of misclassification cost effectively improved the classification accuracy for noisy and imbalanced data. Riccardi *et al.* [24] introduced a cost-sensitive ELM boosting method based on stagewise additive modeling using a multiclass exponential boosting algorithm. The experimental results showed better performance using their methods compared to existing ensemble ELM methods. Zhang and Zhang [25] proposed an evolutionary cost-sensitive ELM (ECSELM) and applied the ECSELM to various datasets. Experimental results showed 5% to 10% of accuracy improvement by adding cost-sensitive factors to the ELM and 10 times efficient improvement over cost-sensitive subspace learning methods.

B. Fault Diagnosis Approaches for AHUs

Since last century, statistical models, such as auto-regressive with exogenous variables (ARX) model and Kalman filter (KF) are utilized for fault diagnosis of HVAC components, such as AHUs. Both the ARX model and KF are statistical models which are capable to convert original sample data with noises to a stationarized time series parameters. Yoshida *et al.* [26] modeled the AHU data using KF. Fault types were recognized by setting different thresholds. Mulumba *et al.* [27] proposed a fault diagnosis technique combining KF and a rule-based system. The KF converts the original data samples into parameter vectors including lag and residual values. The rule-based system was used as thresholds to guard the normal

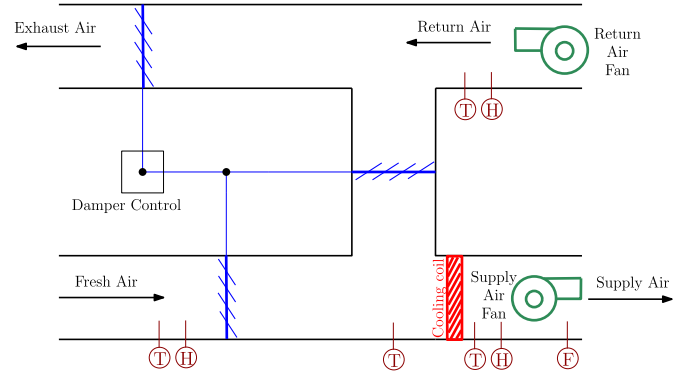
operational parameters of the AHU; any violation of the rules indicated typical faults in a predefined domain. EKF including the unscented KF is capable to capture the nonlinear nature of input data with stochastic uncertainties [28]. For example, Tudoroiu *et al.* [29] used an interactive multiple model augmented unscented KF for FDD of HVAC components including AHU.

Machine learning techniques, which is also known as data-driven approaches in the engineering field, represent another category of methods for fault diagnosis problem. Du *et al.* [30], [31] proposed a series of work which embeds the wavelet analysis into the neural network to find sensor faults in AHUs. The proposed wavelet neural network preprocessed the measurement data using wavelet analysis and later classified the data using neural networks. Zhu *et al.* [32] improved Du *et al.*'s works, and proposed a novel AHU fault diagnosis method, called neural network preprocessed by wavelet and fractal. Three-level wavelet analysis was applied to separate the measurement data; and the fractal analysis was employed to capture the fault characteristics. Comparing with Du *et al.*'s works, Zhu *et al.*'s work increases the diagnosis efficiency up to 15%. Dey and Dong [33] proposed a rule-based decision support system to diagnose faults of AHUs based on the Bayesian belief network. They performed the fault diagnosis practice on a real-world campus building, and showed that the proposed decision support system correctly recognized all possible faults. Cai *et al.* [34] introduced a real-time fault diagnosis method using object-oriented Bayesian networks.

More recently, hybrid methods combining statistical models and machine learning techniques attract wide attentions because of the high and stable classification accuracy dealing with imbalanced, noisy, and large scale data [35]–[37]. Juang [38] proposed a hybrid algorithm combining GA and particle swarm optimization for recurrent network design. Yan *et al.* [39] and Mulumba *et al.* [8] introduced a hybrid fault detection and diagnosis method for HVAC subsystems by integrating ARX model and SVM. The ARX model was utilized to preprocess the measurement data and convert the data into parameter vectors. In the classification phase, the SVM was used to enumerate different fault types. Experimental results showed that the hybrid method combining ARX model and SVM worked for both the AHU and chiller, which are in principal two most important subsystems of a HVAC system.

Compared to existing machine learning approaches for AHU fault diagnosis, the proposed CS-D-ELM method has the following three advantages.

- 1) It provides competitive diagnosis accuracy, which we show in the results section (Section IV).
- 2) ELM and its extensions are reported to have much faster learning speed compared to traditional neural networks and SVM. The fast learning speed is crucial for real-time fault diagnosis of AHUs.
- 3) False positive and false negative cost differently for AHU diagnosis in real-world applications. Most existing approaches did not consider this factor. In contrast, the CS-D-ELM differentiate these two cases and assign different misclassification costs to them.



T: Temperature sensor; H: Humidity sensor; F: Fan speed (air flow rate) sensor

Fig. 1. Structure of an AHU used in ASHRAE project 1312-RP, and the top important sensors.

