Discover Data

Research

A data-centric approach to terminal unit's fault categorization and optimal positioning in building HVAC systems using ensemble learning

Maitreyee Dey[1] · Preeti Patel[1] · Soumya Prakash Rana[2]

Received: 5 November 2024 / Accepted: 23 April 2025
Published online: 08 May 2025

Abstract
This paper focuses on the fault detection and diagnosis of terminal units (TUs) in a building located in London, utilizing real operational historical data to assess their performance and optimal placement across multiple floors. While precise locations of the TUs are unavailable, our method analyzes their operational behaviour for one month, applying popular machine learning models to detect and analyze faults effectively. By examining each TU individually and in the aggregate, we identify behavioural patterns that inform decisions regarding their positioning within the building. The dataset comprises over 2 million data points collected from 730 TUs, enabling a comprehensive analysis of their functionality and the impact of suboptimal thermostat placements. Our study employs three machine learning models-traditional multi-class Support Vector Machines and two ensemble methods: Random Forest (RF), and Adaptive Boosting (AdaBoost)-to classify TU behaviors into normal operation, heating faults, and cooling faults. Results indicate that RF outperforms the other models with an accuracy of 99.89%, while AdaBoost achieves an accuracy of 85% and SVM shows 47% accuracy. The findings underscore the potential of a data-driven approach to inform retrofitting decisions and enhance the reliability of HVAC systems. This research contributes valuable knowledge toward optimizing TU placement, ultimately leading to improved energy efficiency and indoor environmental quality.

Keywords Fault categorisation · Building energy · HVAC · Terminal unit · Statistical data analysis · Ensemble learning

1 Introduction

The complexity of modern buildings, which integrate various systems such as heating, ventilation, air conditioning (HVAC), lighting, and power control, highlights the need for a thorough understanding of their operational dynamics [2]. Among these systems, HVAC plays a crucial role in maintaining indoor environmental quality and energy efficiency. These systems are responsible for regulating air quality, temperature, and humidity, all of which are essential for occupant comfort and well-being. Given that individuals spend a significant amount of time in these environments, it is important to investigate how factors affecting comfort can influence energy consumption patterns in conjunction with broader well-being initiatives.

Unfortunately, equipment failures and performance degradation within HVAC systems often go unnoticed until they negatively impact occupant comfort, trigger alarms, or result in excessive energy use and increased operational

✉ Maitreyee Dey, m.dey@londonmet.ac.uk; Preeti Patel, p.patel@londonmet.ac.uk; Soumya Prakash Rana, s.rana@greenwich.ac.uk
| [1]GENESIS Research Lab, Computer Science and Applied Computing, London Metropolitan University, 166-220 Holloway Rd, London N7 8DB, UK. [2]School of Engineering, University of Greenwich, Medway Campus, Central Avenue Gillingham, Chatham ME4 4TB, UK.

costs. Terminal units (TUs), in particular, encounter various operational challenges that can hinder their efficiency and overall effectiveness [7]. Problems such as improper thermostat placement, malfunctioning components, and inadequate maintenance can lead to higher energy consumption and diminished system reliability. Therefore, the capability to detect and diagnose faults in TUs is essential for improving their operational performance. Early identification of these faults can prevent larger system failures, lower operational costs, and enhance overall energy efficiency [17].

The paper is structured as follows: Sect. 2 outlines the related work in the context of different machine learning models applied to building fault detection and diagnosis, along with their associated challenges. Section 2.6 presents the contributions and motivation of this study. Section 3 focuses on various faults and associated challenges in terminal units. Section 4 describes the proposed methodology, including terminal unit data descriptions, feature transformation, and data labeling. Additionally, it details the ensemble learning models and validation techniques. Section 5 discusses the data analytics and results obtained by applying ensemble and machine learning models. Finally, Sect. 6 concludes the paper and provides future research directions for this work.

2 Related work on fault detection and diagnosis

Automated fault detection and diagnosis methods for HVAC systems have been in use for several years, with recent advancements in machine learning and data mining accelerating their development. Historically, Fault Detection and Diagnosis (FDD) approaches have been classified into three primary categories: quantitative models, process history-based methods, and rule-based techniques. The process history-based category is further divided into two subgroups: knowledge-based methods and data-driven approaches [39]. Figure 1 gives the categorisation of these two subgroups. Knowledge-based methods require extensive prior information to process data, whereas data-driven models bypass this need by extracting patterns and insights directly from the data. While knowledge-driven methods, particularly those based purely on expert knowledge, remain prevalent in the field of fault detection and diagnostics, data-driven methods are gaining increasing popularity. This rise is largely attributed to their capacity to utilize classification, unsupervised learning, and regression techniques, often enhanced by artificial intelligence [81]. Accordingly, the following sections will review relevant literature on the growing prominence of data-driven FDD methods, as this study employs machine learning techniques to improve the performance of the terminal unit/fan coil unit.

In the past few years, numerous reviews have been conducted with the aim of providing insights into the prevalent methods and approaches for fault detection and diagnosis in HVAC systems [7, 8, 31, 32, 41, 45, 48, 49]. These reviews frequently classify the work of researchers based on the use of knowledge-based or data-driven methods for FDD. In many cases, data-driven methods are further categorized into supervised, unsupervised, semi-supervised, or hybrid learning approaches. Notably, [8] provides a comprehensive overview of the data-driven approach, detailing the sequential steps of data collection, cleansing, and pre-processing, leading to fault detection, diagnosis, and prognosis.

2.1 Supervised learning

In supervised learning, each training observation contains both input and output values to classify unseen data [22], allowing the model to predict and correct output values for unseen inputs by learning the relationship between them

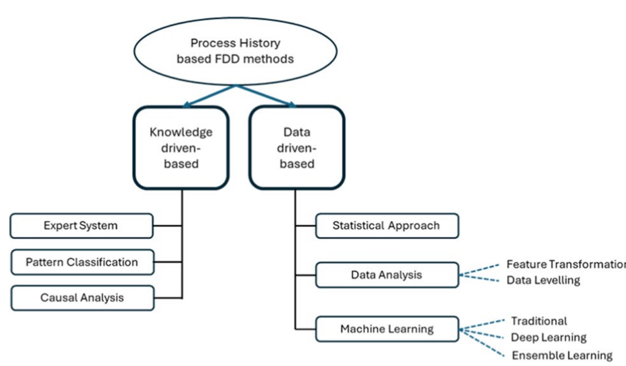


Fig. 1 Process History-based classification scheme for fault detection and diagnosis methods for HVAC units

[59]. This type of learning is often divided into classification (for discrete data) and regression (for continuous data) and is highly interpretable, which contributes to its reliability [34]. Supervised learning techniques are widely used in fault detection and diagnosis, in the HVAC domain, common supervised models include Multi-Layer Perceptrons (MLPs) [1, 56, 58], Convolutional Neural Networks (CNNs) [21, 42, 44, 50], and Recurrent Neural Networks (RNNs) [63]. Some studies combine MLPs with other models such as regression trees [58]. For example, Aguilar et al. [1] developed an autonomous data analysis cycle using Random Forests (RF) and linear regression for binary classification in HVAC systems, followed by MLP and RF for behaviour prediction. While linear regression is useful in some cases, it often struggles with nonlinear FDD problems, making logistic regression or nonlinear support vector regression (SVR) more effective [64]. Neural networks, such as artificial neural networks (ANN), are frequently used in FDD for analysing residuals-the difference between measured and predicted values [64]. However, determining appropriate thresholds for fault isolation is challenging [19]. Bayesian algorithms are common in supervised learning, including Bayesian classifiers [18, 38], diagnostic Bayesian networks [61, 62], and Naive Bayes combined with decision trees and RF [71]. Other techniques include decision trees [57, 66], hybrid RF-SVM models [65], and less common models such as extreme gradient boosting (XGBoost) [6, 80], supervised autoencoders (SAE) [76], and hidden Markov models (HMM) [40]. SVM, commonly used for classification, excels at finding optimal separating hyperplanes, for example, Namburu et al. [52] used SVM, Principal Component Analysis (PCA), and Partial Least Squares (PLS) to detect faults in chiller systems. In [11, 29], multi-label SVM (ML-SVM) was used to detect simultaneous faults in HVAC systems. SVM with Multiscale Principal Component Analysis (MSPCA) have also been applied for feature extraction and fault diagnosis [26]. Deep learning, particularly deep neural networks (DNNs), is gaining attention in FDD due to its ability to handle limited labelled data [35]. However, challenges such as automatic feature selection can omit essential data and reduce accuracy. Many supervised learning methods rely on faulty samples from older components, leading to reduced accuracy in real-world conditions [51]. Additionally, these methods are less effective in transient states, limiting their applicability [2].

