PROCEEDINGS OF ECOS 2026 - THE 39TH INTERNATIONAL CONFERENCE ON
EFFICIENCY, COST, OPTIMIZATION, SIMULATION, AND ENVIRONMENTAL IMPACT OF ENERGY SYSTEMS
28 JUNE – 03 JULY, 2026, CONSTANTA, ROMANIA

A Hybrid Machine Learning Approach for Performance Analysis and Anomaly Detection in CHP Systems for Non-Residential Buildings

Mahmoud M. Elshanshoury[a, c]*, Guido Francesco Frate[a]*, Nicola Bosi[b], Silvia D'Agostino[b], Lorenzo Ferrari[a], and Umberto Desideri[a]

a Department of Energy, Systems, Territory and Constructions Engineering (DESTEC), University of Pisa, Italy.
b TECNO SERVICE SRL, Facility and Energy Management, Lucca, Italy.
c Department of Mechanical Power Engineering, University of Tanta, Egypt.
** Corresponding authors (m.elshanshoury@studenti.unipi.it; guido.frate@unipi.it)*

Abstract:

This paper presents a comprehensive machine-learning framework for analyzing and diagnosing the operational behavior of energy systems in non-residential buildings, demonstrated through a detailed case study of an Italian hospital. Using a 17-month dataset recorded at 15-minute intervals—including electrical and thermal energy production, fuel consumption, multiple efficiency indicators, and ambient temperature—the methodology integrates systematic data preprocessing, multi-stage clustering, and supervised classification. Missing-value removal, filtering, and Density-Based Spatial Clustering (DBSCAN) are first applied to segment the data and isolate outliers; hierarchical clustering is then performed on the largest cluster, yielding six interpretable operational groups that reflect distinct CHP performance regimes. These groups serve as reference states for a multi-model classification comparison, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). Among all models, XGBoost achieved the highest predictive accuracy and generalization capability and was deployed for real-time inference on unseen operational data. The comparison between the original and predicted clusters revealed stable behavior for most groups and meaningful shifts in specific regimes, as supported by feature-importance interpretations. The combined clustering–classification approach enables robust operational state identification and early visibility of inefficiencies, contributing to the development of data-driven, intelligent energy-management systems for hospital facilities.

Keywords:

CHP System; Hospital energy systems; Anomaly detection; Hierarchical clustering; Machine learning.

1. Introduction

Hospitals rank among the most energy-intensive building types due to the continuous operation of HVAC systems, medical equipment, lighting, and specialized clinical services, which makes efficient energy management a strategic priority for sustainability and resilience [1, 2]. Modern Building Management Systems (BMS) now provide high-frequency data streams from thermal and electrical subsystems, creating opportunities for intelligent, data-driven operation. Yet these datasets are often underutilized due to heterogeneity, noise, and limited integration of analytics into routine workflows [1, 2].

Recent literature highlights that machine-learning methods can improve monitoring, control, and anomaly detection in CHP and building-energy systems; surveys report that AI-assisted models enhance efficiency, fault detection, and supervisory decision-making compared with traditional rule-based approaches [1]. System-level reviews of anomaly detection similarly emphasize adaptive ML—including tree-based ensembles—for identifying abnormal conditions in real-time operation [2]. In related thermal-energy applications, gradient-boosting methods have demonstrated unmatched predictive performance across heterogeneous installations, supporting their robustness for real-world deployment [3].

Motivated by these developments, this study presents a data-driven framework for analyzing and classifying operational behavior in a hospital CHP system using 17 months of BMS data. The methodology integrates systematic data preprocessing, DBSCAN, and hierarchical clustering to identify operational patterns, and a comparative evaluation of several supervised classifiers; among these, XGBoost emerged as the most accurate and was subsequently deployed for real-time inference on unseen data—providing state predictions, anomaly visibility, and interpretable operational insights.

1.1. Literature Review

Data-rich Building Management Systems (BMS) have accelerated the transition from rule-based supervision to data-driven monitoring and control in building-energy and CHP contexts, with surveys reporting that ML methods—tree-based ensembles, deep networks, and hybrids—improve fault diagnosis, forecasting, and operational decision-making [4,5].

For forecasting and operational prediction, gradient boosting families (XGBoost/LightGBM) frequently achieve state-of-the-art accuracy and robustness on energy data across multiple scales (community, building, and system), often outperforming recurrent deep models when features are engineered and tuned effectively [6-8].