III. HYBRID FAULT DIAGNOSIS METHOD FOR AHU

A. Measurement Data Description

The input measurement data is employed from American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) project 1312-RP entitled “Tools for evaluating fault detection and diagnostic methods for air-handling units” [40]–[42]. In the indicated ASHRAE project, two duplicate AHUs were used and operating under investigation for a certain period with the same settings. The two duplicate AHUs were named as AHU-A and AHU-B, where AHU-A simulated measurement data under normal operational conditions; and AHU-B simulated measurement data under various artificial faulty operational conditions. The inside structural view of the AHU is shown in Fig. 1. All data samples measured by remote sensors were recorded in a fixed time interval of one minute. We label different faulty types as F_1, F_2, \dots, F_n . Therefore, there are in total $n + 1$ labels including the normal condition *Normal*. Since the normal conditional data is obtained from AHU-A and all faulty conditional data is obtained from AHU-B, the number of data samples labeled with *Normal* is approximately n times more than the number of samples for each fault type.

B. Feature Selection of AHU Data

Feature selection is a necessary preprocessing step for machine learning techniques. In engineering problem, such as the AHU fault diagnosis, eliminating unnecessary attributes can also be beneficial which leads to a cut-off of the number of installing sensors.

In the original dataset provided by ASHRAE project 1312-RP, over one hundred features were recorded, including supply air temperature (T_{SA}), return air temperature (T_{RA}), etc. In this paper, we selected the most important attributes from the original dataset using ReliefF combining genetic algorithm [43]. The top four most important features are listed in Table I.

The values of T_{SA} and T_{RA} are captured by resistance temperature detectors that have the tolerance $\pm 0.3^\circ\text{C}$ under room temperature. Therefore, $\pm 0.3^\circ\text{C}$ is set as the threshold to differentiate the temperature readings at the same positions,

TABLE I
TOP FOUR MOST IMPORTANT FEATURES

Index	Feature	Description
1	Q_{coil}	energy consumption of cooling/heating coil
2	T_{SA}	Supply air temperature
3	T_{RA}	Return air temperature
4	H_{SA}	Supply air relative humidity

e.g., T_{SA} , using the same model of sensors for AHU-A and AHU-B. The H_{SA} values are captured by a humidity transmitter that has the accuracy $\pm 2\%$ under the circumstance of relative humidity less than 90%, and $\pm 3\%$ otherwise. The power is measured by a precision AC Watt transducer (model GW5) that has the accuracy within $\pm 0.2\%$ of the reading. The super heat protections of the compressors are enabled during the fault tests. For more system control settings details, readers may refer to the original ASHRAE 1312-RP report [40]–[42].

C. Extended Kalman Filter Model

EKF stationarizes the original measurement dataset, filters out noises and makes the samples more distinguishable for the classifiers by approximating measurements using mathematical model. It is proved to be useful for increasing the classification accuracy for machine learning techniques [44]. It is usually represented in a recursive manner and applied to nonlinearly co-related systems

$$X(t) = f[X(t-1)] + \sigma(t) \quad (1)$$

$$Y(t) = h[X(t)] + \nu(t) \quad (2)$$

where $X(t)$ represents the state vector predicted by the EKF model; $Y(t)$ is the actual measurement record with noises; $X(t-1)$ is the corresponding $X(t)$ value in one time step before; $\sigma(t)$, $\nu(t)$ are system noise and measurement noise.

In this paper, the measurement data of AHU is input into the EKF to produce a series of predicted state vectors $X(t)$. The approximated state vectors $X(t)$ are treated as inputs for the classifiers, and used in both training and testing phase of the classification process. Our previous works showed that the approximated state vectors are potentially more distinguishable than the raw data, and therefore produce better classification results compared to traditional methods [8].

D. Cost-Sensitive Dissimilar ELM

In this paper, we employed the CS-D-ELM [7] to produce the final classification results. The proposed classifier provides competitive classification accuracy rates compared to traditional approaches (shown in Section IV). Moreover, ELM and its extensions are famous for their fast learning speed, which can be crucial for real-time fault diagnosis. The CS-D-ELM also assigns different misclassification costs for false positive and false negative. In real-world applications, the false positive and false negative cost different damages; and the CS-D-ELM can balance the two to minimize the overall misclassification cost.

1) *Dissimilar ELM*: The D-ELM algorithm proposed in our previous work [12] is an ensemble ELM, which is

demonstrated to be suitable for imbalanced data classification. It provides more stable classification results compared to the single ELM and V-ELM, since it removes individual ELMs from the ensemble pool based on the dissimilarity measure. Suppose there are N sub-ELMs and M training samples, the dissimilarity of the i th sub-ELM and j th sub-ELM for a specific sample k is defined by $Dif(f_{ik}, f_{jk})$; and the overall dissimilarity value is $Dis_{i,j} = \sum_{k=1}^N Dif(f_{ik}, f_{jk})$. For each i th sub-ELM, the overall dissimilarity value is $\eta_i = \sum_{j=1}^N Dis_{i,j}$.