2.2 Unsupervised learning

Unsupervised learning identifies patterns in unlabelled data, focusing on relationships between data points rather than predicting outcomes. It is divided into clustering (grouping similar data points) and association (finding rules that explain data structures). Common clustering methods include k-means [13], Ward's linkage, and Gaussian mixture models (GMM) [15, 28, 46], while association rule mining (ARM) techniques include FP-growth, Apriori, and cSpade [55, 66, 77, 78]. Other methods such as Slow Feature Analysis (SFA) [80], conditional Wasserstein Generative Adversarial Nets [70], and autoencoders with LSTM [4] have also been used. In FDD, unsupervised learning helps uncover hidden structures in unlabelled data [12]. ARM identifies frequent patterns, revealing inherent regularities, as seen in Yu et al. [75], who used ARM to detect faults in ventilation systems. Clustering is also useful for understanding system behaviours and detecting anomalies. Reddy et al. [33] used clustering to adjust energy consumption models, while Xue et al. [67] combined clustering with ARM to identify seasonal patterns and detect faults in district heating systems. However, unsupervised methods face limitations, such as generating vast amounts of redundant rules in ARM, which require post-mining to filter important insights. Additionally, clustering struggles when new faults fall outside predefined groups, and complex relationships between features pose challenges compared to supervised methods [82]. Self-organized neural networks speed up the learning process [37], but more research is needed to apply these methods at scale [70].

2.3 Semi-supervised learning

Semi-supervised learning combines small amounts of labelled data with larger amounts of unlabelled data, improving accuracy over unsupervised methods while reducing time and costs compared to supervised learning. Generative Adversarial Networks (GANs) are often used to generate synthetic datasets, balancing class sizes and increasing data availability [17, 23, 27, 60, 69]. Single-class classifiers, such as SVMs trained with GAN-generated healthy datasets, are also popular [47, 60]. Semi-supervised learning can be combined with unsupervised methods like ARM or supervised approaches like Classification and Regression Trees (CART) [55]. For instance, one approach used a multi-class SVM model with a novel feature extraction technique to detect system faults [16]. In FDD, semi-supervised learning is particularly useful when fault data is scarce. It begins with non-fault class labels and gradually incorporates faulty data through iterative updates, outperforming supervised learning with limited fault data, though at a higher computational cost [2, 14]. Yan et al. [68] demonstrated that semi-supervised SVM achieved 93% accuracy using only 6.66% faulty samples, outperforming other methods like KNN and CART. To enhance performance, GAN extensions can be used to generate more faulty samples for systems like AHUs, improving accuracy but requiring parameter adjustments for different configurations [70].

2.4 Hybrid learning

Hybrid models, or grey-box models, combine white-box (physics-based) and black-box (data-driven) approaches, addressing the limitations of each. Physics-based models, such as TRNSYS, can model building thermodynamics, but obtaining accurate building data can be difficult. By combining this with data-driven approaches, where measurements supplement sensory data, model complexity is reduced, and performance improves [72]. Hybrid approaches enhance resilience by reducing noise and uncertainty, often outperforming purely data-driven techniques [39]. Many studies use physics-based models to generate synthetic data, followed by supervised learning for fault prediction [6, 18, 24–26, 54, 57]. For instance, Chintala et al. [9] used a simple physical model combined with a Kalman filter to detect equipment deterioration, while Dowling and Zhang [19] found that neural networks outperformed Bayesian classifiers for fault detection. Hybrid models can also integrate knowledge discovery and pattern recognition techniques like fuzzy logic [10], clustering [66], and rule mining [77, 78]. Visualization tools like the Structural Similarity Index Measure (SSIM) combined with RadViz visualizations can also detect anomalies in HVAC systems [53]. In FDD, hybrid models combining supervised and unsupervised techniques have proven more accurate. Du et al. [20] used neural networks and clustering to enhance energy efficiency and occupant comfort in HVAC systems by detecting anomalies. Another study combined CART and k-means with density-based clustering to reduce false positives in power consumption data [5]. Despite their advantages, hybrid models can inherit the limitations of both supervised and unsupervised methods, such as increased computational cost. However, with careful design, they offer improved accuracy and efficiency in FDD processes, making them particularly promising for large-scale applications [36].

2.5 Ensemble learning

Several ensemble learning models have been applied to FDD in building HVAC systems, demonstrating effective performance across various applications. An Ensemble Diagnostic Model (EDM) using majority voting with KNN, SVM, and RF was developed to identify faults such as refrigerant leakage and lubricant excess in refrigeration systems, achieving a diagnostic accuracy of 99.88% on ASHRAE data [30]. Tree-based ensemble methods, including LightGBM-MEWMA-a combination of Light Gradient Boosting Machine (LightGBM) and Multivariate Exponentially Weighted Moving Average (MEWMA)-have been applied to centrifugal chillers. This approach identified seven common faults in chillers using ASHRAE RP-1043 data with an accuracy of 82.3% [73]. Additionally, a stacking ensemble learning method was developed using four single models: PCA, One-Class SVM, K-Means, and autoencoder. This approach demonstrated high generalization capabilities when applied to sensor faults in a Ground Source Heat Pump (GSHP) system, working with both real and simulated data [43]. Another ensemble approach integrated four models to detect seven chiller fault types from ASHRAE RP-1043 data, eliminating 84% of redundant features and focusing on optimal feature selection for fault detection [3]. For variable refrigerant flow (VRF) systems, an ensemble model combining a Variational Autoencoder (VAE) with a Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) was designed to address data imbalance challenges in fault detection, though it used laboratory-generated data rather than real-world data [79]. Further, GAN combined with bagging and boosting were employed to reduce model variance and bias. This model integrated diagnostic outputs from multiple models and achieved 98.54% accuracy in detecting five types of faults in AHUs, VAV-Box terminal devices, and distribution equipment [74].

2.6 Contribution

Methods from literature targets specific faults in specific datasets, demonstrating the flexibility of ensemble learning for FDD tasks. In contrast, our model operates on real-world terminal unit data without a predetermined fault type, focusing on identifying any abnormal patterns to support building operators in their diagnostic efforts and streamline their analysis.

Our research analyses data from 730 TUs in a commercial building in London to evaluate their performance and identify potential faults. By examining a month's worth of operational data, we can assess whether TUs are optimally positioned. For this study, data from the winter month of January was used. This selection was made to specifically analyze heating and cooling patterns when HVAC usage is higher in UK buildings. This analysis also helps identify factors contributing to performance issues, such as malfunctions or inadequate placement. Diagnosing the root causes of anomalies aims to reduce unnecessary energy consumption and improve system reliability.

Employing popular machine learning models, we analyze the operational behaviour of TUs, assessing them both individually and collectively to uncover behavioural patterns that inform optimal positioning decisions. Despite lacking precise TU locations, we investigate their performance within the building.

Through this historical data-driven approach, we aim to provide insights for building management and retrofitting efforts, ultimately enhancing HVAC system reliability and indoor environmental quality. The expected outcomes include remote identification of TU faults to support operational efficiencies and contribute to the broader goal of creating smart, responsive building environments.

The primary motivation of this study is to analyze the diverse operational patterns of terminal units (TUs) within a commercial building. By understanding these behavioral patterns, building engineers can make informed decisions to optimize TU placement during retrofitting efforts. This, in turn, can minimize unnecessary energy consumption, enhance system efficiency, and improve overall indoor environmental quality. Our findings aim to provide actionable insights that support smarter, data-driven building management practices.

3 Overview of terminal unit and challenges in its operation

A Terminal Unit (TU) also known as a fan-coil-unit (FCU) is a ceiling-mounted device generally located in rooms, corridors, and open areas, controlled by local thermostats. It contains a heating coil, a cooling coil, and a fan or damper. The unit recirculates air by drawing in either room air or a combination of fresh and recirculated air, adjusting the air release based on the thermostat's settings. Figure 2 shows the external structure and schematic of a TU. In a typical setup, a central chiller and boiler plant supplies chilled water to all cooling coils and hot water to all heating coils. When the temperature rises, the local thermostat detects the increase and triggers the chilled water valve to send cool water through the cooling coil, where the fan blows cool air into the space. Conversely, if the room becomes too cold, the thermostat activates the heating coil, blowing warm air until the room temperature reaches the desired set point.

Temperature inconsistencies in terminal units often arise due to issues like clogged plenums, dirty filters, and unclean coils, which increase resistance, reduce airflow, and lead to ineffective heating or cooling. Several air distribution problems can also impair performance.

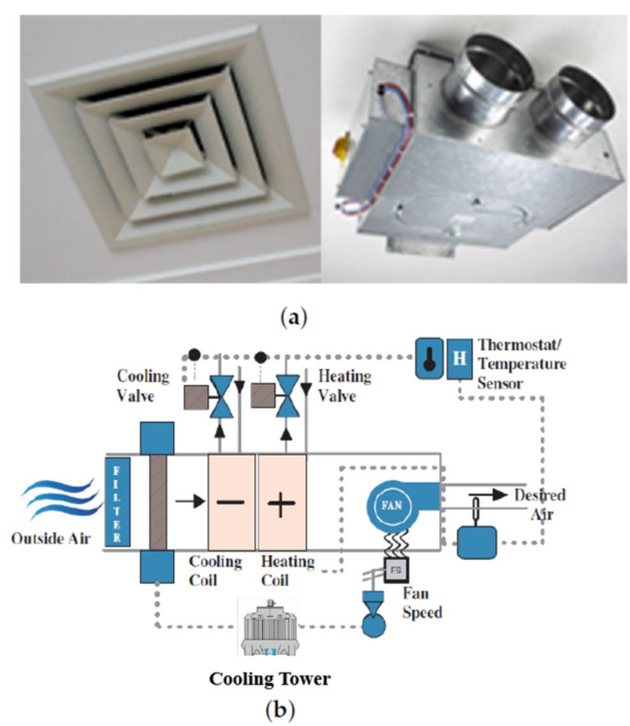


Fig. 2 **a** Inner and outer structure of a TU. **b** Schematic of a TU [15]

Another common issue is suboptimal thermostat placement during the building's construction. Poor thermostat positioning can result in inaccurate temperature readings, leading to excessive power use to reach the setpoint. For example, if a thermostat is installed near a refrigeration area, it may continuously detect a cooler temperature than the actual room temperature. This misreading can cause the system to overcompensate, running unnecessarily and consuming extra energy to reach the desired setpoint.