In anomaly detection and fault detection and diagnosis (FDD) for building energy systems, system-wide reviews report extensive use of supervised machine-learning methods, particularly tree-based ensemble models and support vector machines. Random Forest and gradient-boosting methods are widely adopted for their robustness to noise and ability to capture nonlinear operational behavior. At the same time, SVMs are commonly used for reliable classification in high-dimensional BMS data [9–12].

When labels are scarce, clustering remains essential: DBSCAN captures irregular regimes and isolates noise, while hierarchical clustering reveals nested structure and multi-resolution operation patterns in equipment and building datasets [13,14].

Despite these advances, relatively few studies integrate multi-stage clustering, supervised classification, and deployable inference on unseen hospital BMS data, and fewer still quantify distributional shifts between original and predicted operational states—gaps this work addresses [4-9,13-14].

The objective of this research is to develop and validate a data-driven machine learning framework to analyze, classify, and diagnose the operational behavior of a hospital's combined heat and power (CHP) system using high-resolution BMS data. The study aims to:

1. Identify distinct operational patterns through multi-stage clustering.
2. Build and compare supervised classification models for reliable state recognition.
3. Deploy the best-performing model for inference on unseen data to assess generalization.
4. Provide interpretable insights into performance shifts and potential inefficiencies to support intelligent, sustainable energy management in healthcare facilities.

2. Methodology

This section describes the methodology for evaluating CHP system performance and detecting anomalies. The proposed framework includes several stages: data filtering and preprocessing; a multi-stage clustering process; and the application of multi-class classification models. This structured approach enables comprehensive evaluation of system behavior and supports identification of abnormal operating conditions.

2.1. Case study and data collection

This study examines only the CHP subsystem of the Pontedera Hospital BMS, using a small subset of the larger building-wide dataset. CHP data were collected over 17 months (19 April 2024–15 September 2025) at 15-minute intervals. The dataset includes electric and thermal energy production, gas consumption, and ambient temperature, while thermal, electric, and total efficiencies are engineered variables derived from these primary measurements.

2.2. Data preparation

The data preparation section includes processes such as data filtering, missing-value extraction, and data normalization. The following pipeline was implemented:

- **Missing Data Removal**: A portion of the dataset consisted of missing values. These were removed to ensure data integrity.
- **Data filtering**: Data points outside the normal operational ranges were removed in accordance with system specifications, such as efficiency values below 0.0 or above 1.0.

- **Data Normalization:** Before applying machine learning algorithms, the data were standardized using the Z-score transformation. This transformation ensures that variables with different scales contribute equally to the analysis, thereby improving the performance of clustering algorithms.

2.3. Clustering techniques

2.3.1 DBSCAN Clustering

After preparing the dataset, the DBSCAN algorithm was used as the first clustering technique to identify operational patterns and isolate noisy or abnormal data, as shown in Figure 1. DBSCAN was selected because it can detect clusters of arbitrary shapes and automatically identify outliers, making it particularly suitable for CHP operational data. The algorithm relies on two parameters: the radius ε, which defines the neighborhood around each point, and MinPts, the minimum number of points required to form a dense region. A grid search was conducted to optimize (ε, MinPts), and each configuration was evaluated using the Silhouette score. The optimal values—ε = 0.9 and MinPts = 105—yielded a Silhouette score of 0.6912 and produced four clusters, with one dominant cluster representing typical operating conditions. To support visual interpretation, Principal Component Analysis (PCA) was applied to reduce dimensionality before plotting. Finally, boxplots of all features were generated to compare the distributions across clusters and highlight differences in operating behavior.

