A Case Study Based Approach for Remote Fault Detection Using Multi-Level Machine Learning in Smart Building

Maitreyee Dey[1], Soumya Prakash Rana[1], Sandra Dudley[1]
[1]School of Engineering, London South Bank University, 103 Borough road London, UK
Corresponding Author Email: deym@lsbu.ac.uk, dey.maitreyee@gmail.com

Abstract:

Due to the increased awareness of issues ranging from green initiatives, sustainability and occupant well-being buildings are becoming smarter, but with smart requirements come increasing complexity and monitoring, ultimately carried out by humans. Building heating ventilation and air-conditioning (HVAC) units are one of the major units that consume large percentages of a building's energy, for example through their involvement in space heating and cooling, the greatest energy consumption in buildings. By monitoring such components effectively entire energy demand in buildings can be substantially decreased. Due to the complex nature of building management systems (BMS), many simultaneous anomalous behaviour warnings are not manageable in timely manner, thus many energy related problems are left unmanaged which causes unnecessary energy wastage and deteriorate equipment's lifespan. This study proposes a machine learning based multi-level automatic fault detection system (MLe-AFD) focusing on remote HVAC fan coil unit (FCU) behaviour analysis. The proposed method employs sequential two-stage clustering to identify abnormal behaviour of FCU. The model's performance has been validated by implementing well-known statistical measures and further cross-validated via expert building engineering knowledge. The method has been experimented on commercial building based in central London, UK as a case study and allows to remotely identify three types of FCU faults appropriately and inform building management staff proactively when it occurs; this way the energy expenditure can be further optimized.

Keywords:

Smart building, Fan coil unit, Fault detection, Multi-level clustering, Statistical validation.

---

1. Introduction:

With the increasing demands of smart building infrastructure and plant maintenance, automatic fault detection has gained attention in both academic and industrial fields [1]. There is a growing importance placed upon the development and execution of smart grids and smart buildings in order to encounter electricity demands and building material sustainability in an efficient and cost-effective manner whilst minimizing CO_2 emissions, which account for around three-quarters of total greenhouse gas emissions [2]. Commercial buildings are responsible worldwide for approximately 41% of primary energy consumption including United States, Europe, and Asia, however experts anticipate that will rise over the next 20 years [3]. Improved demand and control strategies require incorporation into the existing infrastructure to sustain the collective electricity demand of commercial buildings. One of the feasible and well accepted moves is to extract the information from historic electricity consumption data of different units and identify the causes that widely effect the demand/supply scale. This efficient functioning of such systems is expected to improve the economy and deliver sustainable solutions for energy production in smart buildings [4, 5]. A significant amount of energy is misused by malfunctioning or poorly maintained building units and often building operators are unaware that units are malfunctioning and wasting energy. A building's energy consumption is a complex system which comprises several elements such as heating ventilation and air-conditioning (HVAC) unit, lighting, elevation, security, etc. Figure 1 shows a pie chart which depicts the different areas that are responsible for extensive energy consumption within a building. It has been found that the HVAC consumes approximately 41% energy of the building's total energy consumption, whereas lighting is the second highest energy consumable unit, which expends around 29% of the total energy. Subsequently, water heating, office equipment, and others consumes 9%, 13%, and 8% respectively of the building's total energy [7]. Faults relating to HVAC systems represent between 1% and 2.5% of total commercial building consumption [6]. Multi-agent based systems depends on the specific areas of the buildings such as, demand response, human behaviour and it has great effect on energy optimisation [35]. Typical building performance monitoring and fault identification are performed by building experts which is a slow process and leaves many problems undetected or worse, ignored. The efficient integration of automatic and remote fault detection methodologies able to detect faults immediately when they occur would be a game changer. Also, it is communicating the fault to the owner or maintenance personnel with an agreed simple language describing the fault, if they are of enough severity, is highly desirable. This system pipeline would eliminate the scheduled maintenance costs, reduce diagnostic labour, reduce wasted energy, reduce peak electricity demand, and minimize downtime.

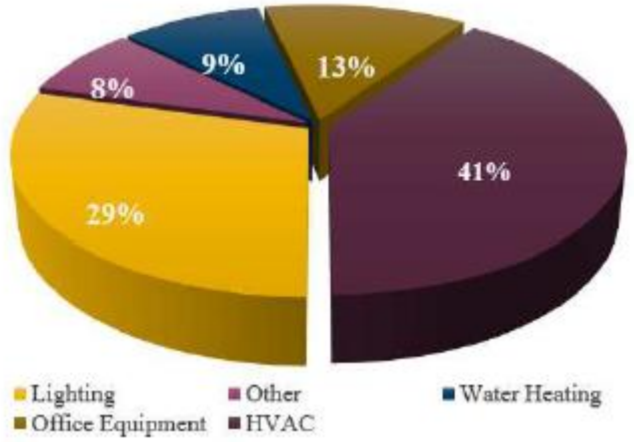


Figure 1: Energy consumption by different HVAC units

The paper is outlined as follows; Section 2 presents the literature review on published HVAC fault detection methods. Section 3 overviews the structural details of FCU units and their associated faults. Section 4 illustrates the proposed technical framework and methodology that have been developed to deal with real building problems. The outcomes of the proposed method are then detailed in Section 5. Finally, the conclusions and impact of this study are discussed in Section 6.

2. Literature review:

Automatic fault detection and diagnosis methodologies for HVAC systems have evolved with notable advancements implementing data mining and machine learning (ML) techniques. However, practical limitations, such as scalability and complexity of HVAC systems have made fault detection extremely challenging since the beginning of dynamic research and exploration in 1980's [8, 9]. Fault detection and diagnosis (FDD) research is classified into quantitative models, process history, and rule based groups, as shown in Figure X. This proposed FDD study focusses on employing machine learning techniques to improve building performance which is a part of process history based FDD which is categorised into two further groups: knowledge based and data-driven based [new ref- Kim et.al-34.] Knowledge based methods require vast amounts of prior information to process the data whereas the data-driven model precludes the need of any prior information, but discovers this information buried within the data itself. In this study, data-driven based automatic fault detection (AFD) has been performed on historic FCU datasets (highlighted within the red box shown in Figure X) where the prior label information of the FCU data are unavailable. The relevant literature of both the groups is discussed in the following sections.