Suppose the average classification accuracy is \bar{p} , the construction process of D-ELM can be concluded in three steps.

- 1) Eliminate unwanted sub-ELMs and reduce the total number of sub-ELMs to \tilde{N} according to two rules: sub-ELMs are eliminated with small η and $\bar{p} < 0.5$; and sub-ELMs are eliminated with big η and $\bar{p} > 0.5$.
- 2) Train the remaining \tilde{N} again using the same node number in the single hidden layer for each sub-ELM.
- 3) Take the voting process and for each classification of sample x with class i based on \tilde{N} sub-ELMs, the weight of that class $W(i)$ is increased by 1. The final probability of sample x belonging to the class i is calculated by

$$P(i|x) = \frac{W(i)}{\tilde{N}}, i \in \{c_1, c_2, \dots, c_m\}$$

where m is the number of classes.

The D-ELM algorithm is essentially still a majority voting ensemble algorithm that is built based on the V-ELM algorithm proposed by Cao *et al.* [21] in 2012. The V-ELM provides more stable and accurate classification results compared to the single ELM algorithm. However, we found that there are similar/redundant ELMs in the ensemble pool, which makes the ensemble process less significant. In the D-ELM algorithm, we remove the redundant ELMs by evaluating the dissimilarity measure and classification accuracy. According to the experimental results, the D-ELM is found to be more stable in terms of generalization performance than the V-ELM [12].

2) *Adding Misclassification Cost to Dissimilar ELM*: The CS-D-ELM algorithm added the misclassification cost to the D-ELM algorithm. In real-world AHU fault diagnosis scenarios, the damage of a false negative case is much higher than a false positive case. Therefore, it is demanded to take the misclassification cost into consideration for false diagnosis of AHUs. In CS-D-ELM, the goal of D-ELM becomes calculating

$$\arg \min \{R(i|x)\} = \arg \min \left\{ \sum_j P(j|x) \cdot C(i, j) \right\}$$

where $\{R(i|x)\}$ is the risk that x is classified to class i ; and $C(i, j)$ represents the risk that a sample belongs to class j is misclassified to class i .

The final probability for sample x belonging to the class i becomes

$$P(i|x) = \frac{R(i|x)}{\tilde{N}}, i \in \{c_1, c_2, \dots, c_m\}$$

where m is the number of classes.

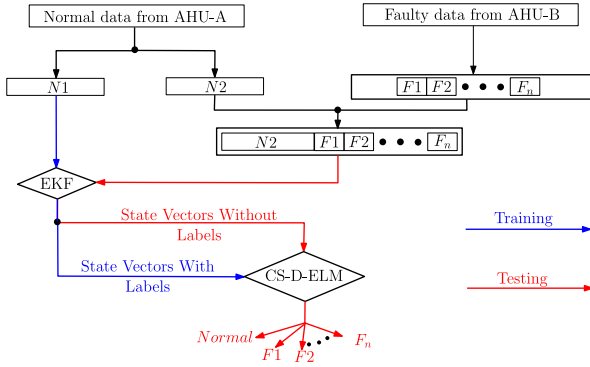


Fig. 2. Overall flowchart of the proposed EKF-CS-D-ELM algorithm.

TABLE II
FAULT DIAGNOSIS ACCURACY RATES OF VARIOUS FAULTS IN THE SUMMER DATASET USING DIFFERENT CLASSIFIERS

	Acc. Bay. (%)	Net.	MLP	J48	SVM	ELM	D-ELM	CS-D-ELM	Our meth.
F1	42.7	21.8	64.2	73.3	63.6	77.5	79.6	91.3	
F2	50.3	25.4	76.0	75.8	67.6	81.1	84.8	95.2	
F3	42.3	29.7	78.0	67.3	53.4	68.3	71.4	84.1	
F4	23.9	15.1	72.2	71.3	65.2	74.5	79.3	93.1	
F5	31.3	25.3	65.2	73.9	67.3	78.1	80.4	94.8	
F6	29.7	26.4	71.2	77.7	72.1	80.3	85.3	91.3	
F7	27.5	31.1	75.9	80.4	75.3	81.3	84.2	96.4	
F8	40.3	28.8	65.4	78.7	67.4	75.0	86.4	95.9	
F9	47.3	35.1	66.1	81.2	70.8	82.4	89.4	96.7	
F10	57.0	37.2	83.0	84.7	75.3	86.8	92.5	99.1	
F11	35.2	15.8	68.9	83.5	74.2	84.3	86.9	92.4	
F12	25.2	13.5	61.4	83.9	78.0	73.2	80.3	94.4	
F13	32.7	26.4	64.0	81.5	78.8	88.4	84.6	97.6	

The CS-D-ELM intends to balance the misclassification cost for false positive cases and false negative cases. In other words, it minimizes the overall misclassification cost, which potentially increases the classification stability as well as the classification accuracy.