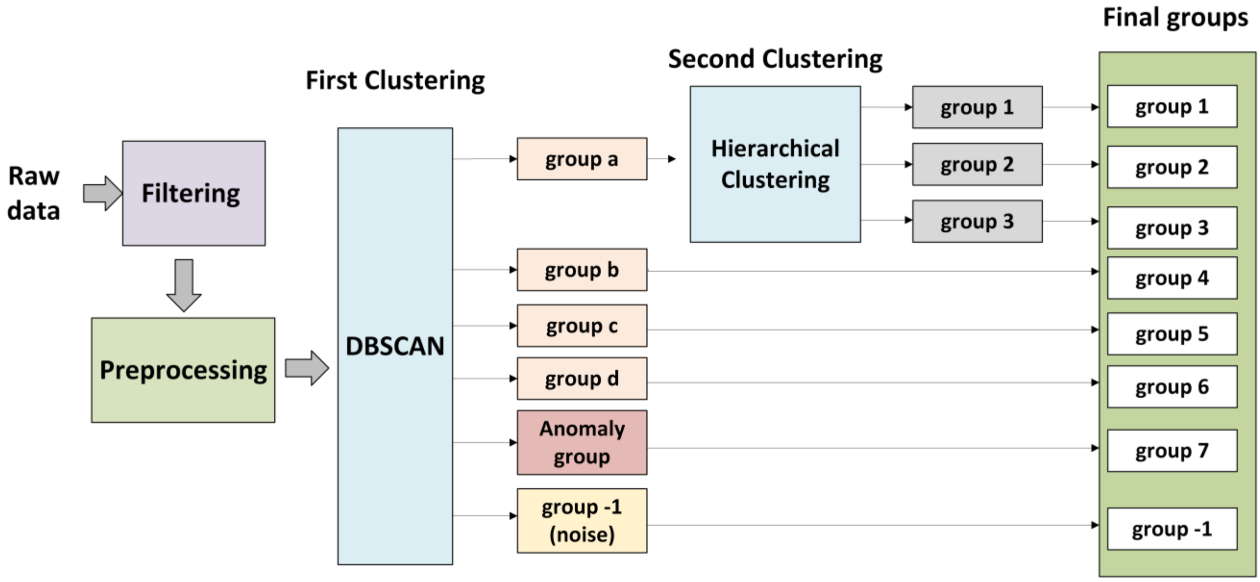

Figure 1. *Framework for clustering.*

2.3.2 Hierarchical Clustering

Following the DBSCAN analysis, hierarchical (agglomerative) clustering was applied to the largest cluster (23,110 data points) to resolve operational sub-patterns further. Multiple linkage methods and distance metrics, as implemented in standard Python clustering libraries (e.g., scikit-learn), were tested, and the configuration with the highest Silhouette score was selected. The best performance was obtained using Ward linkage with Euclidean distance, which achieved a Silhouette score of 0.6307 at the number of classes (k) = 3 across the tested range (k = 3–12).

Cluster characteristics were then examined using boxplots for all features, enabling comparison of operational differences between groups. Ultimately, the hierarchical analysis produced six distinct operational groups. Among these, Group 2 accounted for the majority of observations (70.73%) and was characterized by a high electric power output (≈530 kW) and a moderate ambient temperature (≈16.3 °C).

2.4. Model Classification

A multi-model framework was developed to classify operational groups based on energy parameters, as shown in Figure 2. The methodology began with data preprocessing, in which the timestamp column was removed, and the features were separated from the target variable. The dataset was subsequently divided into training and testing subsets using an 80/20 stratified split. Feature normalization was performed using the StandardScaler, fitted exclusively on the training data to avoid information leakage.

Four classification algorithms were considered: Random Forest, Gradient Boosting (XGBoost), Support Vector Machine (SVM), and a Multilayer Perceptron (MLP) neural network. Model performance was optimized through hyperparameter tuning using a grid search within a five-fold stratified cross-validation framework. This approach ensured proportional representation of each class across folds, thereby reducing sampling bias and achieving a balanced bias–variance trade-off.

2.4.1. Evaluation Metrics

We evaluated the model using overall accuracy and other common performance metrics to gain a clearer picture of how well it classified each group. In addition to accuracy, we examined precision, recall, and F1-score, and we also examined confusion matrices to see where the model made correct or incorrect predictions.