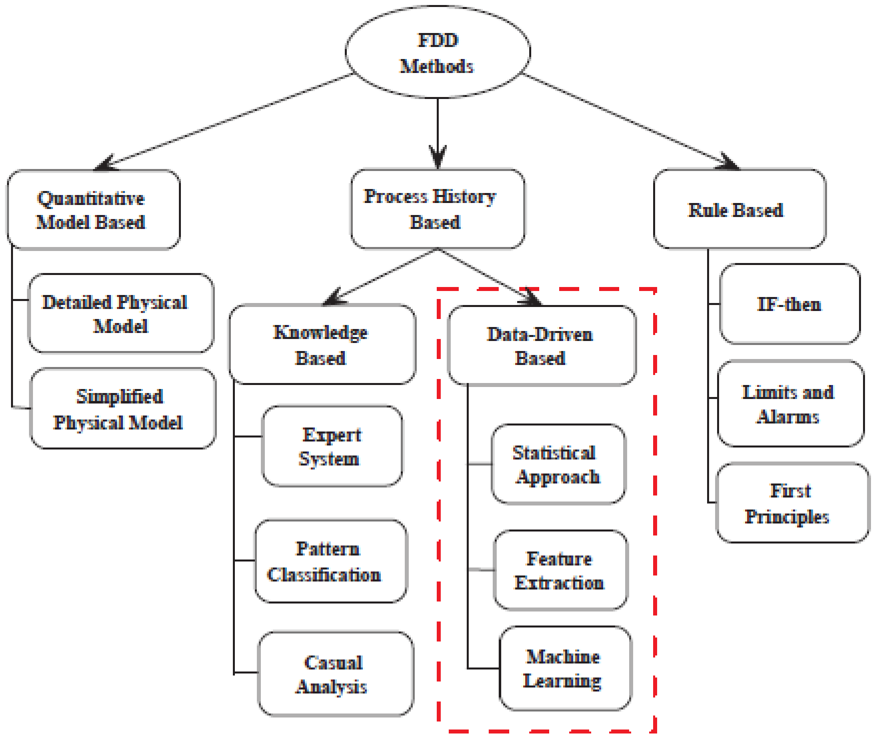


FIGURE 2.2: Classification scheme for fault detection methods.

2.1. Knowledge based method for AFD

Zhao et. al. reviewed artificial intelligence (AI) based fault detection and diagnosis (FDD) systems of building energy systems (BES) and published work [10] describing FDD trends between the years 1998 and 2018. The authors detail the benefit and pitfalls of the existing AI based FDD methods and highlight possible future research directions in the field. The review classifies FDD into two categories i.e., data driven and knowledge driven FDDs. Data driven FDDs mainly rely on the available training data as they are abundant, include supervised, unsupervised, and regression-based learning practices but, problems arise in terms of reliability and robustness. Knowledge-based methods influence the diagnostic process employing human expert's knowledge or expertise to support the decision making. However, in the era of automation engineers and maintenance staff are required to handle the huge tasks that require both human expertise and derived knowledge to supervise an AI based algorithm (i.e., supervised ML). An adaptive Gaussian mixture model (AGMM) approach applying time-varying probabilistic ML model for non-linear systems has been proposed by Karami et. al. [11]. An Unscented Kalman filter (UKF) is integrated with Gaussian mixture model regression for adjusting the model parameters with the help of feedback of residuals between observation and model prediction which is limited only to chiller fault identification. An automatic fault detection technique assembling rapid centroid estimation (ERCE) has been proposed to select illustrative features automatically that are unique in nature to the faults of HVAC system and is able to address different types of air-handling unit (AHU) faults in commercial buildings [12]. Recently, Ranade et. al. proposed a fault diagnosis scheme for FCUs by applying a grey-box model. The work follows a systematic procedure to obtain a simplified model of a heat exchanger coil using polynomial regression to generate residuals. The method shows that the residuals from this model facilitate fault diagnosis by certain rules [36].

2.2. Data-driven based method for AFD

Two-stage data-driven FDD strategy has been modelled with linear discriminant analysis (LDA) followed by a multi-class classification procedure. The LDA reduces the dimension of the in-hand data and clustered faults using the predefined Manhattan distance range to detect and diagnose chiller faults. The clustered information is further used to make fault identification decision solving multi-class classification problem. The method experimented in ASHRAE Research Project 1043 (RP-1043) data for identifying seven types of chiller faults [13]. Gao et. al. proposed association rule-based approach for six different types of air handling units (AHU) behaviour analysing time series data which are instrumented in several buildings of United States of America. Here, twelve performance and assessment rules (APAR) have been inferred for rule based FDD application in AHU and 75% accuracy has been obtained for new or unseen data [14]. A hybrid multi-label classification algorithm assembling clustering and generalized linear mixed model (GLMM) has been proposed by the researchers. Here, clustering groups the available data and reduce the computational complexity of manual labelling whereas, the GLMM figures out the dependency of a subject with multi-labels in training data. The results indicate its suitability towards the large number of labelled information [15]. A laboratory generated air-water heat pump data and their series of features have been analysed with different ML approaches. The fault detection results show that the method performs well with laboratory generated training data set but failed to with real-world data set [16]. Austin et. al. presented a model to estimate air-side capacity based on the specific parameters such as, airflow rate, cooling capacity, system efficiency, and refrigerant mass flow of air handling systems. The sensitivity has been compared with other existing models. Additionally, the model is able to evaluate the uncertainty of input parameters and their sensor requirements. But the model is limited to commercial air-handling units only [17]. A machine learning based anomaly detection and irregular building energy consumption tracking framework has been proposed by Xu and Chen. Here, the recurrent neural network (RNN) has been executed to identify the faulty interval and its energy consumption and the outcomes have been evaluated by the quantile regression range. The framework has been only applied to three different residential houses. This is an unsupervised framework where no prior knowledge is required to detect anomalous behaviours and building managers are benefitted to assess the level of anomalies and spot opportunities in energy conservation [18]. Lee et. al. proposed a real-time deep learning (DL) supported fault diagnostic model for AHUs. Initially, the EnergyPlus simulation software has been employed to establish different types of fault references for DL implementation and behaviour learning. The successful execution of this method shows improvement in the diagnostic process with 95.16% accuracy but has not been tested on real data which lacks the reliability of the model [19]. Zhao et. al. fused wavelet transformation (WT) followed by principle component analysis (PCA) to discover behavioural knowledge and diagnose the HVAC AHU [20]. Beghi et. al. encountered the high dimensional data space problem of buildings and performed a dimensionality reduction technique that maps the data to the lower interested space. The reduced building data have been fed into hybrid model to make an efficient FDD solution for large buildings. However, the work demonstrates that in practice the FDD technique is more appropriate for fault detection rather than fault diagnosis [21]. Magoules et. al. developed an artificial neural network (ANN) utilising recursive deterministic perceptron (RDP) to implement FDD for an entire building level. Remarkably, this new FDD prototype detects and ranks the faulty equipment according to the fault risk [22]. A recent prototype has been developed by Shang and You exercising stochastic model predictive control (SMPC) that provides promising solution to the complex control problems under uncertain disturbances. The SMPC approach actively learns the uncertainty from data driven pipeline involving ML framework [23]. A similar type of energy optimisation pipeline has been implemented by Sonta et. al. along with learning the occupant's behaviour in buildings to improve energy efficiency [24].

2.3. Problem Statement

The literature of building management and its energy handling demonstrate that the researchers have devoted great effort in identifying and developing proficient methods to resolve the real challenges in buildings to optimise their behavioural performance and save the energy wastage mostly at acomponent level via smart building systems. However, many problems are left undetected/unresolved due to large amount of data and complex nature of building management systems (BMSs). The BMS produces vast amounts of data at every minute time interval, most of which have not been analysed and understood due to the lack of building experts increaseing overhead costs as well as time complexity. Thus, the employability of data mining and ML methods are attracting attention for BMS data analysis, but it is complex pathway to discover effective methods due to the fact that learning and knowledge discovery methods are comprehensively data dependent where each type of data represents specific behaviour. This makes the field a highly interesting and vast are for more focused research. Pivotally, although the most numerous units used in building, the fan-coil unit (FCU) has not been explored compared to other major units such as, AHU, chillers, boilers.