IV. EXPERIMENT AND RESULTS

A. Experiment Setup

Three datasets are tested, namely summer (dated from August 19, 2007 to September 10, 2008), spring (dated from May 02, 2008 to June 01, 2008), and winter (dated from January 29, 2008 to February 17, 2008), which were collected by ASHRAE project number 1312-RP at Philadelphia, USA. The summer dataset contains 13 faulty subsets; the spring dataset contains 19 faults; and the winter dataset includes 10 faults. The detailed descriptions of all faults are listed in Tables VII–IX in the Appendix. Each faulty subset generated by AHU-B consists of 1440 data samples (1 day, 1 min per sample), and is accompanied by a corresponding normal dataset that is generated by AHU-A. Therefore, there are in total 37440 data samples for the summer dataset, 54720 data samples for the spring dataset, and 28800 data samples for the winter dataset.

The first phase of the experiment is to rearrange the original measurement data. The normal conditional data collected by AHU-A is evenly divided into two sub-datasets: $N1$ and

TABLE III
CLASSIFICATION ACCURACY RATES FOR THE SPRING DATASET

	Acc. Bay. (%)	Net.	MLP	J48	SVM	ELM	D-ELM	CS-D-ELM	Our meth.
F1	39.0	49.5	84.7	82.1	73.0	83.9	85.3	90.4	
F2	53.5	49.2	87.6	83.7	75.1	84.1	85.9	92.7	
F3	60.0	66.8	90.3	88.3	80.7	89.3	90.2	95.3	
F4	48.2	74.7	86.8	85.6	79.5	85.1	87.3	92.8	
F5	58.9	52.4	93.1	84.3	75.2	86.5	88.0	92.1	
F6	47.3	64.6	87.9	77.6	68.8	78.3	79.0	88.8	
F7	59.2	62.9	88.7	78.8	70.4	81.3	82.5	89.9	
F8	73.9	76.9	94.2	92.0	84.1	89.8	91.7	96.3	
F9	37.7	26.9	80.1	78.5	73.4	80.1	83.3	87.6	
F10	57.1	64.3	88.9	86.3	80.0	87.9	89.1	93.5	
F11	46.7	37.2	85.0	78.7	70.6	78.4	79.6	87.5	
F12	82.9	54.8	91.7	88.0	81.0	89.1	90.1	97.4	
F13	55.6	74.3	81.3	83.9	77.5	85.2	86.9	91.8	
F14	52.4	59.6	87.5	81.8	75.4	82.1	83.0	89.0	
F15	60.0	53.2	86.8	88.1	80.5	89.6	90.7	96.4	
F16	58.2	63.1	89.8	76.1	67.9	77.5	79.5	90.6	
F17	45.0	42.0	81.1	73.2	67.1	72.9	73.9	89.6	
F18	53.5	36.1	86.0	85.2	79.6	86.1	87.2	93.3	
F19	57.6	52.8	89.7	87.9	80.3	88.6	89.7	95.7	

TABLE IV
CLASSIFICATION ACCURACY RATES FOR THE WINTER DATASET

	Acc. Bay. (%)	Net.	MLP	J48	SVM	ELM	D-ELM	CS-D-ELM	Our meth.
F1	82.5	61.7	94.7	89.5	81.4	90.4	91.2	96.4	
F2	68.2	54.8	90.4	88.8	79.3	88.2	89.7	95.1	
F3	74.4	63.5	93.3	87.5	79.6	89.6	91.5	94.8	
F4	85.7	88.0	88.2	83.9	74.9	85.3	87.0	91.3	
F5	79.9	68.6	94.9	90.6	83.6	92.4	93.8	96.6	
F6	79.6	73.5	96.2	95.9	87.5	96.0	96.3	99.5	
F7	72.8	54.1	91.3	77.5	71.5	80.5	82.2	92.8	
F8	70.9	57.7	92.0	84.6	75.0	86.2	88.4	93.2	
F9	62.5	52.7	86.0	88.0	81.7	88.3	89.6	93.0	
F10	61.4	57.3	89.0	88.2	82.2	88.9	90.1	94.4	

$N2$. The first sub-dataset $N1$ is used to construct the EKF model and $N2$ joins the faulty dataset for the second part of the experiment. The data order of the joint dataset is $N2, F1, F2, \dots, F_n$.

In the second phase, the actual measurement of selected features are converted to state variables by the EKF model. The converted steady state vector set is again evenly divided into two subsets. The first subset with labels is used as training dataset for classifiers. The second subset after removing the labels is used as testing dataset for evaluation. For each of the original subdatasets: $N2, F1, F2, \dots, F_n$, half of the converted state vectors belong to the training dataset and the rest converted state vectors belong to the testing dataset. In addition, as we mentioned in Section III-D, the misclassification costs for false positive and false negative are different. We assign the misclassification cost for false positive to be $C(1, -1) = 1$, and false negative to be $C(1, -1) = 5$.