In this paper, we focus on data-driven analysis of Terminal Unit (TU) data through statistical methods and ensemble machine learning to identify optimal placement issues and detect, classify, and categorize TU operational faults. This study utilizes real-world data from a commercial building in London, UK, to enhance HVAC system efficiency and reliability.

4 Proposed methodology

This paper presents a dual approach, focusing on both exploratory data analysis (EDA) to assess thermostat placement effectiveness and data-driven machine learning to classify control temperature faults in Terminal Units (TUs). Each aspect addresses a specific challenge in HVAC system optimization, enabling better energy management, occupant comfort, and overall system reliability.

In the first part, exploratory data analysis (EDA) is utilized to evaluate the placement and performance of thermostats across the building. Descriptive statistics, such as distributions, means, and standard deviations, are calculated for each TU to reveal patterns in their temperature regulation behaviour. Analyzing these statistical measures provides insights into TU performance variability and identifies any deviations that may indicate suboptimal thermostat placement. This analysis serves as a diagnostic tool, shedding light on operational consistency across the building over the month-long period of data collection.

In the second part, machine learning is applied to classify abnormality in TU control temperature settings. By using the labelled information, which categorizes TU operation as a normal, heating fault, or cooling fault, the ML models automate fault detection, improving the speed and accuracy of identifying operational inefficiencies. Together, these methods support a holistic approach to enhancing HVAC functionality, focusing both on existing setup efficiency and proactive fault management.

4.1 Data description

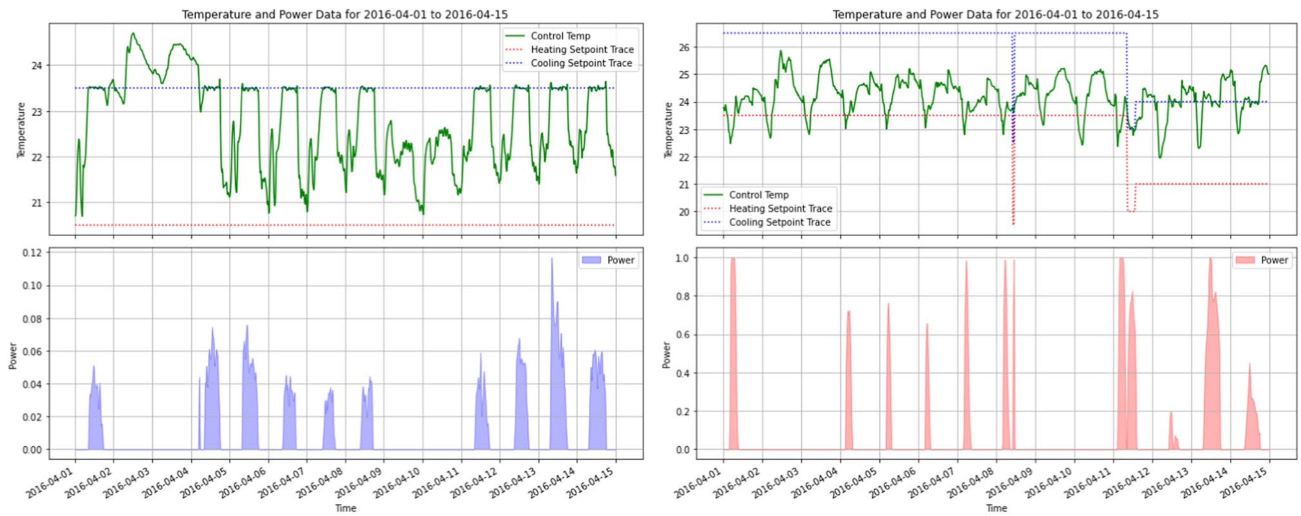
This study focuses on a retrofitted commercial building in London, where data from the building's Terminal Units (TUs) were collected and analyzed. This multi-floor building spans seven levels and includes 730 TUs distributed across these floors. For this case study, data for one month from each TU, with measurements recorded every 10 min were gathered. The data was collected using a data acquisition device developed by engineers from a UK-based company. The data collection and processing methods have been detailed in the authors' previous works [15, 17].

Each TU monitors 26 distinct parameters, providing comprehensive insights into the system's performance. Given this frequency and range of parameters, each TU contributes approximately 560,000 data entries over the month, amounting to a total dataset of around 410 million data points. This extensive dataset enables a detailed analysis of TU behaviour, facilitating accurate fault detection, energy use assessment, and identification of areas for operational improvement in the building's HVAC system.

Figure 3a, b illustrate two distinct behavioral patterns of different TUs over a 15-day period. In Fig. 3a, the control temperature exceeds the cooling setpoint, leading to an increase demand for cooling power. Conversely, Fig. 3b shows the control temperature dropping below the heating setpoint, which increases heating power demand. These TU data patterns have been analyzed in this research to identify TUs that persistently operate outside the deadband, resulting in excessive and abnormal power demand.

4.2 Feature transformation and data labelling

The raw TU data underwent pre-processing and feature transformation to prepare it for machine learning applications. Initially, each file contained only three columns: date, parameter, and value, resulting in unstructured, unlabeled data unsuitable for direct model training. To address this, each TU data file was transformed by setting time as a constant



(a) Control temperature and its cooling power demand.

(b) Control temperature and its heating power demand.

Fig. 3 The sample TU data of cooling and heating pattern's for 15 days period

column and pivoting the various parameters into separate columns, each representing a distinct feature with its corresponding values. Missing values within these features were addressed using an interpolation method to ensure data continuity.

For data labelling, we focused on several key features critical to temperature regulation. The "control temperature" feature reflects the current room temperature, while the "cooling setpoint" and "heating setpoint" features indicate whether heating or cooling is required. Another important feature, the "deadband," defines the allowable range of variation for the control temperature. Based on these features, the data was labelled as follows:

- Class 0 (Normal Operation): When the control temperature remains within the deadband range, the system is considered to be functioning normally.
- Class 1 (Heating Fault): When the control temperature falls below the heating setpoint, the system demands heating, raising the control temperature to reach the desired range.
- Class 2 (Cooling Fault): When the control temperature exceeds the cooling setpoint, the system activates cooling to bring the control temperature back within the acceptable range.

Using this classification approach, the entire dataset was labelled into three categories: Class 0 for normal operation, Class 1 for heating faults, and Class 2 for cooling faults. Examples of each class are shown in Fig. 4. Figure 4a illustrates the expected behaviour for Class 0, representing non-faulty or normal TU operation, where the control temperature remains within the deadband limit, resulting in minimal or no power demand. Figure 4b shows the behaviour for Class 1, which indicates an abnormal TU operation associated with heating faults. Here, the control temperature exceeds the heating setpoint, triggering an increased demand for heating power. Similarly, Fig. 4c represents Class 2, another abnormal TU operation type, associated with cooling faults. In this case, the control temperature surpasses the cooling setpoint, resulting in a higher cooling power demand.

Since this operational information was not initially provided by the building operator, labelling was essential to identify and categorize these faults accurately. After labelling, machine learning models were applied to automate fault detection across the dataset.

4.3 Machine learning

After transforming and labelling the raw data, three machine learning models were applied to classify TU behaviour into three distinct operational states (classes). The chosen models include a traditional multi-class Support Vector Machine (SVM) and three ensemble models: Random Forest (RF), andAdaBoost. Each model's performance was evaluated, focusing