2.4. Contribution

The authors have focused on this this small but influential fan-coil unit (FCU)unit and created a method that can be adopted for different buildings and equipment. The proposed work has been performed on a real building based in London, UK as a case study. A FCU is a specific sub-unit of a HVAC system and the main unit of interest for this investigative research work. A machine learning based multi-level automatic fault detection (MLe-AFD) framework is proposed and developed for FCU fault identification and their performance analysis. The study emphasizes the successful utilization of machine learning provided by two-stage learning in the presented multi-level framework. This proposed model allows to utilise the unstructured and unlabelled building data in such a way that fits with this model in the experiment stage. The results of this study show the ability of the presented work to improve the fault identification while having limited amount of information about the FCU behaviour.

3. Overview of FCU and associated faults

A FCU is a ceiling-mounted unit commonly found inside rooms, corridors, open space areas and controlled by local thermostats. It comprises heating coil, cooling coil, and a fan or damper. The return air recirculates internal air or fresh air along with recirculated air and release fresh air to the room depending upon the thermostat. An outer structure and the schematic of FCU are shown in Figure 2. Commonly, the central chiller and boiler plant distributes cold water to all the cooling coils and hot water to all the heating coils. If the environment becomes too warm, the local thermostat senses the rise of temperature and signals the chilled water valve to flow cold water through the cooling coil then cool air being blown by the fan. If the room temperature becomes too cold (depending on the local set point or the user preferred temperature setting), the heating coil starts working in similar way and blows the hot air until the room temperature reaches the anticipated level or set point.

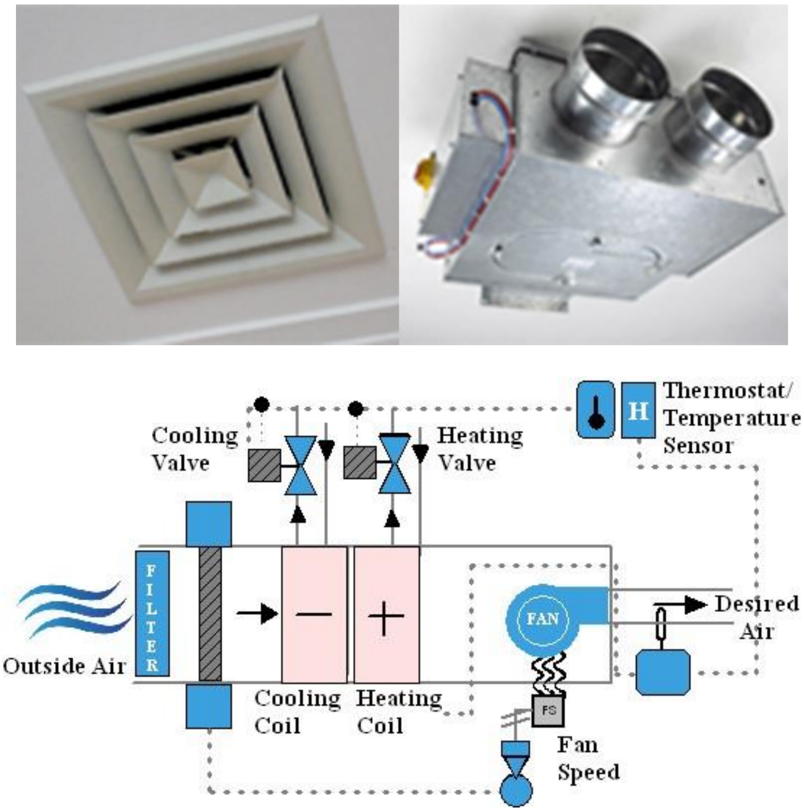
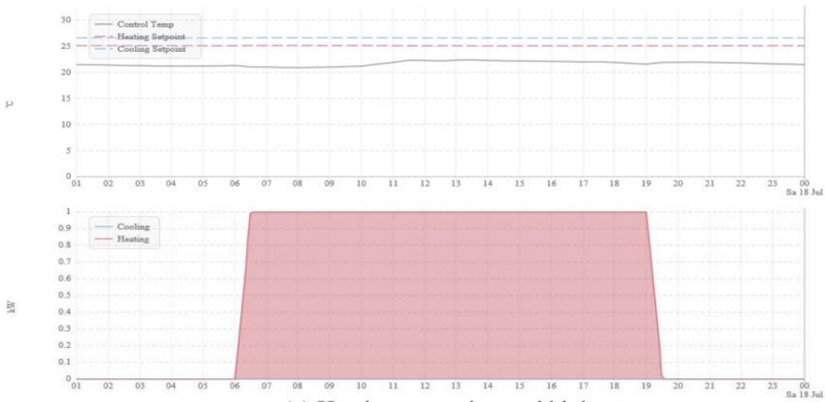
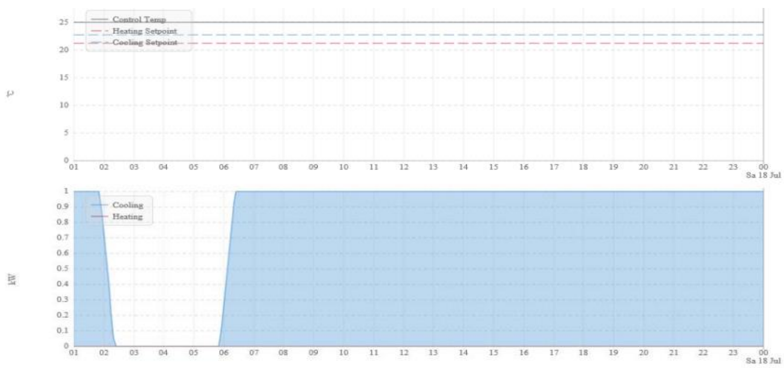


Figure 2: (a) FCU inner and outer structure, (b) Schematic of FCU

Due to dirty plenums, filters, and coils, the resistance increases lowering the air volume causing inappropriate cooling or heating. There are several problems related to air distribution creating performance issues. Thus, three types of FCU performance issues: (a) saturation, (b) on-ness, and (c) hunting have been investigated in the proposed work. These three malfunctioning behaviours have been thoroughly examined and learned by ML methods for fault detection that aims fast maintenance response. Figure 3 visually illustrates the three types of faults that shows raw control temperature and related power demands during different types of faults. Each subfigure has been divided into two parts, the upper part denotes the control temperature variation in degrees centigrade and the lower part displays the associated power demand in kilowatt. Here, the pink and blue dotted lines in the control temperature graphs denote the heating and cooling setpoints respectively. Similarly, the blue colour in the power demand represents cooling power whereas the pink colour represents the heating power. These graphs show the control temperature and corresponding power consumption of a single unit for a whole day, where it is observed that the control temperature could not reach any of the set points (heating/ cooling) though the power demand was continuously high. Figure 3(a) shows the ‘saturation’ type of heating (indicated by pink colour) power trends where the temperature was still struggling to reach its set point. Though, the high-power consumption is found the temperature couldn’t achieve the set point for that instance, illustrating the occurrence of fault which needs to be identified and addressed. Figure 3(b) shows the ‘on-ness’ behaviour in the cooling (indicated by blue colour) power trend. In general, the FCUs are enabled during the daytime of 6:00 am to 6:00 pm to cover the office hours but in this case the power consumption is continuously high and saturated even in the out of operational hours due to this defective behaviour. It wastes significant amount of energy in building operation. Figure 3(c) shows the third type of fault case here, the heating trends and ‘hunting’ nature of the FCU behaviour reveals in terms of power and temperature, while the unit is on even after the operational hours. These three faults of FCU behaviour have been addressed and analysed here using the proposed multi-level clustering (MLC) to learn and identify the fault patterns automatically. The data have been collected in every 10 min timestamp with the control temperature, heating power, cooling power, set point, dead band, and enable signal information from the FCUs of the case study building.