The overall flowchart of the experiment setup is shown in Fig. 2, and it is also viewed as an overall flowchart for the proposed EKF-CS-D-ELM algorithm.

B. Results

We compare the classification results of the proposed hybrid method with traditional approaches, including Bayesian

TABLE V
COMPARISON OF AVERAGE MISCLASSIFICATION COST FOR
EACH DATASET USING DIFFERENT CLASSIFIERS

Ave. cost	Bay. Net.	MLP	J48	SVM	ELM	D-ELM	CS-D-ELM	Our meth.
Sum.	1.97	2.41	1.13	0.89	1.07	0.70	0.33	0.16
Spr.	1.53	1.57	0.57	0.73	0.92	0.68	0.30	0.15
Win.	0.95	1.02	0.49	0.58	0.63	0.57	0.19	0.10

TABLE VI
TRAINING TIME COMPARISON FOR DIFFERENT CLASSIFIERS

Trai. Time (sec.)	Bay. Net.	MLP	J48	SVM	ELM	D-ELM	CS-D-ELM	Our meth.
Sum.	0.08	25.7	1.1	7.3	0.20	0.51	0.55	0.58
Spr.	0.12	38.1	0.73	8.5	0.33	0.68	0.69	0.73
Win.	0.04	18.6	0.57	3.1	0.17	0.35	0.37	0.43

network, multilayer perceptron (MLP), J48 decision tree, SVM, ELM, D-ELM, and CS-D-ELM. The first three approaches are tested using WEKA [45]; SVM is tested using LibSVM [46]; and all ELM approaches are implemented in MATLAB. All tests are performed on a 32-core DELL server machine with 128 GB RAM. The number of sub-ELM for D-ELM and CS-D-ELM is 100. All accuracy rates are averaged over 30 times of repeated runs.

The classification accuracy rates of different faults in the summer, spring, and winter datasets are shown in Tables II–IV, respectively. The results show that: 1) the CS-D-ELM provides stable classification results for imbalanced data, such as the AHU fault diagnosis datasets and 2) by adding the EKF model in the preprocessing phase, the classification accuracy rates can be improved approximately from 3% to 15%.

The embedding of misclassification cost in CS-D-ELM reduces the misclassification cost, and therefore balances the misclassification rate for false positive and false negative. In other words, the classification accuracy for faulty types increases. We show the average misclassification cost for each dataset using different classifiers in Table V. The proposed method demonstrates a large improvement on the average misclassification cost for AHU fault diagnosis, which may lead to huge energy savings in the real-world applications.

It is noted that the misclassification cost improvement highly depends on the predefined costs for false positive and false negative. In our case, we set the misclassification cost of false negative to be the value of 5 times the cost of false positive. If the ratio increases, the misclassification cost improvement increases as well.

Last, we show the training time comparison of the eight algorithms (Table VI). Although the CS-D-ELM is an ensemble ELM algorithm, which is generally slower than the ELM algorithm, it is still much faster than the SVM. Therefore, in overall, our method is more suitable for real-time fault diagnosis for AHU than SVM.

V. CONCLUSION

In this paper, we propose a novel hybrid method to diagnose faults in AHUs, which is named as EKF-CS-D-ELM. The proposed method is demonstrated to be a fast and accurate

TABLE VII
DESCRIPTIONS OF ALL FAULTS IN THE SUMMER DATASETS

Index	Fault Description
F1	Exhausted air (EA) damper stuck (fully open)
F2	EA damper stuck (fully close)
F3	Return fan at fixed speed (30%spd)
F4	Return Fan complete failure
F5	Cooling coil valve control unstable
F6	Outdoor air (OA) damper stuck (fully closed)
F7	Cooling coil valve stuck (fully closed)
F8	Cooling coil valve stuck (fully open)
F9	Cooling coil valve stuck (partially open - 15%)
F10	Cooling coil valve stuck (partially open - 65%)
F11	OA damper leak (55% open)
F12	AHU duct leaking (after supply fan (SF))
F13	AHU duct leaking (before SF)

TABLE VIII
DESCRIPTIONS OF ALL FAULTS IN THE SPRING DATASETS

Index	Fault Description
F1	OA temperature sensor bias (+3F)
F2	OA temperature sensor bias (-3F)
F3	OA damper stuck (fully close)
F4	OA damper stuck (40% open)
F5	EA damper stuck (fully open)
F6	EA damper stuck (fully close)
F7	EA damper stuck (40% open)
F8	Cooling coil valve stuck (fully closed)
F9	Cooling coil valve stuck (fully open)
F10	Cooling coil valve stuck (partially open 50%)
F11	Return fan complete failure
F12	Return Fan at fixed speed 20%spd
F13	Return Fan at fixed speed 80%spd
F14	Air filter area block fault (10%)
F15	Air filter area block fault (25%)
F16	Mixed air damper unstable
F17	Cooling coil control unstable
F18	Sequence of heating and cooling unstable
F19	Supply fan control unstable

classification method to be applied in the field of fault diagnosis of AHUs. In a real-world fault diagnosis scenario, the measurement dataset is always imbalanced, where the number of normal conditional data is usually more than the number of faulty conditional data. Direct application of conventional classifiers results in low classification accuracy and high cost for not identifying the faults (false negatives). We show that: 1) the original noisy and imbalanced dataset can be stationaryized by an EKF model and 2) stable and accurate classification results can be achieved efficiently using CS-D-ELM.