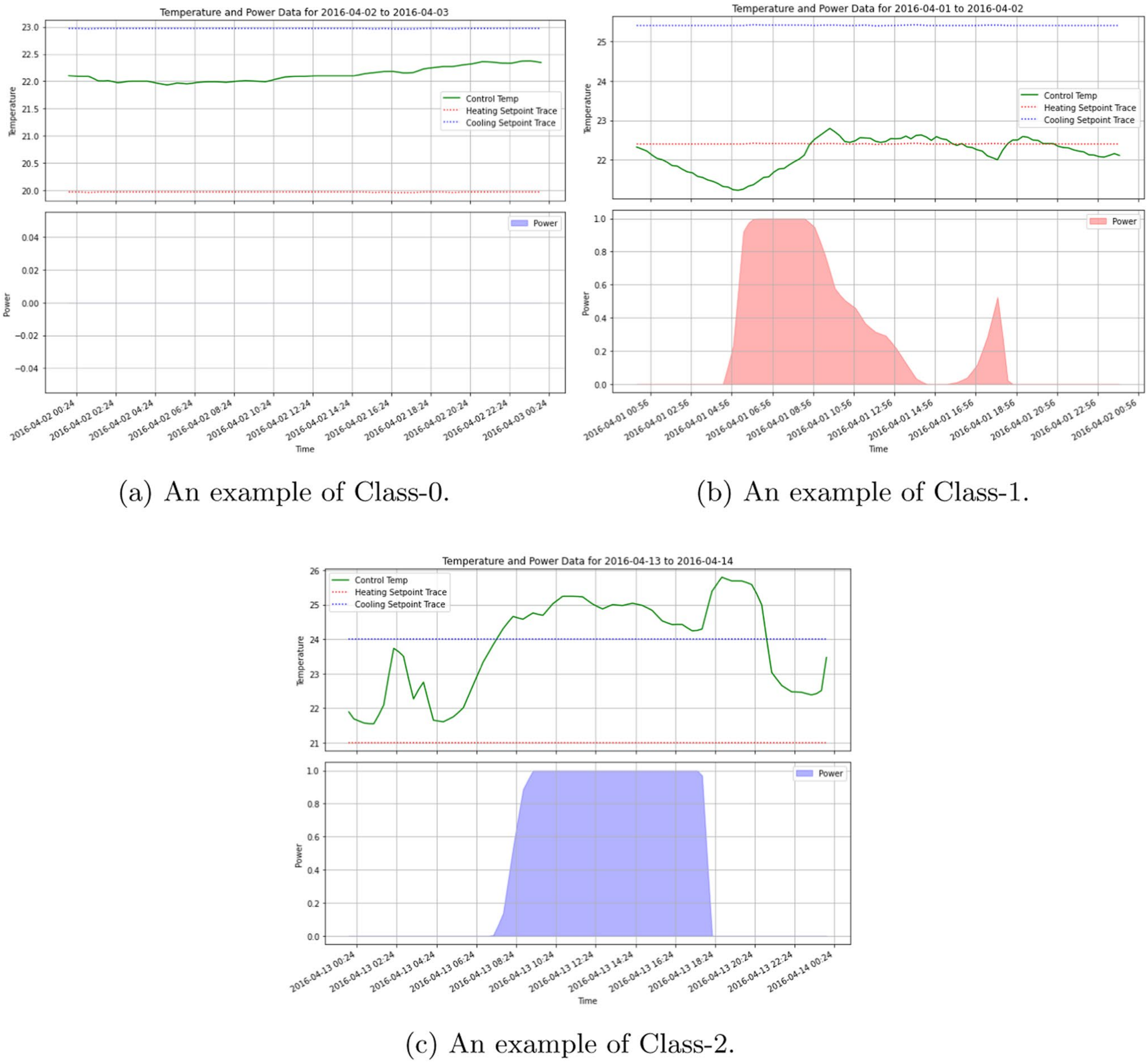


Fig. 4 An example of daily TU behaviour for each of the three classes

on both classification accuracy and computational efficiency to determine the most effective approach for TU fault detection. An overview of each model is detailed below.

Support Vector Machine (SVM): SVM is a supervised learning algorithm that works by finding a hyperplane that best separates data points into classes. In multi-class SVM, a one-vs-rest or one-vs-one approach is used to extend binary SVM classification to multiple classes. The SVM classifier aims to maximize the margin between classes, making it robust for fault detection tasks with distinct behavioural classes. The decision function for SVM is defined as:

$$f(x) = w \cdot x + b \tag{1}$$

where w represents the weights of the hyperplane, x is the input feature vector, and b is the bias term. The objective is to maximize $\frac{1}{\|w\|}$, the margin distance between classes.

SVM performs well with complex decision boundaries and is computationally efficient with small to medium-sized datasets. It serves as a benchmark for assessing performance against ensemble methods in identifying TU fault patterns.

Random Forest (RF): Random Forest is an ensemble method that combines multiple decision trees, each trained on random subsets of the data and features. It uses a majority vote across the trees to make final predictions, reducing overfitting and enhancing accuracy for varied data patterns. The decision function for Random Forest can be represented as:

$$f(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \tag{2}$$

where T is the number of trees, and $h_t(x)$ represents the prediction from the t-th tree.

Random Forest's robustness against overfitting and high interpretability make it well-suited for fault detection, where the underlying patterns can vary. Its ability to handle large, complex datasets and non-linear relationships is ideal for the multi-parameter nature of TU data.

AdaBoost: AdaBoost (Adaptive Boosting) is a boosting technique that sequentially trains weak learners, typically decision trees, by emphasizing misclassified instances in each iteration. This adaptive weighting makes the model sensitive to difficult cases, often resulting in high accuracy for multi-class classification. The prediction function is given by:

$$f(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \tag{3}$$

where T is the number of weak learners, α_t is the weight of each learner based on its error, and $h_t(x)$ is the t-th learner's prediction.

AdaBoost's ability to adjust to complex data distributions and prioritize challenging instances helps in detecting subtle anomalies in TU behaviour. This makes it valuable in fault detection where certain patterns may only slightly deviate from the norm.

In summary, these models were selected to compare traditional and ensemble approaches in detecting TU faults, with SVM providing a reliable baseline and the ensemble models offering robust alternatives for handling large-scale, multi-parameter data. The ensemble models' strengths in managing complex interactions and their general resilience to overfitting make them particularly advantageous for TU fault detection and classification.

To further evaluate model performance, the results were also compared using a balanced dataset. A data balancing technique was applied to the TU dataset to upsample the labels and address class imbalance. Specifically, Synthetic Minority Over-sampling Technique (SMOTE) was used to redistribute the uneven class before training and testing all ML models. SMOTE is a popular method for handling imbalanced datasets. It works by creating synthetic samples of the minority class rather than simply duplicating existing ones. This helps the model learn better from underrepresented classes and improves classification performance. Thereafter, model validation was conducted to check whether balancing the data improved performance and to understand its impact on the results.

4.4 Model evaluation

To evaluate the effectiveness of these implemented ML models in detecting faults in Terminal Units (TUs), three key performance metrics: Precision, Recall, and F1-score were used. These metrics were chosen to provide a well-rounded assessment of each model's performance, particularly in handling fault detection scenarios where distinguishing between fault and non-fault states is critical. All these measures are derived from the confusion matrix and the outcomes have been analysed in the Result Analysis section.

Precision: Precision measures the accuracy of the positive predictions, indicating the proportion of true positives out of all instances classified as positive. High precision ensures that most detected faults are genuine, thereby reducing unnecessary interventions.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{4}$$

Recall: Recall assesses the model's ability to detect actual faults, showing the proportion of true positives out of all actual positives. A high recall value indicates that the model successfully captures most faults, minimizing missed detections. This is essential in TU fault detection to ensure that all operational issues are identified, preventing potential energy inefficiencies or comfort disruptions due to undetected faults.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{5}$$

F1-Score: The F1-score provides a harmonic mean of precision and recall, offering a balanced measure that considers both false positives and false negatives. This score is used to find out the balance between precision and recall, as in TU fault detection, both accurate detection and comprehensive fault coverage are necessary for reliable operation.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

By comparing these metrics across models, we can determine which model not only detects faults with high accuracy but also achieves a balance between correctly identifying faults (high recall) and avoiding false alarms (high precision). This comprehensive evaluation helps in selecting the most effective model for automated TU fault detection, ultimately supporting a reliable and efficient HVAC system.

5 Result analysis and discussion

This section provides a detailed analysis of the results obtained from the applied ML models. The experiment used one month of data from 730 TUs, with each TU providing 2783 rows of data recorded at 10-minute intervals. After preprocessing and removing missing values, a total of 2,002,173 data points remained for analysis. The TU dataset used in this study consists of time-series data, where measurements of control temperature, power consumption, and other related parameters were collected at 10-min intervals. This structured data captures temporal variations in system behavior, and provides a continuous stream of observations over time, enabling models to learn patterns and identify potential faults in TUs. The details about the data were discussed in Sect. 4.1.

Following the labelling process, the dataset was labelled into three classes based on TU operational states:

- Class 0 (Normal Operation): Comprising 1,180,337 rows, this class represents approximately 59% of the data, capturing instances where TUs' control temperature operated within the deadband range.
- Class 1 (Heating Fault): Consisting of 520,263 rows, or 29% of the data, this class includes cases where TUs required heating adjustments.
- Class 2 (Cooling Fault): Representing 15% of the data with 295,990 rows, this class identifies cases where TUs needed cooling adjustments.

This dataset has been used for analyzing the performance of each model in classifying TU behaviours across different fault states, ultimately aiding in the identification and resolution of operational issues in HVAC systems.

5.1 EDA results

This section employs Exploratory Data Analysis (EDA) with distribution analysis to identify operational trends, deviations, and anomalies in TU control temperatures over a one-month period. Figure 5 presents temperature variations of selected 730 TUs, helping assess normal operation or potential faults. If the temperature stays within the deadband levels 20–25 °C, the TU operates normally. However, frequent deviations indicate possible issues. If peaks in the distribution align with expected setpoints, the TU is stable; otherwise, significant peaks outside this range suggest faults requiring further inspection.