(a) Heating saturation and high temperature error



(b) Cooling saturation and on-ness

(c) Heating hunting and on-ness

Figure 3: The control temperature and associated power behaviour of three types of FCU faults

4. Proposed multi-level automatic fault detection:

The multi-level automatic fault detection (MLe-AFD) model has been proposed performing three stages, feature extraction from the raw FCU data, first-level clustering to separate faulty and non-faulty data, and second-level clustering to identify different types of faults. The performance of the proposed work has been evaluated in each clustering level using statistical validation. It notifies the building engineer about the faults and their types automatically so they can proactively perform guided maintenance. The flowchart of the proposed method has been shown in Figure 4 and described in the following sections.

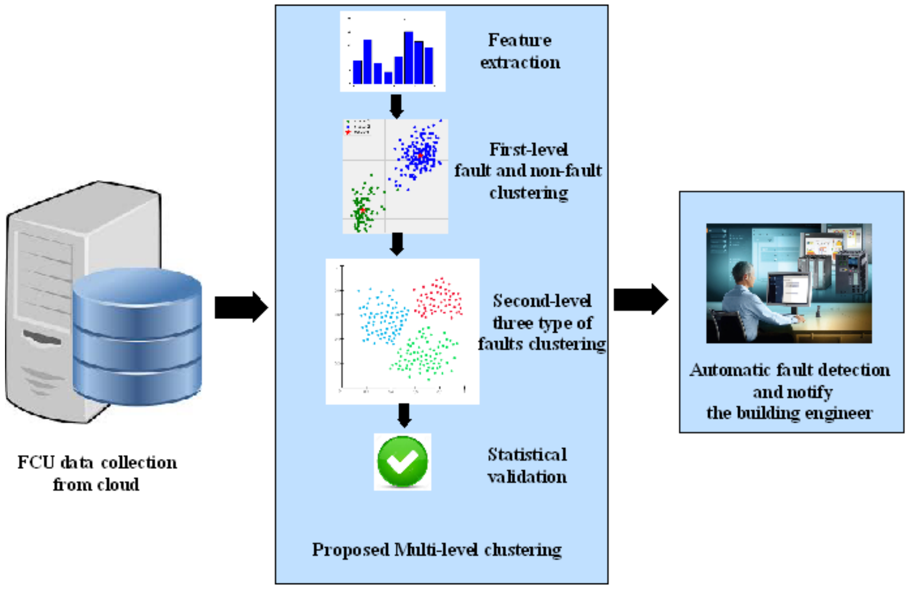


Figure 4: Flow diagram of proposed AFD method

4.1 Data Collection Process:

The data are gathered through the data acquisition device (DAD) installed in the building, which acts as a gateway to connect an existing/resident BMS to a secure internet service by the Demand Logic (DL) team (London). The data gathering process is completed within 24 to 48 hours and creates a virtual asset model (VAM) of all the equipment installed in that BMS network, considering each data point as BMS data.

The BMS data is extracted through a single embedded PC Engine 2D13 ALIX connected through a mobile network router. The embedded PC contains embedded software that is used to: (i) obtain a map of the BEMS network. This includes all the BEMS internetworks and consists of multiple local area networks, devices on a LAN (a single device may relate to one or many services equipment), and data points on a device. These can be binary or analogue control signals, feedback signals or settings. The text label and numerical ID are obtained for each of the LAN, device and data point, (ii) pulling the data points (typically at i 10 minute intervals), (iii) store/buffer the data if the internet connection is lost, and (iv) securely send the data to the cloud servers.

4.2 Feature extraction

An 'intelligent' feature extraction method was proposed and applied by the authors [25, 26] to deal with the high dimensional data and project into reduced data dimension. The feature extraction process generated informative and non-redundant information facilitating subsequent learning and improved the performance of the entire framework. This feature extraction method is performed on the six FCU parameters: control temperature, set point, dead-band, heating power, cooling power, and enable signal. These FCU data are collected at every 10 minute interval from the time series data on a daily basis via a secured gateway. The proposed feature extraction method has been employed to discover different events: event start, respond delay, goal achieved, and event end based on their temperature and power flow during a day (24hrs). The area (A_E) under temperature and power curve ($f(x_i)$) at each time interval (Δ_x) is calculated event wise. This area under the curve calculation is carried out for both heating (H) and cooling (C) actions. There are six different features are measured from each of the heating and cooling events employing the event area calculation,

$$F_{H_k} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{6}\left(F_{H_{k_i}}\right) \qquad (1)$$

$$F_{C_k} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{6}\left(F_{C_{k_i}}\right) \qquad (2)$$

Where, F_H represents heating and F_C represents cooling features, n is the number of occurrences of each event type (heating or cooling) on a day, k is the number of features generated by measuring the area of each events (A_E). The feature extraction method transforms and represents heating-cooling events of a whole day by the twelve-dimensional (12-D) feature vector. On a day there are $(\frac{24*60}{10})$ = 144 data point have been collected for each FCU parameter at 10 minute intervals. Six different parameters have been considered for each FCU, thus altogether (144*6) = 864 data points exist for each FCU, subsequently converted into 12 meaningful features employing the cooling and heating operation information. These feature vectors have further considered for clustering to automatically detect and identify the faults.

4.3 Multi-level clustering:

The faulty and non-faulty behaviour ground truths of the FCUs were unavailable to the authors during this investigation, thus, the unsupervised learning approach i.e., clustering has been incorporated here to discover the FCU behavioural patterns without any prior knowledge. The clustering algorithm categorises the similar types of FCU behaviours into same cluster based on dissimilarity found in the feature space. Thus, the MLe-AFD has been implemented into two stages. The first level clustering has been performed to separate faulty and non-faulty FCU patterns (obtained two clusters), the obtained clusters have been thoroughly analysed to understand each group. The second level clustering has been applied into a faulty cluster to identify further faulty groups (three groups obtained). Three well-known clustering algorithms have been implemented for multi-level application: k-means [27, 28], average linkage hierarchical clustering [28], and Gaussian mixture model (GMM) clustering [29] depending upon the data characteristics.