The proposed EKF-CS-D-ELM method is a novel hybrid approach for AHU diagnosis, extending our previous work [8], where we replace the ARX model by EKF and replace the SVM by CS-D-ELM. In the experiment section, first, we show the effectiveness by adding an EKF model for preprocessing of the raw data before the classification process. Second, we show the highest overall classification accuracy using our method. Third, we balance the misclassification rates for false positive and false negative, which results in much lower misclassification cost. Last, we show that the proposed method is more suitable for real-time fault diagnosis than SVM because of the faster training speed.

The limitation of this paper is that faults can only be recognized with training samples available, which may not be the

TABLE IX
DESCRIPTIONS OF ALL FAULTS IN THE WINTER DATASETS

Index	Fault Description
F1	OA damper stuck (fully close)
F2	OA damper leaking (52% open)
F3	OA damper leaking (62% open)
F4	EA damper stuck (fully open)
F5	EA damper stuck (fully close)
F6	Heating coil fouling stage 1
F7	Heating coil fouling stage 2
F8	Heating coil reduced capacity stage 1
F9	Heating coil reduced capacity stage 2
F10	Heating coil reduced capacity stage 3

TABLE X
NOMENCLATURE

Var.	Description
t	a time instance
$X(t)$	a state vector of the EKF model
f	a non-linear function
σ	system noise
Y	an actual measurement
h	a non-linear function
ν	measurement noise
Dif	dissimilarity measurement
Dis	overall dissimilarity measurement
η	overall dissimilarity measurement
\bar{p}	average classification accuracy
\tilde{N}	number of sub-ELMs
$W(i)$	weight of class i
$P(i x)$	probability of sample x belonging to the class i
c	a class
$R(i x)$	risk that x is classified to class i
$C(i, j)$	risk that a sample belongs to class j is misclassified to class i
$N1$	a subset of the normal dataset
$N2$	a subset of the normal dataset
F_n	a fault type indicating a subset of the fault dataset
Q_{coil}	energy consumption of cooling/heating coil
T_{SA}	Supply air temperature
T_{RA}	Return air temperature
H_{SA}	Supply air relative humidity

case in real-world applications. As a future work, we intend to identify new types of faults using existing faulty type samples as training data. Once the new faulty type samples are confidently identified, the training pool can be iteratively enhanced using semi-supervised learning approaches.

APPENDIX

See Tables VII–X.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Wen and X. Li for the measurement dataset supply from ASHRAE project report No. 1312-RP. They would like to thank Prof. A. Afshari from the Masdar Institute of Science and Technology for his support on the early stage of this project.