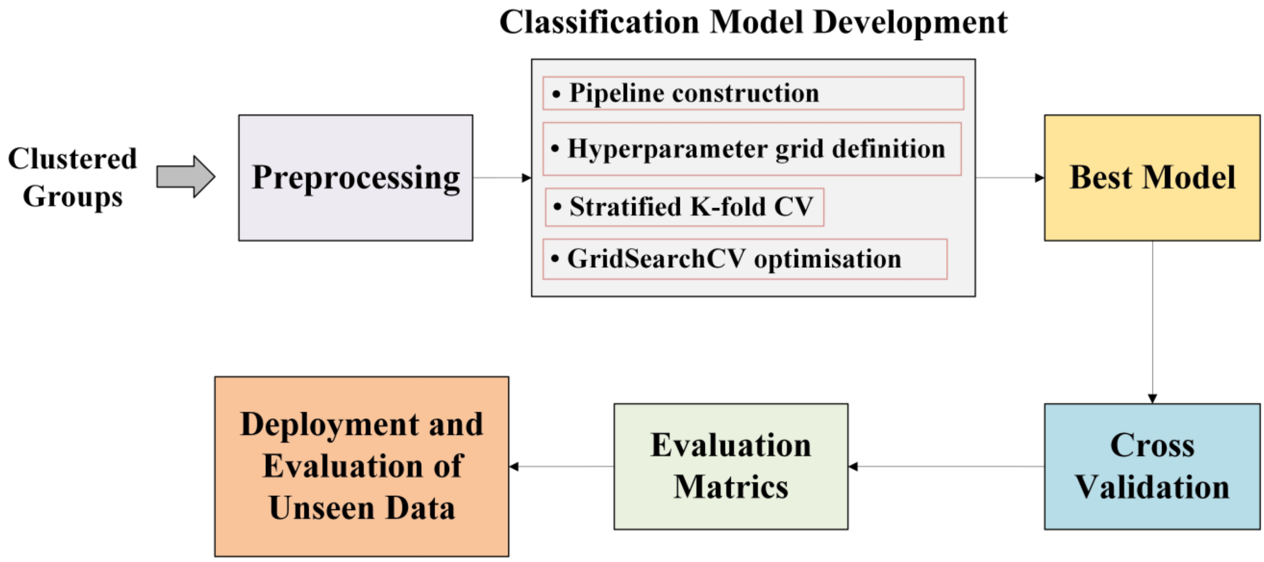

Figure 2. *Framework for model classification.*

2.4.2. Deployment and Evaluation of Unseen Data

After model development, the best-performing classifier (XGBoost) was deployed to evaluate its performance on a new, unseen dataset. The inference pipeline consisted of loading the trained scikit-learn pipeline—which incorporates both preprocessing (scaling and feature alignment) and the XGBoost classifier—along with the corresponding Label-Encoder used during training.

Before prediction, the unseen dataset was validated to ensure consistency with the model's training inputs. Feature columns were reordered to match the exact feature set and column order used during training, and non-feature fields (e.g., timestamps and metadata) were removed. Records containing missing values were excluded to remain consistent with the training configuration. The inference pipeline then applied the same preprocessing and transformation steps used during training, ensuring deterministic and reproducible predictions.

3. Results and Discussion

This section presents the findings from applying the machine learning methodology to the CHP unit in the energy dataset from Pontedera Hospital.

3.1. Clustering techniques

The DBSCAN algorithm initially identified an outlier group characterized by zero electric energy despite measurable thermal energy and fuel consumption. This group was classified as an anomaly, likely reflecting a malfunction in the electric meter sensor. Excluding these anomalies, DBSCAN was applied and produced four distinct clusters with a Silhouette score of 0.69, indicating strong separation and well-defined group structure, as shown in Figure 3. This step isolated noise from the dataset, ensuring subsequent analyses focused on valid operational patterns. Cluster 1, the largest group, has the most frequent operating conditions and was selected for detailed examination. Figure 4 presents the boxplots of the main operational variables across the DBSCAN-derived clusters. Each boxplot illustrates the distribution, variability, and presence of outliers for each feature within each cluster, enabling a direct visual comparison of operating behavior.

By examining the boxplots, the four groups identified can be clearly distinguished. For example, Cluster 1 shows the maximum electric energy output with higher thermal energy output; Cluster 2 reflects intermediate conditions with moderate efficiency values; Cluster 3 corresponds to higher ambient temperatures and lower thermal output; while Cluster 4 exhibits the lowest electric energy output. These visual patterns help validate and interpret the statistical differences that define each operational group.

Visualizing the clusters in this way is particularly useful because it provides an intuitive understanding of how the CHP system behaves under different operating regimes. The boxplots highlight not only the central tendencies but also the variability and extreme values within each cluster, making it easier to identify characteristic signatures, unusual patterns, and potential inefficiencies that may warrant further investigation.