The objective function of k-means is defined in Eq. (3), where $||X_i^{(j)} - \mu_j||^2$ is the Euclidean distance between the data point $X_i^{(j)}$ and the cluster centre μ_j. The distance between the *n* number of data points from their respective cluster centres are defined as 'k' for each cluster. Each data point is assigned to the group that has the nearby centroid. After all the data points are assigned, the positions of the *k* centroids are recalculated. The steps are re-iterated until the centroids no longer move.

$$\mathrm{J} = \sum_{j=1}^{k} \sum_{i=1}^{n} \left|\left|\mathrm{X}_i^{(j)} - \mu_j\right|\right|^2 \quad (3)$$

Hierarchical clustering creates a grading of clusters in the FCU dataset measured by linkage criteria between the sets or the groups of observations. It is the function to measure the pairwise distances between the observations in each set. The objective function for average linkage hierarchical clustering is defined in Eq. (4), where, a and b are the object which belong to the set A and B.

$$\max \mathrm{d}(\mathrm{a}, \mathrm{b}) : \mathrm{a} \in \mathrm{A}, \mathrm{b} \in \mathrm{B} \quad (4)$$

$$\frac{1}{|A|.|B|} \sum_{a \in A} \sum_{b \in B} \mathrm{d}(\mathrm{a}, \mathrm{b})$$

The GMM follows the soft clustering technique for assigning the FCU behaviour data points to the Gaussian distributions. The GMM decides k number of clusters calculating the mean (μ), covariance ($\sum k$) and the density of the distribution (π_i). The working principle of GMM doesn't rely on the shape of distribution, is shown in Eq. (5),

$$p(x) = \sum_{i=0}^{k} \pi_i N\,(x \mid \mu_k, \textstyle\sum k\,) \quad (5)$$

The compactness of clustering outcomes has been evaluated in each level measuring statistical metrics to be confident about the inter-cluster separation and intra-cluster coherence.

4.4 Validation

Three standard clustering validation techniques have been implemented; Gap [30], Silhouette indexing (SI) [31], Davies-Bouldin (DB) [32] to analyse and compare performance. These measures are performance indicators used where the class labels (or, ground truths) are not available. These performance indicators determine the degree of intra-cluster cohesion and inter-cluster separation. Each of these methods has its own numeric range to illustrate the compactness of the clusters where, loosely coupled clusters are required further investigation. The Gap approach is expressed in Eq. (6) as,

$$G_n(k) = E_n^* log(W_k) - log(W_k) \qquad (6)$$

Where, En* denotes expectation under the sample size of n from the distribution, considering as uniform data points with k centres, and the gap statistic measures the deviation of the observed W_k value from its expected value under the null hypothesis. The k is the optimal cluster number where the gap measure maximises. The Silhouette indexing (SI) defined in Eq. (7),

$$\mathrm{SI} = \frac{1}{nk}\sum_{\mathrm{i}\in\mathrm{k}} \frac{\mathrm{b_i} - \mathrm{a_i}}{\mathrm{max}_{(\mathrm{a_i},\mathrm{b_i})}} \qquad (7)$$

Where a_i is the average distance from the i^{th} point to the other points in the same cluster, and b_i is the minimum average distance from the i^{th} point to the points in different clusters. The silhouette value ranges from -1 to +1. A high value indicates that i is well-matched to its own cluster. The clustering solution is considered appropriate if most points have a high silhouette value. The Davies-Bouldin (DB) measure is denoted in Eq. (8),

$$\mathrm{DB} = \frac{1}{k}\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{max}_{\mathrm{j}\neq\mathrm{i}} \frac{\delta_\mathrm{i} + \delta_\mathrm{j}}{\Delta_{\mathrm{i,j}}} \qquad (8)$$

Where, $\Delta_{\mathrm{i,j}}$ is the cluster's distance ratio for the i^{th} and j^{th} data points within and to the other cluster. The ∂_i and ∂_j are the average distances between each point in the cluster from the centroid of that i^{th} cluster and the average distance between each point in the j^{th} cluster and the centroid of the j^{th}. The value ranges from -1 to +1 and a low DB value consider the clustering solution is appropriate.

4.5 Hypothesis Test: Two-sample t-test

A Hypothesis test has been performed to be confirm between the relationships of the FCU behavioural clusters. Two FCU groups have been tested at a time thus a two-sample t-test has been performed here. Here, the null hypothesis (Ho) indicates that the two FCU clusters come from independent random samples from a normal distributions using the two-sample t-test and the alternative hypothesis (H1) is vice versa. The Ho has been accepted or rejected at the 5% significance level (α=0.05).

5. Experimental result analysis:

The proposed MLe-AFD has been examined on a real case study building. The details of the case study building, and outcomes have been discussed in the following section.

5.1 Case study description

The case study building is London based, built in 1960 and renovated in 2009. It covers 149,000 sq. ft. for offices and 8,000 sq. ft. of retail space. The building has seventeen (17) floors with seven hundred and thirty-one (731) FCUs distributed across the different floors. A total of 723 FCUs were operating out of a total of 731. The proposed MLe-AFD has been experimented on using building data gathered since 2015 and continued to 2020. Thus, the 723 FCUs for each day and consecutively 3615 FCUs have been monitored for one week with thorough analysis performed to understand the behavioural patterns of faulty and non-faulty FCUs of the building so that it could useful for future fault anticipation.

These data are then accessed in the University lab for research purposes. The experiment has been carried out using MATLAB R2018b tool on an Intel(R) Core (TM) i5 processor@ 3.30 GHz running Windows 7 Enterprise 64-bit operating system with 7856-MB NVIDIA Graphics Processing Unit (GPU).

5.2 Feature correlation

The correlation between the features obtained from the proposed feature extraction method has been analysed here calculating the Pearson correlation coefficient (PCC) [33]. Pearson's correlation coefficient is known as the best test statistical method for measuring the association and relationship between two continuous variables of interest and is based on the covariance. The coefficient values range between -1 to +1, where +1 defines a perfect correlation between the FCU features. The first six features are related to the heating trends and the next six features are related cooling trends. Additionally, the each FCU features represent the control temperature and power behaviour events [25, 26]. The heating features have the negative correlation with cooling features that represents inverse relationship with each other. It is found from the colour map in Figure 5 that feature variables are perfectly correlated with their own pair and moderately correlated with pair variables where the value ranges from -0.0003234 to +1.

Figure 5: Feature correlation colour map

5.3 Clustering results

The proposed MLe-AFD has been performed in two stages using three different methods and the results compared to identify the best fitted clusters with the case study FCU dataset. The first-level clustering has been performed on each day using 12-D feature vector for all the one day and one week FCUs and clustered them into two groups, faulty and non-faulty. Three different clustering algorithms have been performed here two find these two groups. The results have been verified using two methods, i) by applying statistical metrics (describe in Section 4.3) and ii) through analysis by building engineers where, they have confirmed the identified faulty and non-faulty FCU patterns (described in Section 4.3). The number of faulty and non-faulty FCUs found from each clustering model are shown in Table 1 which displays the numbers in each category obtained from the models. The non-faulty groups obtained from all three clustering techniques contained more FCUs than the faulty groups. The outcomes indicated that, most of the FCUs were working properly which has been confirmed later by the building engineers.