The first two patterns, shown in Fig. 5a, b, exhibit a normal distribution, indicating that the control temperature generally remains within the deadband limit. However, occasional deviations beyond this range are observed, after which the temperature returns to normal. This suggests that the system self-corrects periodic fluctuations over time. In contrast, Fig. 5c, d display a skewed distribution, where the control temperature frequently drifts towards the cooling setpoint, surpassing the deadband range. This pattern signals potential cooling faults in the TU's operation, necessitating further inspection of the TU or thermostat placement. Similarly, Fig. 5e, f depict a uniform distribution, where the control temperature fluctuates continuously between 20–25 °C without stabilization. Such behavior suggests potential operational inefficiencies, requiring a reassessment of the TU's positioning or functionality. Figure 5g, h also exhibit distribution patterns with distinct peaks. In Fig. 5g, the temperature primarily remains within the required range, with only

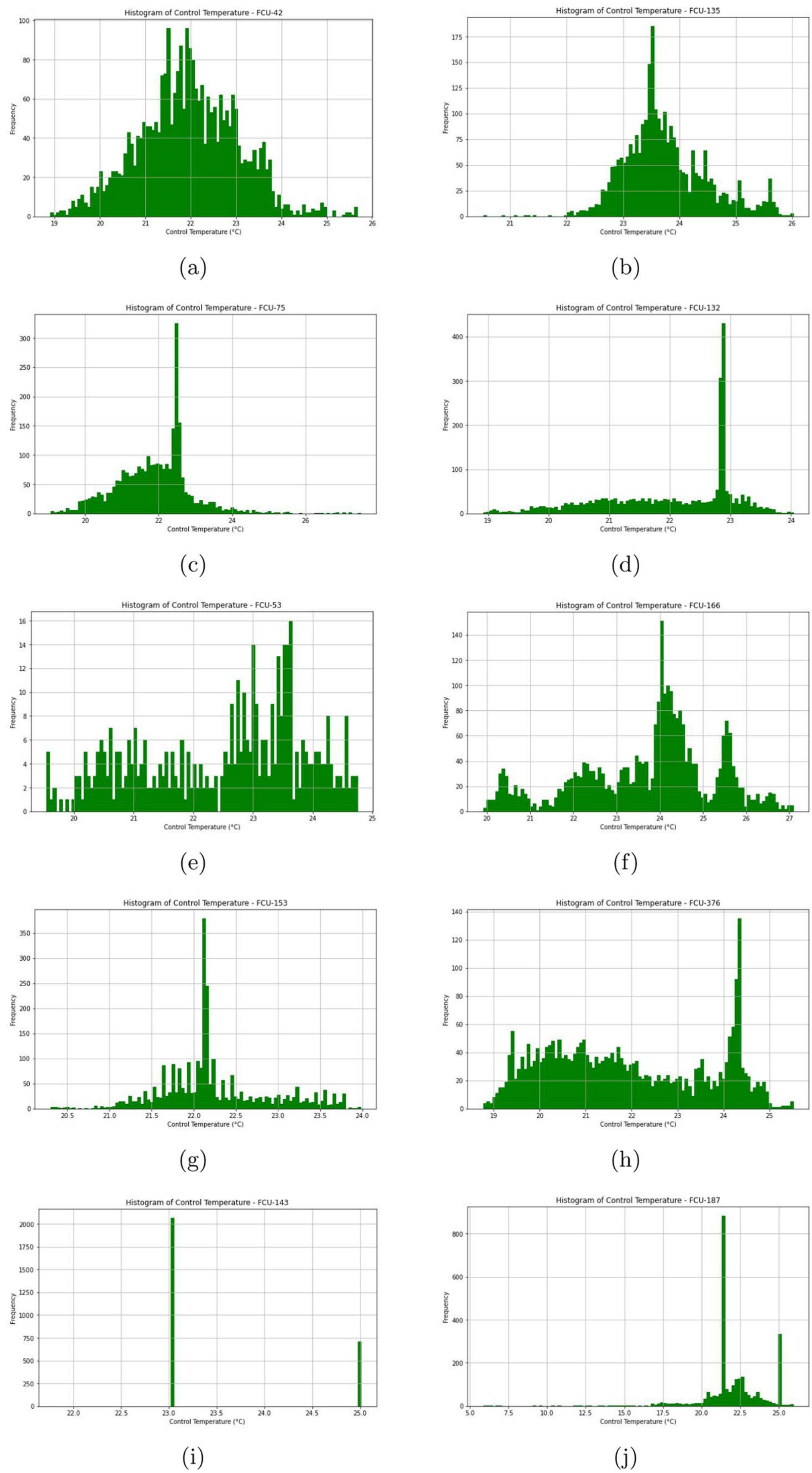


Fig. 5 Monthly data distribution of control temperature patterns for 10 different terminal units (TUs), **a**–**j** represents the temperature variation of an individual TU, highlighting operational trends and potential anomalies

occasional variations, indicating stable TU operation. However, Fig. 5h is right-skewed, meaning the temperature, while mostly within limits, frequently reaches the cooling setpoint, suggesting a gradual shift towards higher temperatures. Finally, Fig. 5i, j illustrate bi-modal distributions, significantly deviating from expected patterns. Figure 5i shows a stable temperature for prolonged periods, reflecting minimal fluctuations. In contrast, Fig. 5j presents a left-skewed binomial distribution with sharp peaks and a low control temperature around 5 °C, signaling potential anomalies. This TU behavior warrants further investigation to determine whether the issue stems from mechanical faults or external factors affecting temperature control.

After analyzing the TU distributions, the standard deviation was calculated from each TU distribution, as shown in Fig. 6. The red dotted line represents a 1.5-sigma shift, a statistical adjustment accounting for natural process drift over time. TUs with standard deviation values exceeding this threshold may require further investigation by operators, as these deviations could indicate suboptimal placement or operational issues that prevent the TU from functioning as expected.

5.2 ML results

The dataset was initially prepared by transforming it into a columnar format, imputing missing data points, and labeling the data for machine learning, as detailed in Sect. 4.2.

To implement machine learning, the dataset was split in two stages to ensure robust training, validation, and testing. First, 80% of the data was allocated for training and validation, while the remaining 20% was set aside for testing. This two-stage split strategy provides effective model training, fine-tuning, and thorough testing, allowing for an accurate performance assessment on an independent dataset and helping to prevent overfitting. During the training and validation stage, a fivefold cross-validation approach was used, followed by testing on the reserved dataset.

The SVM model results are presented in Table 1, where precision, recall, and F1-score are calculated for each class, along with overall model accuracy. As observed, precision is high for class-0 (normal class) but lower for class-1 and class-2, representing heating and cooling faults. Conversely, recall is lower for class-0 but higher for class-1 and class-2, indicating a better ability to detect fault classes accurately rather than detecting the non-faults properly.

The model achieved an average training accuracy of 50.094% and a testing accuracy of 46.998%. To optimize performance, we experimented with various Support Vector Machine (SVM) kernel functions, including linear and Gaussian (RBF) kernels. Among these, the linear kernel demonstrated the best performance, which is why the results presented here reflect the linear kernel application only.

To enhance classification accuracy in detecting TU faults, two ensemble models were implemented in addition to the traditional machine learning model. Ensemble methods, known for combining multiple decision models to improve predictive performance, offer a more robust solution for fault detection by reducing variance and enhancing generalization. These ensemble models provide a comprehensive approach to refining classification accuracy for complex fault detection scenarios, ensuring improved reliability and resilience over traditional ML techniques.

Similar to the SVM approach, the Random Forest algorithm was applied to the same datasets using the same data partitioning process. The results are presented in Table 2, which demonstrates a significant improvement over the traditional machine learning methods. The precision, recall, and F1-score achieved values exceeding 99%, with overall training and validation accuracy at 99.94% and testing accuracy at 99.89%.

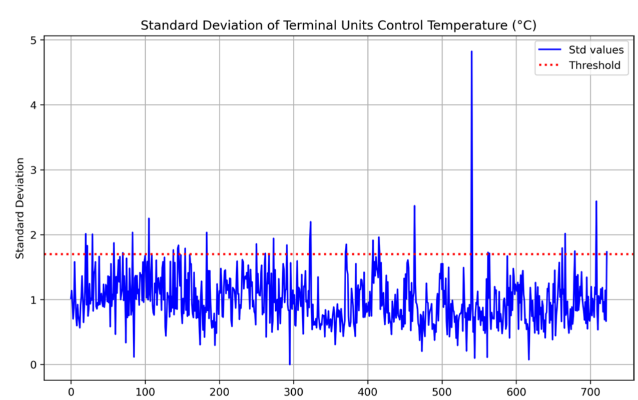


Fig. 6 Standard deviation plot of monthly data from 730 terminal units (TUs)

Table 1 The evaluation matrices for fivefold validation and testing result for the SVM model

Training and validation results	Performance metrics	Class-0	Class-1	Class-2
Fold-0	Accuracy	63.629		
	Recall	37.936	99.995	98.522
	Precision	99.448	67.166	85.879
	F1 Score	54.848	68.455	71.520
Fold-1	Accuracy	68.322		
	Recall	45.581	99.988	99.97
	Precision	99.989	62.356	95.365
	F1 Score	62.616	65.423	88.795
Fold-2	Accuracy	70.468		
	Recall	49.231	99.873	99.988
	Precision	99.920	63.791	96.631
	F1 Score	65.953	66.396	91.590
Fold-3	Accuracy	75.051		
	Recall	57.729	98.689	99.988
	Precision	99.174	77.216	90.749
	F1 Score	72.926	75.220	79.807
Fold-4	Accuracy	50.094		
	Recall	14.344	99.988	99.984
	Precision	99.994	62.100	74.111
	F1 Score	25.088	65.197	58.603
5-Folds average	Accuracy	50.094		
	Recall	40.964	99.712	99.695
	Precision	99.705	66.506	88.547
	F1 Score	56.286	68.138	78.063
Testing results	Accuracy	46.998		
	Recall	15.479	99.988	99.988
	Precision	99.988	55.219	77.937
	F1 Score	26.808	59.446	56.607

To further validate the improvement in results and confirm the efficacy of ensemble models on this type of dataset, the AdaBoost algorithm was applied, with the same data partitioning strategy used as before. The results are detailed in Table 3.