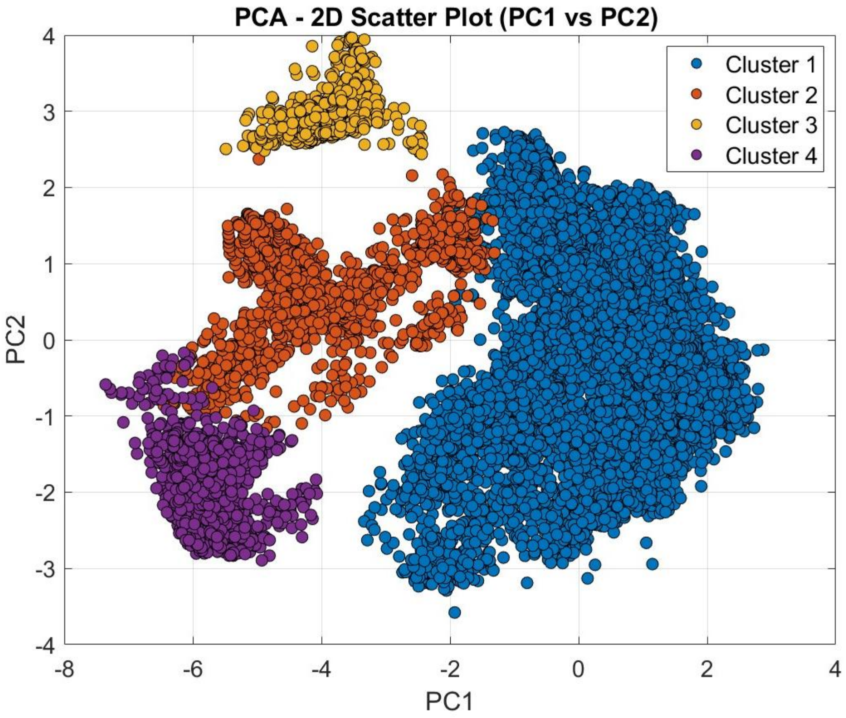

__Figure 3.__ Scatter plot for the initial clusters.

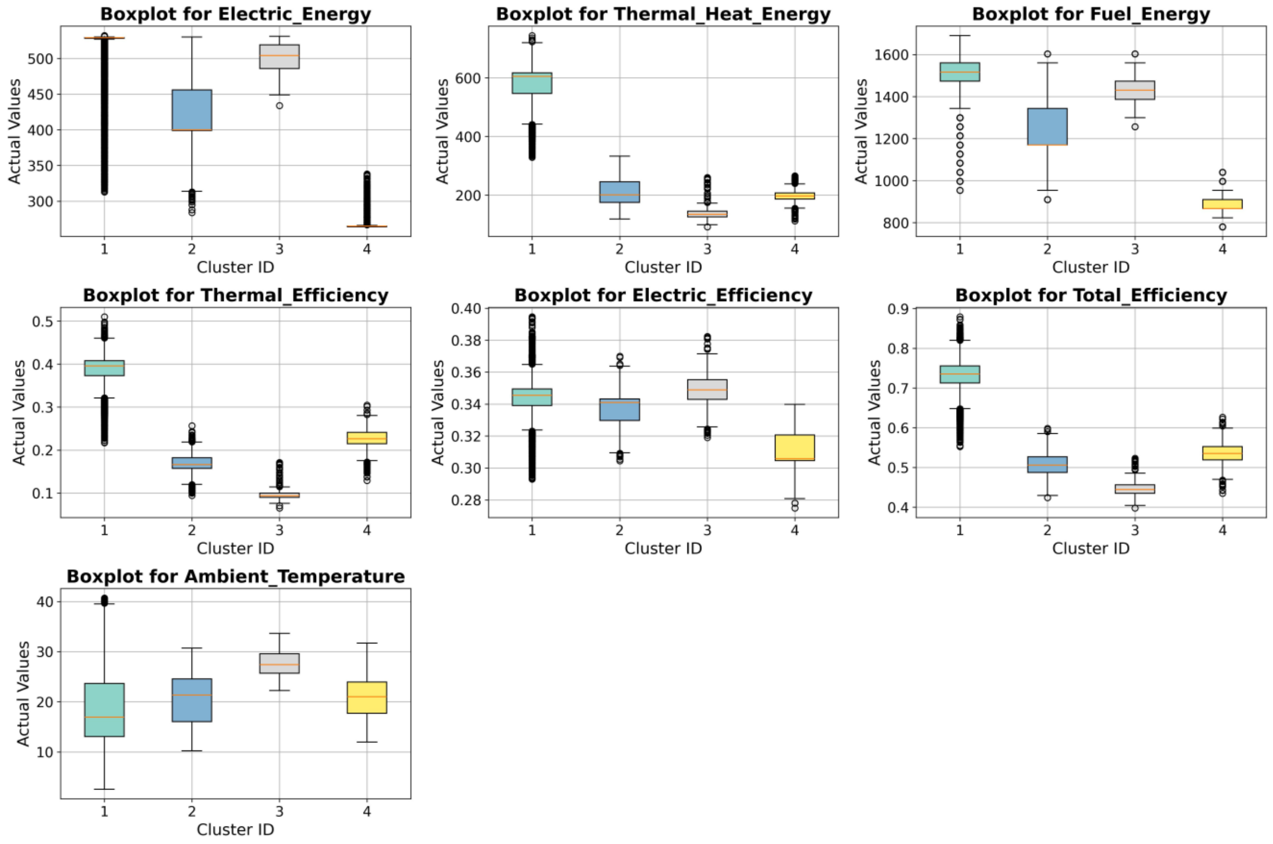

__Figure 4.__ Boxplot for each feature across the initial clustered groups.

Analysis of Cluster 1, comprising 23,110 data points, using hierarchical (agglomerative) clustering revealed more granular operational patterns. Ward linkage combined with Euclidean distance produced the most coherent grouping, achieving a Silhouette score of 0.63. Based on cluster cohesion, an optimal solution was selected, which, when integrated with the broader dataset, resulted in six distinct operational groups.

The scatter plot for the final clustered groups from the hybrid clustering techniques based on the principal components (PC1 and PC2) is shown in Figure 5. Table 1 summarizes the distinct energy-consumption profiles of these six groups. By identifying these profiles, particular operational states—such as periods of reduced efficiency and opportunities for load shifting—can be recognized, providing practical insights into system performance.

Table 1. The feature for the final clustered groups.

Group	Feature	Percentage, %
1	Maximum electric power, high temperature, and average thermal load	8.71
2	Maximum electric power, low temperature, high thermal efficiency, high thermal load, high electric efficiency, and maximum total efficiency	70.73
3	High electric power, high Temperature, and very low thermal efficiency	2.40
4	Average electric power, low temperature, high thermal efficiency, and high thermal load	6.77
5	Average electric power, average temperature, low thermal efficiency, and high electric efficiency	6.35
6	Low electric power, average temperature, and average thermal efficiency	5.04