REFERENCES

- [1] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, Budapest, Hungary, 2004, pp. 985–990.
- [2] A. M. Omer, "Energy, environment and sustainable development," *Renew. Sustain. Energy Rev.*, vol. 12, no. 9, pp. 2265–2300, 2008.
- [3] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy Build.*, vol. 40, no. 3, pp. 394–398, 2008.
- [4] S. Wang and X. Yao, "Multiclass imbalance problems: Analysis and potential solutions," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, pp. 1119–1130, Aug. 2012.
- [5] S. Wang and J. Cui, "Sensor-fault detection, diagnosis and estimation for centrifugal chiller systems using principal-component analysis method," *Appl. Energy*, vol. 82, no. 3, pp. 197–213, 2005.
- [6] T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "Comparing boosting and bagging techniques with noisy and imbalanced data," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 3, pp. 552–568, May 2011.
- [7] Y. Liu, H. Lu, K. Yan, H. Xia, and C. An, "Applying cost-sensitive extreme learning machine and dissimilarity integration to gene expression data classification," *Comput. Intell. Neurosci.*, vol. 2016, Aug. 2016, Art. no. 19.
- [8] T. Mulumba, A. Afshari, K. Yan, W. Shen, and L. K. Norford, "Robust model-based fault diagnosis for air handling units," *Energy Build.*, vol. 86, pp. 698–707, Jan. 2015.
- [9] G.-B. Huang and L. Chen, "Convex incremental extreme learning machine," *Neurocomputing*, vol. 70, nos. 16–18, pp. 3056–3062, 2007.
- [10] G.-B. Huang and L. Chen, "Enhanced random search based incremental extreme learning machine," *Neurocomputing*, vol. 71, no. 16, pp. 3460–3468, 2008.
- [11] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.
- [12] H.-J. Lu, C.-L. An, E.-H. Zheng, and Y. Lu, "Dissimilarity based ensemble of extreme learning machine for gene expression data classification," *Neurocomputing*, vol. 128, pp. 22–30, Mar. 2014.
- [13] J.-H. Zhai, H.-Y. Xu, and X.-Z. Wang, "Dynamic ensemble extreme learning machine based on sample entropy," *Soft Comput.*, vol. 16, no. 9, pp. 1493–1502, 2012.
- [14] P. Liu, Y. Huang, L. Meng, S. Gong, and G. Zhang, "Two-stage extreme learning machine for high-dimensional data," *Int. J. Mach. Learn. Cybern.*, vol. 7, no. 5, pp. 765–772, 2016.
- [15] C. Lian, Z. Zeng, W. Yao, and H. Tang, "Ensemble of extreme learning machine for landslide displacement prediction based on time series analysis," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 99–107, 2014.
- [16] B. Mirza, Z. Lin, and N. Liu, "Ensemble of subset online sequential extreme learning machine for class imbalance and concept drift," *Neurocomputing*, vol. 149, pp. 316–329, Feb. 2015.
- [17] J. Cao, K. Zhang, M. Luo, C. Yin, and X. Lai, "Extreme learning machine and adaptive sparse representation for image classification," *Neural Netw.*, vol. 81, pp. 91–102, Sep. 2016.
- [18] Y. Yang and Q. M. J. Wu, "Multilayer extreme learning machine with subnetwork nodes for representation learning," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2570–2583, Nov. 2016.
- [19] Y. Yang, Q. M. J. Wu, and Y. Wang, "Autoencoder with invertible functions for dimension reduction and image reconstruction," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published.
- [20] H. Liu, J. Qin, F. Sun, and D. Guo, "Extreme kernel sparse learning for tactile object recognition," *IEEE Trans. Cybern.*, to be published.
- [21] J. Cao, Z. Lin, G.-B. Huang, and N. Liu, "Voting based extreme learning machine," *Inf. Sci.*, vol. 185, no. 1, pp. 66–77, 2012.
- [22] H. Lu, *A Study of Tumor Classification Algorithms Using Gene Expression Data*, China Min. Univ., Xuzhou, China, 2012.
- [23] H. Lu, W. Chen, X. Ma, and L. Yi, "A dataset splitting based neural network ensemble method for tumor classification," *Int. J. Digit. Content Technol. Appl.*, vol. 6, no. 5, pp. 167–173, 2012.
- [24] A. Riccardi, F. Fernández-Navarro, and S. Carloni, "Cost-sensitive adaboost algorithm for ordinal regression based on extreme learning machine," *IEEE Trans. Cybern.*, vol. 44, no. 10, pp. 1898–1909, Oct. 2014.
- [25] L. Zhang and D. Zhang, "Evolutionary cost-sensitive extreme learning machine," *IEEE Transactions on Neural Networks and Learning Systems*, to be published. doi: 10.1109/TNNLS.2016.2607757
- [26] H. Yoshida, T. Iwami, H. Yuzawa, and M. Suzuki, "Typical faults of air conditioning systems and fault detection by ARX model and extended Kalman filter," presented at the *Winter Meeting Amer. Soc. Heating Refrigeration Air Conditioning Eng.*, vol. 102, Atlanta, GA, USA, Feb. 17–21, 1996, paper CONF-960254.