The findings indicate that AdaBoost outperformed the traditional machine learning model. Precision values were found to be lower for class 0 compared to recall and F1-score, while for classes 1 and 2, precision was higher. Conversely, recall was highest for class 1 but lower for class 2, leading to a continuous decrease in F1-score from class 0 to class 2. This trend suggests that the model may be experiencing underfitting, likely due to the abundance of class 0 data in the dataset. Nevertheless, the training, validation, and testing accuracy reached approximately 85%, indicating that while the model performs better than the SVM, it still falls short compared to the Random Forest model.

For further insights, a confusion matrix for testing outcomes is provided to illustrate the SVM model's classification performance across different classes as shown in Fig. 7a. Similarly, to provide insights into the RF model's performance, confusion matrix shown in Fig. 7b, and the Adaboost's confusion matrix is presented in Fig. 7c.

The confusion matrix provides a detailed view of the model's classification performance, showing true and false class assignments across all three classes: normal operation, heating faults, and cooling faults. It highlights where the model accurately identified each class and where misclassifications occurred, aiding in the evaluation of the model's reliability in fault detection for TUs. This analysis is instrumental in diagnosing areas of strength and identifying potential improvements, ensuring more effective classification between normal and faulty states.

Table 2 The evaluation matrices for fivefold validation and testing result for the RF model

Training and validation results	Performance metrics	Class-0	Class-1	Class-2
Fold-0	Accuracy	99.946		
	Recall	99.952	99.949	99.919
	Precision	99.938	99.981	99.984
	F1 Score	99.954	99.948	99.915
Fold-1	Accuracy	99.945		
	Recall	99.953	99.944	99.918
	Precision	99.934	99.976	99.989
	F1 Score	99.953	99.938	99.928
Fold-2	Accuracy	99.94		
	Recall	99.954	99.929	99.909
	Precision	99.922	99.981	99.984
	F1 Score	99.949	99.938	99.912
Fold-3	Accuracy	99.942		
	Recall	99.95	99.938	99.919
	Precision	99.931	99.973	99.989
	F1 Score	99.95	99.932	99.929
Fold-4	Accuracy	99.942		
	Recall	99.954	99.932	99.911
	Precision	99.924	99.978	99.987
	F1 Score	99.95	99.936	99.92
5-Folds average	Accuracy	99.942		
	Recall	99.953	99.939	99.915
	Precision	99.93	99.978	99.987
	F1 Score	99.951	99.938	99.921
Testing results	Accuracy	99.888		
	Recall	99.912	99.879	99.782
	Precision	99.846	99.946	99.984
	F1 Score	99.91	99.857	99.835

5.3 SMOTE results

The Synthetic Minority Over-sampling Technique (SMOTE) was applied to the TU dataset, increasing the number of data points from approximately 160 K to 270 K. After balancing the data, SVM, Random Forest (RF), and AdaBoost were applied, and their results were evaluated using accuracy, precision, recall, and F1 score metrics. The training and testing outcomes are presented in Table 4.

The findings show that SVM's accuracy improved from 50 to 82% in training and from 46 to 64% in testing. Additionally, recall for class-0 (non-faulty class) improved in both training and testing phases, as this class had the highest number of data points. Similarly, AdaBoost also showed improved performance after applying SMOTE.

However, Random Forest (RF) outperformed over SVM and AdaBoost, for both cases with and without SMOTE. This suggests that RF is the optimal choice for TU fault classification, as real-world faults occur infrequently and often unpredictably. Since RF performs well even on the original imbalanced dataset, it is more suitable for real-world deployment, where models must detect faults without relying on artificial data balancing.

Table 3 The evaluation matrices for 5-fold validation and testing result for the AdaBoost model

Training and validation results	Performance metrics	Class-0	Class-1	Class-2
Fold-0	Accuracy	87.522		
	Recall	98.479	79.323	60.256
	Precision	72.248	98.835	99.968
	F1 Score	90.188	86.886	75.117
Fold-1	Accuracy	88.366		
	Recall	98.726	74.141	73.610
	Precision	73.939	99.734	99.354
	F1 Score	90.807	84.786	83.112
Fold-2	Accuracy	88.635		
	Recall	99.093	76.087	70.805
	Precision	74.133	99.852	99.505
	F1 Score	91.017	86.217	81.619
Fold-3	Accuracy	87.716		
	Recall	98.597	79.425	60.873
	Precision	72.564	98.923	99.973
	F1 Score	90.332	87.071	75.608
Fold-4	Accuracy	85.976		
	Recall	99.152	72.527	59.212
	Precision	67.586	99.641	99.729
	F1 Score	89.176	83.588	73.695
5-Folds average	Accuracy	85.976		
	Recall	98.809	76.310	64.949
	Precision	72.094	99.397	99.706
	F1 Score	90.304	85.710	77.830
Testing results	Accuracy	85.936		
	Recall	99.232	65.553	59.701
	Precision	63.579	99.823	99.601
	F1 Score	89.847	78.936	73.491

6 Conclusion and future work

This study highlights the effectiveness of a data-centric approach for diagnosing faults in terminal units and optimizing their placement within building HVAC systems. Since different HVAC components exhibit distinct operational characteristics and fault types, a direct comparison with these studies is challenging. Additionally, many of these works benefit from prior knowledge of fault characteristics, whereas our study operates with raw TU data, making fault identification more complex. This highlights the novelty of our approach- using ensemble ML techniques to detect TU faults without predefined fault characteristics.

By utilizing ensemble learning techniques, we achieved significant improvements in fault classification accuracy compared to traditional machine learning methods. The analysis revealed key insights into the behaviour of TUs, including the identification of suboptimal thermostat placements and the impact of class imbalance on model performance. A data balancing technique also applied to improve the model's performance. However, the RF model outperformed the others, demonstrating its robustness in handling imbalanced datasets, while AdaBoost provided valuable insights despite its lower accuracy. The incorporation of confusion matrices allowed for a deeper understanding of classification outcomes, which is crucial for refining fault detection processes. This research contributes

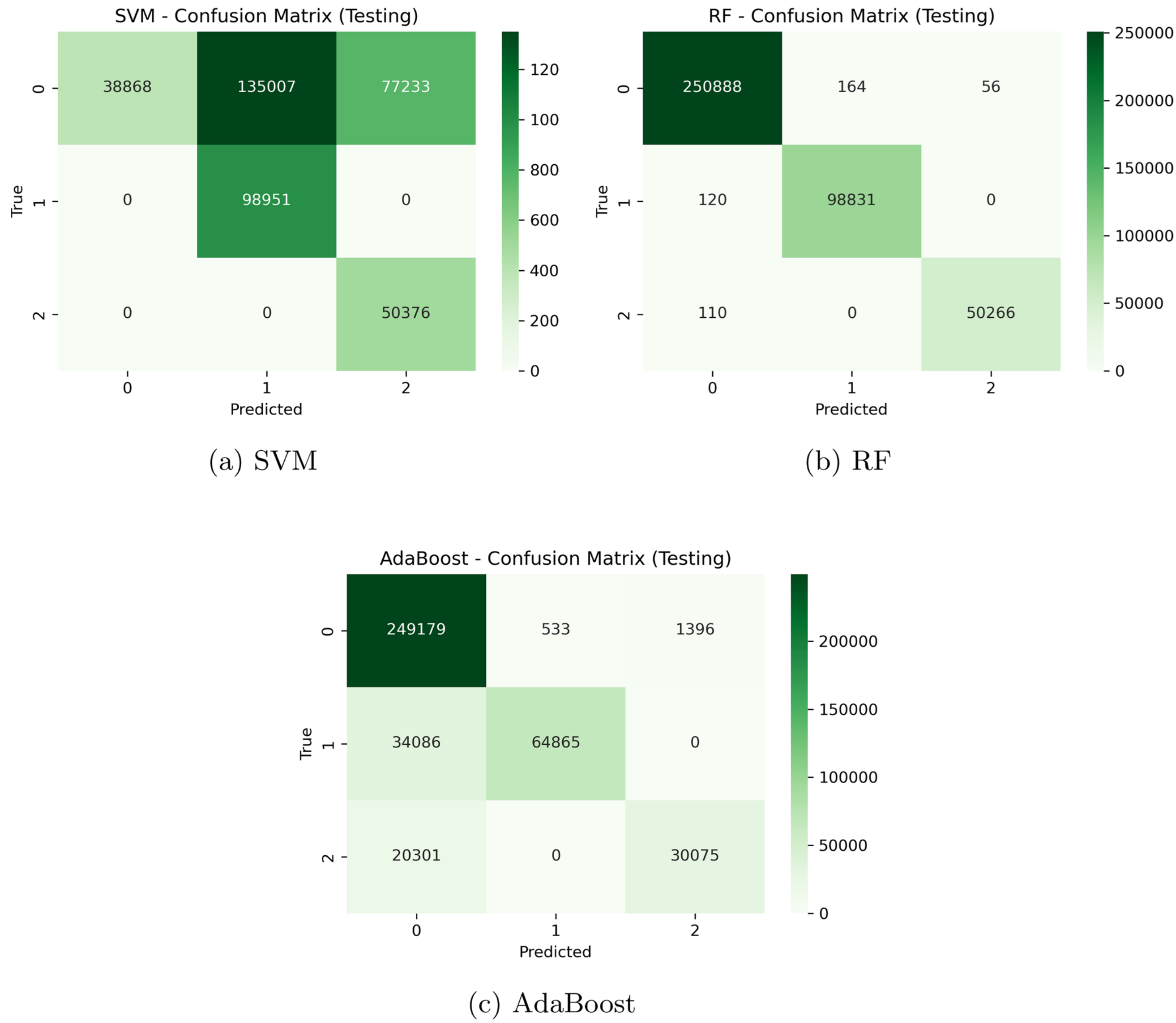


Fig. 7 Testing confusion matrix for all three compared ML models

to the ongoing effort to enhance the efficiency and effectiveness of building HVAC systems, ultimately leading to reduced energy consumption and improved indoor environmental quality.