Table 1: The number of data points accumulate into each cluster after employing first-level clustering			
A day FCU data outcomes	Methods	Non-faulty	Faulty
	k-means	598	125
	Linkage	592	131
	Gaussian mixture	596	127
A week FCU data outcomes	k-means	3185	430
	Linkage	3190	425
	Gaussian mixture	3177	438

Now, these two clusters obtained from each model have been validated through the internal evaluation schemes and results are displayed in Table 2. The validation schemes assess how well the clustering has been performed using the quantities and features inherited from the FCU dataset. In the case of the Gap algorithm, it finds the optimal number of clusters first which might provide a best fit for the FCU data in hand and a maximal Gap metric score to indicate the clustering performance. The number of FCU behaviours are known to the authors from building engineers i.e., two and three groups respectively in first and second level clustering. The first two groups (faulty and non-faulty) are common but, three types of faults in the next level have been informed by engineers collaborating on this research and this information was particularly supportive towards the optimal number of clusters. It was found from the k-means clustering that there were 598 non-faulty and 125 faulty FCUs, where the Gap criterion achieved 1.10 and expresses very good compactness of the faulty and non-faulty FCU clusters. The SI measure scores 0.9649 which is very near to +1 also defines decent clustering outcomes and 0.2582 from k-means. The DB scored 0.2582 where the metrics close to 0 indicates promising outputs from k-means. There are 592 FCUs found in non-faulty cluster and 131 FCUs into faulty cluster. Here, the Gap index scores are found same as of k-means, but less SI score achieved than the k-means and Gaussian mixture clustering. Also, the obtained DB index is 0.1595 which is moderately low and indicates good linkage clustering performance. In case of GMM clustering, 596 and 127 FCUs have been identified in non-faulty and faulty groups respectively and all the considered indexes indicate decent clustering outcome. Similarly, the experiment has been tested on a week's worth of data that comprises 3616 FCUs altogether. Table 1displays how many of these week long FCUs have been assigned into two groups by the chosen algorithms. It can be observed that a greater number of data that has been grouped together belongs to the non-faulty behavioural group. Also, in Table 2 the internal validation results from the three measures have been incorporated to investigate the compactness of these clustering algorithms and it was found that all three methods achieved optimal scores based on their evaluation criteria.

Table 2: The validation score of different methods obtained by first level clustering				
	Methods	Gap	Silhouette	Davis Boulding
A day FCU data outcomes	k-means	1.100	0.9649	0.2582
	Linkage	1.100	0.9421	0.1595
	Gaussian mixture	1.090	0.9652	0.2517
A week FCU data outcomes	k-means	2.248	0.970	0.3279
	Linkage	2.369	0.9715	0.3138

	Gaussian mixture	2.246	0.9715	0.3138

Figure 6 shows the bar plot to compare all the first level clustering performances for daily and weekly FCUs. In the daily data, it has been found that GMM performed better than other two clustering methods in first level clustering to identify faulty and non-faulty FCUs. All the methods achieved good performance criteria in the weekly data analysis. However, the Linkage and GMM clustering achieves the same score in the Silhouette and Davis Boulding indexing calculations. Our partner building engineers have verified the clustering outcomes. Subsequently, the faulty FCU groups have been further clustered using corresponding algorithms to categorise different types of faults in the second level clustering.

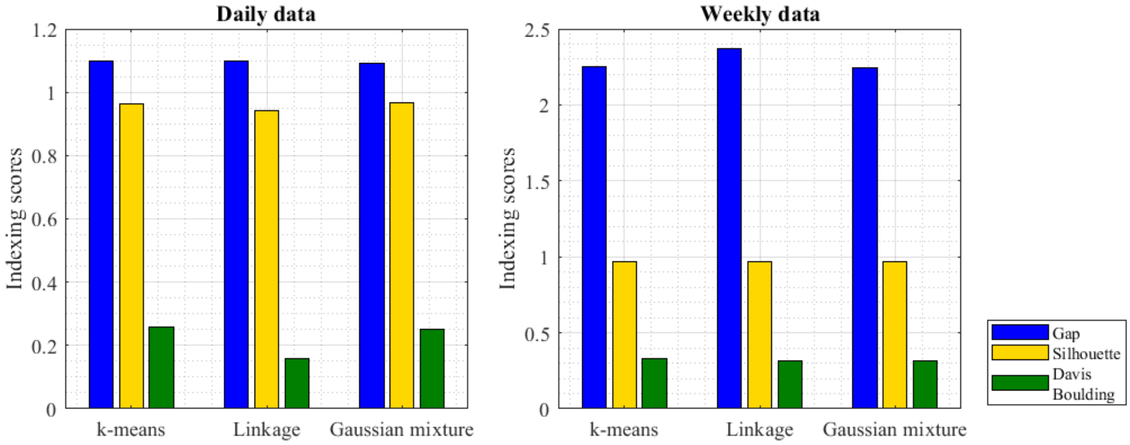


Figure 6: Bar plot for first level clustering validation comparison

The second level clustering has been performed only on the faulty FCU groups, i.e., 125 FCU patterns for k-means clustering, 131 FCU patterns for linkage, and 127 FCU patterns for GMM clustering have been considered at this stage for day FCU experiment. Subsequently, 430, 425, 438 FCU patterns have for second level clustering using k-means, linkage, GMM clustering respectively in week data investigation. As the building engineers informed previously there are three types of faults, hunting, saturation, and on-ness noted in FCU behaviour. Thus, the clustering aimed to break each faulty cluster into three more groups. The allocation of FCUs into each fault type are tabulated in Table 3 with the corresponding clustering algorithm. These clusters also further verified through the expert building engineers to understand and map their categories into the obtained faults. The faults included in Table 3 denotes, fault-1 as hunting, fault-2 as saturation, and fault-3 as on-ness patterns. Alike the first level, this second-level clustering also validated statistically and through engineers to check their compactness and system performance, is tabulated in Table 4.

Table 3: The number of FCU data accumulated from each clustering to perform second-level clustering

Experimental Time Frame	Methods	Total no of FCU data	Fault-1	Fault-2	Fault-3
A whole day	k-means	125	35	24	66
	Linkage	131	37	25	69
	Gaussian mixture	127	38	27	62
A whole week	k-means	430	175	47	208
	Linkage	425	174	46	205
	Gaussian mixture	438	177	51	207

It has been found from daily data analysis employing k-means that 35, 24, 66 FCUs were identified as displaying hunting, saturation, and on-ness, respectively. Linkage demonstrated that, 37, 25, 69 FCUs were grouped as displaying hunting, saturation, on-ness respectively. The GMM identified that, 38, 27, 62 FCUs displayed hunting, saturation, and on-ness in nature respectively. Similarly, it has been observed from the weekly data analysis that k-means separated 175, 47, 208 FCUs into three distinct faulty behaviours, whereas linkage grouped 174, 46, 205 FCUs into different clusters, and 177, 51, 207 FCUs wereseparated by GMM algorithm. The validation scores have been summarized in Table 4. The k-means and linkage clustering models achieved identical scores for Gap criterion, whereas linkage and GMM achieved better scores than k-means for both SI and DB indexing for daily analysis. In the case of weekly analysis k-means achieves different performance score for all the validation methods whereas linkage and GMM achieve similar scores for silhouette and Davis Boulding. All these implemented methods have achieved good performance scores for all the internal validation criteria which are acceptable and considered as good clustering for these FCU behaviour analysis.