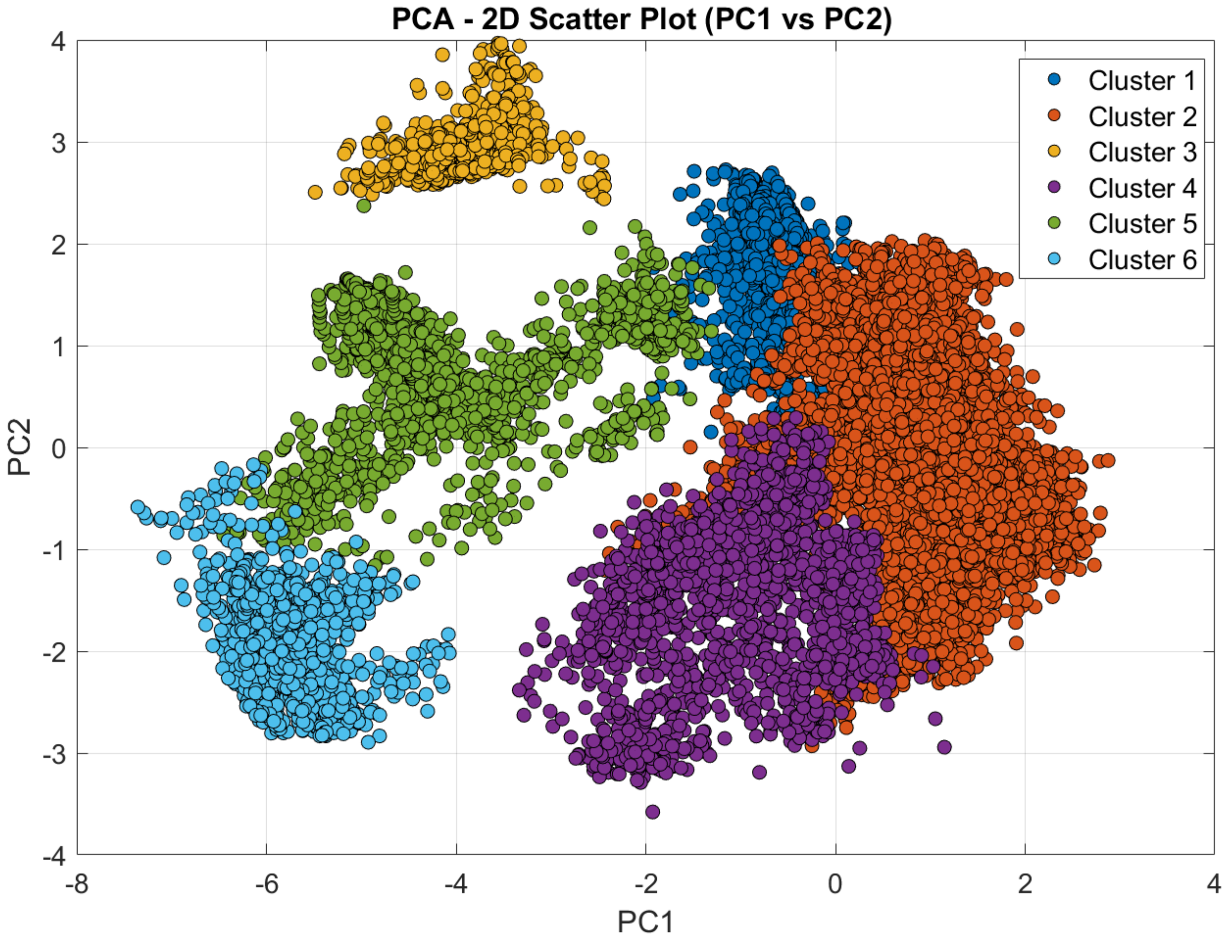

Figure 5. *Scatter plot for the final clusters.*

Figure 6 presents boxplots of the key features across the final clustered groups. Clear differences are observed among clusters, particularly in electricity consumption, thermal energy use, and ambient temperature. These variations highlight distinct operational profiles, reinforcing the validity of the clustering approach and providing insight into how energy demand and environmental conditions shape group behavior.