This study relies on data from a single winter month data, without detailed location of TUs within the building. This limits the ability to analyze the impact of spatial positioning on performance, airflow distribution, and thermal comfort. Thus, future work will concentrate on incorporating more real data to capture seasonal variations or long-term trends in TU performance for improving fault detection capabilities and system performance, particularly through floor-wise analyses. Additionally, the research will explore how patterns vary across different floors and how they change in relation to the building's cardinal directions. We will also test the models with a broader and more diverse dataset to address class imbalance issues effectively.

Table 4 Comparison of training and testing results for all three experimented models after applying SMOTE

Models	Training & testing	Performance metrics	Class-0	Class-1	Class-2
SVM	Training	Accuracy	82.569		
		Recall	45.852	99.982	99.695
		Precision	99.839	77.113	95.813
		F1 Score	62.188	81.4	95.901
	Testing	Accuracy	64.352		
		Recall	43.153	100	100
		Precision	100	55.215	97.792
		F1 Score	60.289	59.444	92.875
RF	Training	Accuracy	99.97		
		Recall	99.939	99.983	99.99
		Precision	99.987	99.983	99.987
		F1 Score	99.956	99.974	99.982
	Testing	Accuracy	99.878		
		Recall	99.854	99.925	99.905
		Precision	99.918	99.932	99.953
		F1 Score	99.903	99.86	99.791
AdaBoost	Training	Accuracy	89.902		
		Recall	96.75	84.364	84.194
		Precision	84.279	99.544	98.833
		F1 Score	84.808	91.056	90.238
	Testing	Accuracy	93.16		
		Recall	99.305	82.045	84.362
		Precision	82.826	99.881	99.604
		F1 Score	94.794	89.958	90.171

Author contributions MD—Maitreyee Dey, PP—Preeti Patel, SPR—Soumya Prakash Rana MD collected the data through the building operators and conducted the data processing and transformation. SPR and MD performed the machine learning algorithms and analysed the FDD outcomes for the building TU data. PP conducted the literature review. MD managed the experiments performed, and along with co-authors. MD and PP wrote the manuscript. MD, PP, and SPR contributed to writing the paper and reviewing the manuscript. All the authors have reviewed the revised manuscript.

Funding This research received no external funding.

Data availability The data is subject to commercial restrictions, thus the building's data supporting this study cannot be made available publicly.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Aguilar J, Garcés-Jiménez A, Gómez-Pulido JM, et al. Autonomic management of a building's multi-HVAC system start-up. IEEE Access. 2021;9:70502–15.
2. Beghi A, Brignoli R, Cecchinato L, et al. Data-driven fault detection and diagnosis for HVAC water chillers. Control Eng Pract. 2016;53:79–91.
3. Bi J, Wang H, Hua M, et al. An interpretable feature selection method integrating ensemble models for chiller fault diagnosis. J Build Eng. 2024;87: 109029.
4. Bouabdallaoui Y, Lafhaj Z, Yim P, et al. Predictive maintenance in building facilities: a machine learning-based approach. Sensors. 2021;21(4):1044.
5. Capozzoli A, Lauro F, Khan I. Fault detection analysis using data mining techniques for a cluster of smart office buildings. Expert Syst App. 2015;42(9):4324–38.
6. Chakraborty D, Elzarka H. Early detection of faults in HVAC systems using an XGboost model with a dynamic threshold. Energy Build. 2019;185:326–44.
7. Chen J, Zhang L, Li Y, et al. A review of computing-based automated fault detection and diagnosis of heating, ventilation and air conditioning systems. Renew Sustain Energy Rev. 2022;161: 112395.
8. Chen Z, O'Neill Z, Wen J, et al. A review of data-driven fault detection and diagnostics for building HVAC systems. Appl Energy. 2023;339: 121030.
9. Chintala R, Winkler J, Jin X. Automated fault detection of residential air-conditioning systems using thermostat drive cycles. Energy Build. 2021;236: 110691.
10. Ciani L, Guidi G, Patrizi G, et al. Condition-based maintenance of HVAC on a high-speed train for fault detection. Electronics. 2021;10(12):1418.
11. Dey M, Data-driven remote fault detection and diagnosis of HVAC terminal units using machine learning techniques. PhD thesis, London South Bank University; 2020
12. Dey M, Gupta M, Rana SP, et al A pid inspired feature extraction method for HVAC terminal units. In: 2017 IEEE conference on technologies for sustainability (SusTech), IEEE; 2017a. p. 1–5.
13. Dey M, Gupta M, Turkey M, et al. Unsupervised learning techniques for HVAC terminal unit behaviour analysis. In: 2017 IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications. Internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), IEEE: Cloud & Big Data Computing; 2017. p. 1–7.
14. Dey M, Rana SP, Dudley S, Semi-supervised learning techniques for automated fault detection and diagnosis of hvac systems. In: 2018 IEEE 30th international conference on tools with artificial intelligence (ICTAI), IEEE; 2018. p. 872–7.
15. Dey M, Rana SP, Dudley S. A case study based approach for remote fault detection using multi-level machine learning in a smart building. Smart Cities. 2020;3(2):401–19.
16. Dey M, Rana SP, Dudley S. Smart building creation in large scale HVAC environments through automated fault detection and diagnosis. Future Generation Comput Syst. 2020;108:950–66.
17. Dey M, Rana SP, Dudley S. Automated terminal unit performance analysis employing x-RBF neural network and associated energy optimisation-a case study based approach. Appl Energy. 2021;298: 117103.
18. Dowling CP, Zhang B, Transfer learning for hvac system fault detection. In: 2020 American control conference (ACC), IEEE; 2020. p. 3879–85.
19. Du Z, Fan B, Chi J, et al. Sensor fault detection and its efficiency analysis in air handling unit using the combined neural networks. Energy Build. 2014;72:157–66.
20. Du Z, Fan B, Qiao Jin X, et al. Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis. Build Environ. 2014;73:1–11.
21. Elnour M, Meskin N. Novel actuator fault diagnosis framework for multizone HVAC systems using 2-D convolutional neural networks. IEEE Trans Autom Sci Eng. 2021;19(3):1985–96.
22. Fan C, Xiao F, Li Z, et al. Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: a review. Energy Build. 2018;159:296–308.
23. Fan C, Liu Y, Liu X, et al. A study on semi-supervised learning in enhancing performance of ahu unseen fault detection with limited labeled data. Sustain Cities Soc. 2021;70: 102874.
24. Gálvez A, Diez-Olivan A, Seneviratne D, et al. Fault detection and rule estimation for railway HVAC systems using a hybrid model-based approach. Sustainability. 2021;13(12):6828.
25. Gálvez A, Seneviratne D, Galar D. Hybrid model development for HVAC system in transportation. Technologies. 2021;9(1):18.
26. Gharsellaoui S, Mansouri M, Refaat SS, et al. Multivariate features extraction and effective decision making using machine learning approaches. Energies. 2020;13(3):609.
27. Gharsellaoui S, Mansouri M, Trabelsi M, et al. Interval-valued features based machine learning technique for fault detection and diagnosis of uncertain HVAC systems. IEEE Access. 2020;8:171892–902.
28. Gunay HB, Shi Z. Cluster analysis-based anomaly detection in building automation systems. Energy Build. 2020;228: 110445.
29. Han H, Gu B, Hong Y, et al. Automated FDD of multiple-simultaneous faults (MSF) and the application to building chillers. Energy Build. 2011;43(9):2524–32.
30. Han H, Zhang Z, Cui X, et al. Ensemble learning with member optimization for fault diagnosis of a building energy system. Energy Build. 2020;226: 110351.
31. Himeur Y, Ghanem K, Alsalemi A, et al. Artificial intelligence based anomaly detection of energy consumption in buildings: a review, current trends and new perspectives. Appl Energy. 2021;287: 116601.
32. Hosseini Gourabpasi A, Nik-Bakht M. Knowledge discovery by analyzing the state of the art of data-driven fault detection and diagnostics of building HVAC. Civil Eng. 2021;2(4):986–1008.