Table 4: The validation score obtained from different methods in second level clustering

Experimental Time Frame	Methods	Gap	Silhouette	Davis Boulding
A whole day	k-means	1.310	0.9517	0.207
	Linkage	1.310	0.9527	0.1925
	Gaussian mixture	1.300	0.9527	0.1925
A whole week	k-means	2.720	0.9877	0.2443
	Linkage	4.281	0.9930	0.2440
	Gaussian mixture	4.157	0.9930	0.2440

Figure 7 shows the bar plot comparison of internal validation scores obtained from all three clustering methods for day and week FCU data analysis. It has been realised from second-level cluster analysis that GMM provides optimal performance for FCU automatic fault detection of s over the other two algorithms tested. Henceforth, the proposed MLe-AFD employing GMM have been considered to capable of detecting distinct FCU fault patterns automatically and without any prior information.

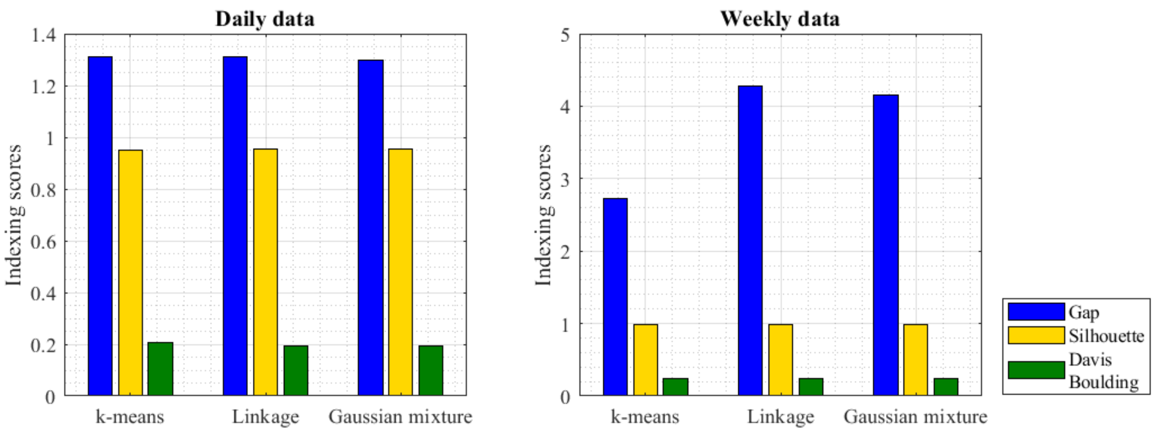


Figure 7: Bar plot for second level clustering validation comparison

5.4. Hypothesis test

A Hypothesis test has been performed for trusting the relationship between different types of FCU populations that have been separated through clustering, this test gives precise criteria for rejecting or accepting a null hypothesis i.e., the obtained results within a significance level [37]. The proposed MLe-AFD has been employed to detect anomalously behaving FCUs and pivotally their types of anomalies without the label information. One day and oneweek data have been initially considered totest the outcomes of the proposed MLe-AFD. Hence, the paired t-test has been implemented to discover the correlation between the predicted FCU clusters and to confirm the FCUs belongingness to a particular cluster. Table 5 shows the comparison of p-values obtained by the MLe-AFD where, the first two clusters obtained from first level clustering represent the non-faulty and faulty patterns considering control temperature and corresponding power variation. Another three clusters have been obtained from the second level clustering representing the different faulty FCU patterns. The significance level 0.05 has been considered for the FCU cluster to justify the null hypothesis. The null hypothesis has been accepted for a predicted cluster where the p-value is greater than the significance level. From the table below, it has been seen from both the daily and weekly data that the first and second level clustering obtained a p-value less than the significance level (p-value < 0.05). Thus, it depicts the null hypothesis is rejected, indicating that the proposed MLe-AFD could cluster the data into appropriate distinct patterns.

Table 5: Hypothesis test obtained for daily and weekly data including the p-value of first and second level clustering populations

	Level-1 clustering	Level – 2 clustering		
	p-value for faulty and non-faulty population	p-value for fault-1 and fault-2 population	p-value for fault-1 and fault-3 population	p-value for fault-2 and fault-3 population
Day test	0.0176	0.0464	0.0184	0.0196
Week test	0.0239	0.0242	0.0508	0.0349

6. Discussions and Conclusions:

It has been concluded from this study that the multi-level machine learning framework is an effective solution for the automatic detection of FCU fault patterns in commercial buildings. The proposed AFD method was developed to illustrate how predictive machine learning methods can be influential in creating useful smart buildings with inherent and automated fault identification in heating and cooling FCUs as is the case investigated here. This study has described the benefits of the proposed approach to identify faults remotely and importantly anticipate their behaviour automatically. Although, it has been realised from the clustering application that, all the FCUs are identified as faulty, some may not actually be faults but a result human interference such as open windows etc where the FCU cannot cope with sudden changes in temperature. Other natural, but not considered changes also cause “faults”, e.g. when the sun shines in a room creating a temperature level that the FCU set point cannot cope with. Again, these are issues that can be identified via the method proposed but text can inform managers of these non- fault issues, reducing workloads and informing building design efforts. Creating systems that can identify, predict and categorise faults and non-faults by costs (determined by the building manger are pivotal to future smart building) will be pivotal to building sustainability, occupant well-being and green efforts. The proposed work has obtained statistically acceptable indexing scores.

The proposed method has a significant impact on energy savings as well. Identification of the faulty behaving units directly affects energy consumption performance of the built environment. This automatic approach will be the first step to reduce building energy wastage. The automatic fault findings would be beneficial for future fault anticipation, ensuring appropriate preventative maintenance strategies . This could significantly reduce the operational energy consumption and cost of the HVAC units.

The proposed MLe-AFD strategy has been executed in two stages, the feature extraction method followed by unsupervised learning techniques. The feature extraction considers the temperature, set point, and corresponding power for FCU characterisation, where three distinct FCU faults have been remotely identified. Further, this investigation has been validated through engagement with building engineer to understand the effectiveness of the proposed framework. This method can reduce the manual workload of fault finding and identification and provides the necessary leap towards useful and applied AFD that can assist in predictive maintenance where necessary. Additionally, this method would help building engineers to look at a single FCU unit along with the whole cluster where they can take necessary actions for all the affected units belonging to that cluster without looking into each of them. Thus, the proposed MLe-AFD method optimizes the building's operational workload identifying abnormally behaving equipment reducing the large amount of energy loss in smart buildings. The method developed and deployed for this paper is currently focuses on a specific type of HVAC unit (Fan Coil Unit) of a single building, but will be extended to different types of units such as air handling unit (AHU), variable air volume (VAV) and chilled beam etc. for widespread and applied validation.