- [27] T. Mulumba, A. Afshari, and L. Friedrich, "Kalman filter-based fault detection and diagnosis for air handling units," in *Proc. 15th Int. Refrigeration Air Conditioning Conf.*, West Lafayette, IN, USA, 2014, pp. 1332–1340.
- [28] X. Kai, C. Wei, and L. Liu, "Robust extended Kalman filtering for nonlinear systems with stochastic uncertainties," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 2, pp. 399–405, Mar. 2010.
- [29] N. Tudoroiu, M. Zaheeruddin, E.-R. Tudoroiu, and V. Jeflea, "Fault detection and diagnosis (FDD) in heating ventilation air conditioning systems (HVAC) using an interactive multiple model augmented unscented Kalman filter (IMMAUKF)," in *Proc. Conf. Human Syst. Interact.*, Kraków, Poland, 2008, pp. 334–339.
- [30] Z. Du, X. Jin, and Y. Yang, "Wavelet neural network-based fault diagnosis in air-handling units," *HVAC R Res.*, vol. 14, no. 6, pp. 959–973, 2008.
- [31] Z. Du, X. Jin, and Y. Yang, "Fault diagnosis for temperature, flow rate and pressure sensors in VAV systems using wavelet neural network," *Appl. energy*, vol. 86, no. 9, pp. 1624–1631, 2009.
- [32] Y. Zhu, X. Jin, and Z. Du, "Fault diagnosis for sensors in air handling unit based on neural network pre-processed by wavelet and fractal," *Energy Build.*, vol. 44, pp. 7–16, Jan. 2012.
- [33] D. Dey and B. Dong, "A probabilistic approach to diagnose faults of air handling units in buildings," *Energy Build.*, vol. 130, pp. 177–187, Oct. 2016.
- [34] B. Cai, H. Liu, and M. Xie, "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks," *Mech. Syst. Signal Process.*, vol. 80, pp. 31–44, Dec. 2016.
- [35] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012.
- [36] Z. Ji, B. Wang, S. Deng, and Z. You, "Predicting dynamic deformation of retaining structure by LSSVR-based time series method," *Neurocomputing*, vol. 137, pp. 165–172, Aug. 2014.
- [37] Q. Hui and H. Zhang, "Optimal balanced coordinated network resource allocation using swarm optimization," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 5, pp. 770–787, May 2015.
- [38] C.-F. Juang, "A hybrid of genetic algorithm and particle swarm optimization for recurrent network design," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 2, pp. 997–1006, Apr. 2004.
- [39] K. Yan, W. Shen, T. Mulumba, and A. Afshari, "ARX model based fault detection and diagnosis for chillers using support vector machines," *Energy Build.*, vol. 81, pp. 287–295, Oct. 2014.
- [40] S. Li and J. Wen, "Development and validation of a dynamic air handling unit model—Part I," *ASHRAE Trans.*, vol. 116, no. 1, pp. 45–56, 2010.
- [41] S. Li, J. Wen, X. Zhou, and C. J. Klaassen, "Development and validation of a dynamic air handling unit model—Part II," *ASHRAE Trans.*, vol. 116, no. 1, pp. 57–73, 2010.
- [42] J. Wen and S. Li, "Tools for evaluating fault detection and diagnostic methods for air-handling units," *ASHRAE*, Atlanta, GA, USA, Tech. Rep. 1312, 2011.
- [43] K. Yan, Z. Ji, and W. Shen, "Online fault detection methods for chillers combining extended Kalman filter and recursive one-class SVM," *Neurocomputing*, vol. 228, pp. 205–212, Mar. 2017.
- [44] S. S. Haykin *et al.*, *Kalman Filtering and Neural Networks*. New York, NY, USA: Wiley, 2001.
- [45] M. Hall *et al.*, "The WEKA data mining software: An update," *ACM SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009.
- [46] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 1–27, 2011.
- [47] K. Li, X. Kong, Z. Lu, L. Wenyin, and J. Yin, "Boosting weighted ELM for imbalanced learning," *Neurocomputing*, vol. 128, pp. 15–21, Mar. 2014.



Zhiwei Ji received the bachelor's degree from Zhejiang A&F University (ZAFU), Hangzhou, China, in 2003, the master's degree from Shanghai University, Shanghai, China, in 2009, and the Ph.D. degree from the School of Electronics and Information Engineering, Tongji University, Shanghai, in 2016.

Since 2003, he has been a Lecturer with the School of Information Engineering, ZAFU. His current research interests include pattern recognition, machine learning, computational biology, and bioinformatics.



Huijuan Lu received the Ph.D. degree from the China University of Mining and Technology, Xuzhou, China, in 2012.

She is a Professor with China Jiliang University, Hangzhou, China. She is an Outstanding Member of China Computer Federation. She has published over 80 papers, received two National Natural Science Foundation of China funds and conducted eight Zhejiang provincial science and technology projects. Her current research interests include machine learning, pattern recognition, and bioinformatics.



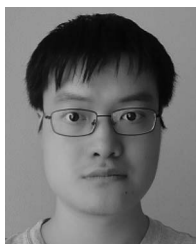
Jing Huang received the master's degree from the University of Paris Descartes, Paris, France, in 2009, and the Ph.D. degree in computer graphics at TELECOM ParisTech, Paris, in 2013.

He is a Researcher with the CNRS, LTCI, TELECOM ParisTech, Paris, France. His current research interests include interactive rendering and animation generation with parallel high performance computing and learning and modeling expressive gestures in interactive communicative systems.



Wen Shen received the B.Eng. degree from the Masdar Institute of Science and Technology, Abu Dhabi, UAE, and the M.Sc. degree from Northwestern Polytechnical University, Xi'an, China. He is currently pursuing the Ph.D. degree with the Department of Informatics, University of California at Irvine, Irvine, CA, USA.

His current research interests include multiagent systems, human-machine interaction, and game theory.



Ke Yan received the bachelor's degree and the Ph.D. degree in computer science in 2012 under the supervision of Dr. H.-L. Cheng from the National University of Singapore, Singapore.

From 2013 to 2014, he was a Post-Doctoral Researcher with the Masdar Institute of Science and Technology, Abu Dhabi, UAE. He currently serves as a Lecturer with China Jiliang University, Hangzhou, China. His current research interests include computer graphics, computational geometry, data mining, and machine learning.



Yu Xue (M'14) was born in 1981. He received the Ph.D. degree from the College of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing, China, in 2013.

He is a Lecturer with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China. He has published nearly 20 journal and conference papers. His current research interests include computational intelligence, Internet of Things, and electronic countermeasures.

Dr. Xue is a member of ACM and CCF.