33. Jalori S, Reddy TA. A unified inverse modeling framework for whole-building energy interval data: daily and hourly baseline modeling and short-term load forecasting. ASHRAE Trans. 2015;121(2):156–69.
34. James G, Witten D, Hastie T, et al. An introduction to statistical learning, vol. 112. Springer; 2013.
35. Jin B, Li D, Srinivasan S, et al Detecting and diagnosing incipient building faults using uncertainty information from deep neural networks. In: 2019 IEEE international conference on prognostics and health management (ICPHM), IEEE; 2019. p. 1–8.
36. Kalhori SRN, Zeng XJ. Improvement the accuracy of six applied classification algorithms through integrated supervised and unsupervised learning approach. J Comput Chem. 2014;02:201–9.
37. Kezunovic M, Rikalo I. Detect and classify faults using neural nets. IEEE Comput App Power. 1996;9:42–7.
38. Kim W, Braun JE. Development, implementation, and evaluation of a fault detection and diagnostics system based on integrated virtual sensors and fault impact models. Energy Build. 2020;228: 110368.
39. Kim W, Katipamula S. A review of fault detection and diagnostics methods for building systems. Sci Technol Built Environ. 2018;24(1):3–21.
40. Li D, Zhou Y, Hu G, et al. Handling incomplete sensor measurements in fault detection and diagnosis for building HVAC systems. IEEE Trans Autom Sci Eng. 2019;17(2):833–46.
41. Li G, Hu Y, Liu J, et al. Review on fault detection and diagnosis feature engineering in building heating, ventilation, air conditioning and refrigeration systems. IEEE Access. 2020;9:2153–87.
42. Li G, Yao Q, Fan C, et al. An explainable one-dimensional convolutional neural networks based fault diagnosis method for building heating, ventilation and air conditioning systems. Build Environ. 2021;203: 108057.
43. Li G, Zheng Y, Liu J, et al. An improved stacking ensemble learning-based sensor fault detection method for building energy systems using fault-discrimination information. J Build Eng. 2021;43: 102812.
44. Liao H, Cai W, Cheng F, et al. An online data-driven fault diagnosis method for air handling units by rule and convolutional neural networks. Sensors. 2021;21(13):4358.
45. Mariano-Hernández D, Hernández-Callejo L, Zorita-Lamadrid A, et al. A review of strategies for building energy management system: model predictive control, demand side management, optimization, and fault detect & diagnosis. J Build Eng. 2021;33: 101692.
46. Markus AA, Hobson BW, Gunay HB, et al. A framework for a multi-source, data-driven building energy management toolkit. Energy Build. 2021;250: 111255.
47. Martinez-Viol V, Urbano EM, Kampouropoulos K, et al Support vector machine based novelty detection and fdd framework applied to building ahu systems. In: 2020 25th IEEE international conference on emerging technologies and factory automation (ETFA), IEEE, 2020. p. 1749–54
48. Matetić I, Štajduhar I, Wolf I, et al. A review of data-driven approaches and techniques for fault detection and diagnosis in HVAC systems. Sensors. 2022;23(1):1.
49. Mirnaghi MS, Haghighat F. Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: a comprehensive review. Energy Build. 2020;229: 110492.
50. Miyata S, Akashi Y, Lim J, et al Model-based fault detection and diagnosis for HVAC systems using convolutional neural network. In: Building simulation 2019, IBPSA; 2019. p. 853–60.
51. Najafi M, Fault detection and diagnosis in building HVAC systems. PhD thesis, UC Berkeley; 2010.
52. Namburu SM, Azam MS, Luo J, et al. Data-driven modeling, fault diagnosis and optimal sensor selection for HVAC chillers. IEEE Trans Autom Sci Eng. 2007;4(3):469–73.
53. Novikova E, Bestuzhev M, Exploration of the anomalies in hvac data using image similarity assessment. In: 2020 9th Mediterranean conference on embedded computing (MECO), IEEE; 2020. p. 1–4.
54. Parzinger M, Hanfstaengl L, Sigg F, et al. Residual analysis of predictive modelling data for automated fault detection in building's heating, ventilation and air conditioning systems. Sustainability. 2020;12(17):6758.
55. Piscitelli MS, Mazzarelli DM, Capozzoli A. Enhancing operational performance of ahus through an advanced fault detection and diagnosis process based on temporal association and decision rules. Energy Build. 2020;226: 110369.
56. Piscitelli MS, Brandi S, Capozzoli A, et al. A data analytics-based tool for the detection and diagnosis of anomalous daily energy patterns in buildings. In: Building simulation. Springer; 2021. p. 131–47.
57. Ranade A, Provan G, Mady AED, et al. A computationally efficient method for fault diagnosis of fan-coil unit terminals in building heating ventilation and air conditioning systems. J Build Eng. 2020;27: 100955.
58. Rosato A, Guarino F, Sibilio S, et al. Healthy and faulty experimental performance of a typical HVAC system under Italian climatic conditions: artificial neural network-based model and fault impact assessment. Energies. 2021;14(17):5362.
59. Shalev-Shwartz S, Ben-David S. Understanding machine learning: from theory to algorithms. Cambridge University Press; 2014.
60. Simmini F, Rampazzo M, Peterle F, et al. A self-tuning KPCA-based approach to fault detection in chiller systems. IEEE Trans Control Syst Technol. 2021;30(4):1359–74.
61. Taal A, Itard L. Fault detection and diagnosis for indoor air quality in dcv systems: application of 4s3f method and effects of DBN probabilities. Build Environ. 2020;174: 106632.
62. Taal A, Itard L, Zeiler W, A diagnostic bayesian network method to diagnose building energy performance. In: Building Simulation 2019, IBPSA; 2019. p. 893–899
63. Taheri S, Ahmadi A, Mohammadi-Ivatloo B, et al. Fault detection diagnostic for HVAC systems via deep learning algorithms. Energy Build. 2021;250: 111275.
64. Tran DAT, Chen Y, Ao HL, et al. An enhanced chiller FDD strategy based on the combination of the LSSVR-de model and EWMA control charts. Int J Refrig. 2016;72:81–96.
65. Tun W, Wong JKW, Ling SH. Hybrid random forest and support vector machine modeling for HVAC fault detection and diagnosis. Sensors. 2021;21(24):8163.
66. Xu Y, Yan C, Shi J, et al. An anomaly detection and dynamic energy performance evaluation method for HVAC systems based on data mining. Sustain Energy Technol Assess. 2021;44: 101092.
67. Xue P, Zhou Z, Fang X, et al. Fault detection and operation optimization in district heating substations based on data mining techniques. Appl Energy. 2017;205:926–40.

68. Yan K, Zhong C, Ji Z, et al. Semi-supervised learning for early detection and diagnosis of various air handling unit faults. Energy Build. 2018;181:75–83.
69. Yan K, Chong A, Mo Y. Generative adversarial network for fault detection diagnosis of chillers. Build Environ. 2020;172: 106698.
70. Yan K, Huang J, Shen W, et al. Unsupervised learning for fault detection and diagnosis of air handling units. Energy Build. 2020;210: 109689.
71. Yang C, Gunay B, Shi Z, et al. Machine learning-based prognostics for central heating and cooling plant equipment health monitoring. IEEE Trans Autom Sci Eng. 2020;18(1):346–55.
72. Yang H, Zhang T, Li H, et al. HVAC equipment, unitary: fault detection and diagnosis. In: Encyclopedia of energy engineering and technology. 2nd ed. Boca Raton: CRC Press; 2014. p. 854–64.
73. Yao W, Li D, Gao L. Fault detection and diagnosis using tree-based ensemble learning methods and multivariate control charts for centrifugal chillers. J Build Eng. 2022;51: 104243.
74. You Y, Tang J, Guo M, et al. Ensemble learning based multi-fault diagnosis of air conditioning system. Energy Build. 2024;319: 114548.
75. Yu ZJ, Haghighat F, Fung BC, et al. A novel methodology for knowledge discovery through mining associations between building operational data. Energy Build. 2012;47:430–40.
76. Yun WS, Hong WH, Seo H. A data-driven fault detection and diagnosis scheme for air handling units in building HVAC systems considering undefined states. J Build Eng. 2021;35: 102111.
77. Zhang C, Xue X, Zhao Y, et al. An improved association rule mining-based method for revealing operational problems of building heating, ventilation and air conditioning (hvac) systems. Appl Energy. 2019;253: 113492.
78. Zhang C, Zhao Y, Lu J, et al. Analytic hierarchy process-based fuzzy post mining method for operation anomaly detection of building energy systems. Energy Build. 2021;252: 111426.
79. Zhang J, Li Z, Chen H, et al. Integrated generative networks embedded with ensemble classifiers for fault detection and diagnosis under small and imbalanced data of building air condition system. Energy Build. 2022;268: 112207.
80. Zhang S, Zhu X, Anduv B, et al. Fault detection and diagnosis for the screw chillers using multi-region XGboost model. Sci Technol Built Environ. 2021;27(5):608–23.
81. Zhao Y, Li T, Zhang X, et al. Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future. Renew Sustain Energy Rev. 2019;109:85–101.
82. Zhao Y, Zhang C, Zhang Y, et al. A review of data mining technologies in building energy systems: load prediction, pattern identification, fault detection and diagnosis. Energy Built Environ. 2020;1(2):149–64.