Author Contributions:

M.D. has performed the data collection, analysis, proposed and design the algorithms, and wrote the paper, S.P.R. has conducted the experiments, guided to develop algorithm and wrote the paper, S.D. has supervised this work throughout and instigated the research and the collaborative work on this paper between teams at DL and LSBU. All authors reviewed the manuscript.

Acknowledgement:

This research project is funded by Innovate UK (EP/M506734/1). The authors are grateful to the Demand Logic team for providing us the data and feedback on building services and related performance issues.

Conflict of Interest:

The authors declare no conflict of interest.

References

1. Robust ensemble learning framework for day-ahead forecasting of household based energy consumption.
2. Dionysia Kolokotsa. The role of smart grids in the building sector. Energy and Buildings, 116:703-708, 2016.

3. Building energy databook.<https://openei.org/doe-opendata/dataset/buildingsenergy-data-book>[last accessed: 2017-11-23].
4. Clastres C. Smart grids: Another step towards competition, energy security and climate change objectives. Energy Policy 2011;39:5399–408.
5. Yu X, Cecati C, Dillon T, Simões MG. The new frontier of smart grids. IEEE Ind Electron Magaz 2011;5:49–63.
6. Roth, K.W., et al, 2005. “Energy Impact of Commercial Building Controls and Performance Diagnostics: Market Characterization, Energy Impact of Building Faults and Energy Savings Potential.” TIAX. http://tinyurl.com/TIAX2005.
7. Yang Zhao, Tingting Li, Xuejun Zhang, and Chaobo Zhang. Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. Renewable and Sustainable Energy Reviews, 109:85-101, 2019.
8. PB Usoro, IC Schick, and S Negahdaripour. An innovation-based methodology for hvac system fault detection. Journal of dynamic systems, measurement, and control, 107(4):284-289, 1985.
9. D Anderson, L Graves, W Reinert, JF Kreider, J Dow, and H Wubbena. A quasi-real-time expert system for commercial building hvac diagnostics. ASHRAE Transactions (American Society of Heating, Refrigerating and Air-Conditioning Engineers); (USA), 95(CONF-890609), 1989.
10. Zhao, Yang, et al. "Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future." Renewable and Sustainable Energy Reviews 109 (2019): 85-101.
11. Karami, Majid, and Liping Wang. "Fault detection and diagnosis for nonlinear systems: A new adaptive Gaussian mixture modeling approach." Energy and Buildings 166 (2018): 477-488.
12. Guo, Ying, et al. "Real-time HVAC sensor monitoring and automatic fault detection system." Sensors for Everyday Life. Springer, Cham, 2017. 39-54.
13. Li, Dan, Guoqiang Hu, and Costas J. Spanos. "A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis." Energy and Buildings 128 (2016): 519-529.
14. Gao, Jingkun, and Mario Bergés. "A large-scale evaluation of automated metadata inference approaches on sensors from air handling units." Advanced Engineering Informatics 37 (2018): 14-30.
15. Lin, Sung-Chiang, Chih-Jou Chen, and Tsung-Ju Lee. "A multi-label classification with hybrid label-based meta-learning method in Internet of Things." IEEE Access (2020).
16. Bode, Gerrit, et al. "Real-world application of machine-learning-based fault detection trained with experimental data." Energy (2020): 117323.
17. Rogers, Austin, Fangzhou Guo, and Bryan Rasmussen. "Uncertainty analysis and field implementation of a fault detection method for residential HVAC systems." Science and Technology for the Built Environment 26.3 (2020): 320-333.
18. Xu, Chengliang, and Huanxin Chen. "A hybrid data mining approach for anomaly detection and evaluation in residential buildings energy data." Energy and Buildings (2020): 109864.
19. Lee, Kuei-Peng, Bo-Huei Wu, and Shi-Lin Peng. "Deep-learning-based fault detection and diagnosis of air-handling units." Building and Environment 157 (2019): 24-33.
20. Yang Zhao, Shengwei Wang, and Fu Xiao. Pattern recognition-based chillers fault detection method using support vector data description (SVDD). Applied Energy, 112:1041-1048, 2013.
21. A Beghi, R Brignoli, L Cecchinato, G Menegazzo, M Rampazzo, and F Simmini. Data-driven fault detection and diagnosis for hvac water chillers. Control Engineering Practice, 53:79-91, 2016.

22. Frederic Magoules, Hai-xiang Zhao, and David Elizondo. Development of an RDP neural network for building energy consumption fault detection and diagnosis. Energy and Buildings, 62:133-138, 2013.
23. Chao Shang and Fengqi You. A data-driven robust optimization approach to scenario-based stochastic model predictive control. Journal of Process Control, 75:24-39, 2019.
24. Andrew J Sonta, Perry E Simmons, and Rishee K Jain. Understanding building occupant activities at scale: An integrated knowledge-based and data-driven approach. Advanced Engineering Informatics, 37:1-13, 2018.
25. Dey, Maitreyee, et al. "A pid inspired feature extraction method for hvac terminal units." 2017 IEEE Conference on Technologies for Sustainability (SusTech). IEEE, 2017.
26. Dey, Maitreyee, Soumya Prakash Rana, and Sandra Dudley. "Smart building creation in large scale HVAC environments through automated fault detection and diagnosis." Future Generation Computer Systems (2018).
27. Dey, Maitreyee, et al. "Unsupervised learning techniques for HVAC terminal unit behaviour analysis." 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation. IEEE, CA, USA, 2017.
28. Sergios Theodoridis and Konstantinos Koutroumbas. Pattern recognition. 2003. Elsevier Inc, 2009.
29. McLachlan, G., and D. Peel. Finite Mixture Models. Hoboken, NJ: John Wiley & Sons, Inc., 2000.
30. Tibshirani, R., G. Walther, and T. Hastie. “Estimating the number of clusters in a data set via the gap statistic.” *Journal of the Royal Statistical Society: Series B*. Vol. 63, Part 2, 2001, pp. 411–423. (For gap)
31. Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20:53-65, 1987.
32. David L Davies and Donald W Bouldin. A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence, (2):224-227, 1979.
33. Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. *Numerical Recipes in C*, 2nd Ed., Cambridge University Press, 1992. (for correlation)
34. Woohyun Kim and Srinivas Katipamula. A review of fault detection and diagnostics methods for building systems. Science and Technology for the Built Environment, 24(1):3{21, 2018.
35. González-Briones, Alfonso, et al. "Multi-agent systems applications in energy optimization problems: A state-of-the-art review." Energies 11.8 (2018): 1928.
36. Ranade, A., Provan, G., Mady, A. E. D., & O'Sullivan, D. (2020). A computationally efficient method for fault diagnosis of fan-coil unit terminals in building Heating Ventilation and Air Conditioning systems. Journal of Building Engineering, 27, 100955.
37. González-Briones, A., Prieto, J., De La Prieta, F., Herrera-Viedma, E., & Corchado, J. M. (2018). Energy optimization using a case-based reasoning strategy. Sensors, 18(3), 865.
38. Igor Sartori, Assunta Napolitano, Karsten Voss, Net zero energy buildings: A consistent definition framework, energy and buildings, vol 48, 2011, Pages 220-232;