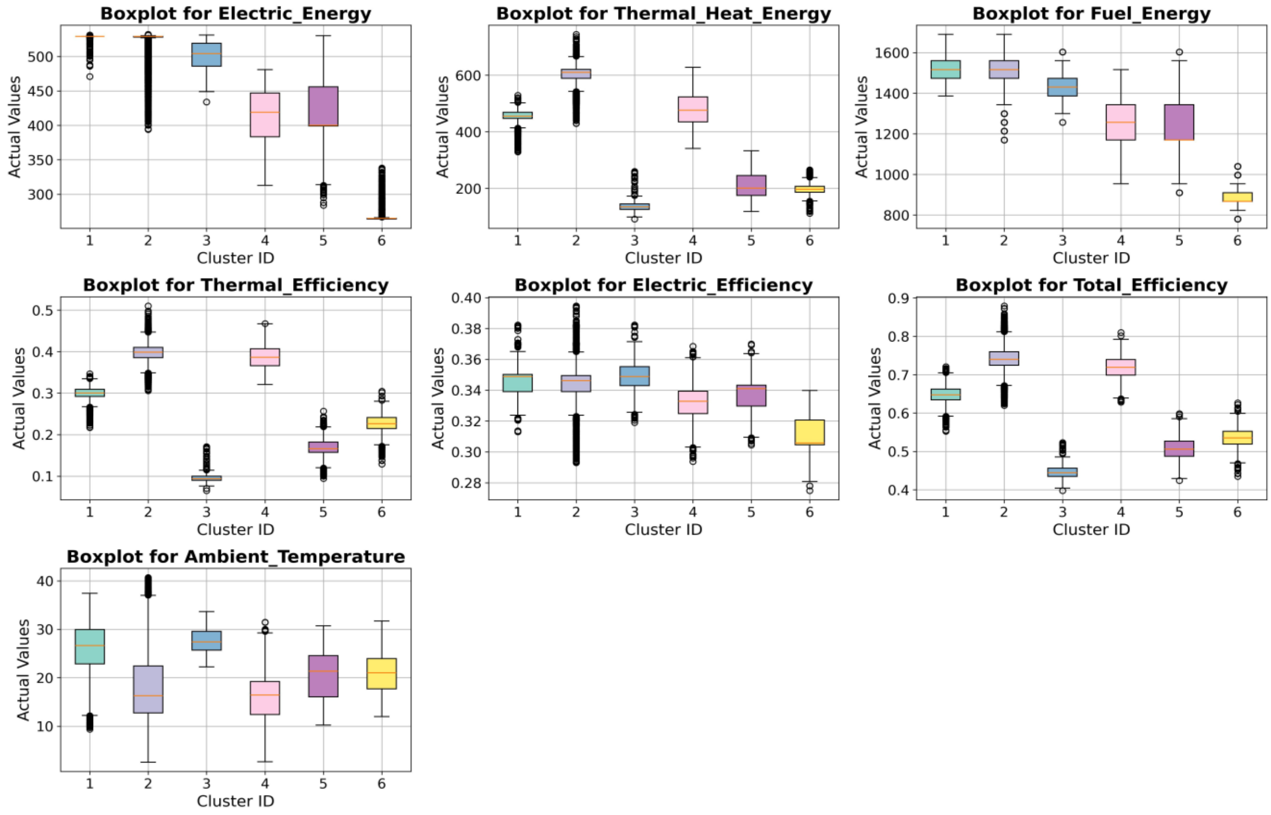

Figure 6. *Boxplot for each feature across the final clustered groups.*

3.2. Model Classification

A multi-model framework was implemented to classify these operational states. Models were trained on a pre-processed dataset using an 80/20 stratified split and a StandardScaler for feature scaling. Four models, Random Forest, Gradient Boosting, SVM, and Multilayer Perceptron (MLP), were evaluated.

3.2.1 Overall Model Performance

Table 2 summarizes the primary performance metrics for the four evaluated models, including accuracy, F1-score, precision, recall, Cohen's kappa, and total training time. Across all metrics, XGBoost consistently delivers the strongest performance, achieving the highest accuracy (0.9938), F1-score (0.9863), and Cohen's kappa (0.99). These results reinforce its ability to model the underlying class structure effectively while maintaining excellent generalization. In contrast, the SVM model achieved the lowest overall performance (accuracy = 0.9838), with lower precision (0.957) despite relatively high recall (0.986), suggesting a tendency toward false positives. Overall, these results highlight XGBoost as the most accurate and reliable model, followed closely by Random Forest, whereas Neural Network and SVM achieve lower accuracy or higher computational cost.

Table 2. *The performance metrics for classification models.*

Model	Accuracy	F1_Score	Precision	Recall	Cohen Kappa	Training Time (s)
XGBoost	0.9938	0.9863	0.987	0.986	0.99	330.17
Random Forest	0.9916	0.9799	0.982	0.978	0.987	336.1
Neural Network	0.9907	0.9777	0.978	0.978	0.985	598
SVM	0.9838	0.9706	0.957	0.986	0.975	445

3.2.2 Confusion Matrix Analysis

Figure 7 shows the Confusion matrices for the classification models on the test set. Each matrix shows the distribution of true versus predicted labels across the eight classes. XGBoost and Random Forest exhibit the strongest diagonal dominance, indicating highly accurate predictions with minimal cross-class confusion. The Neural Network's performance is similar, although it exhibits slightly greater dispersion across the minority classes. Conversely, the SVM model exhibits the highest misclassification rate, particularly for classes with fewer samples. These results provide a detailed view of class-level performance and highlight the robustness of tree-based ensemble methods.

3.2.3 Learning Curve Analysis

To further assess the behavior of the evaluated models under varying data availability, learning curves were generated for SVM, Random Forest, XGBoost, and the Neural Network. These curves plot both training and cross-validation accuracy as the number of training examples increases, providing insight into model bias, variance, and sample efficiency, as shown in the Figure 8. The XGBoost model exhibits the most favorable learning dynamics. Both its training and cross-validation curves rise sharply with the first few thousand samples and converge early, stabilizing with only a minimal gap between them. This pattern indicates excellent generalization, low variance, and strong sample efficiency. Random Forest follows a similar pattern, with validation performance consistently high and closely tracking training accuracy, although it converges slightly later than XGBoost.

In contrast, the SVM exhibits the weakest learning-curve performance. Both training and validation curves remain below those of the other models across all sample sizes, with a consistently small gap between them. This behavior reflects underfitting, suggesting that the SVM model—despite benefiting from additional data—lacks sufficient capacity to capture the dataset's underlying complexity.

Overall, the learning-curve analysis supports the conclusions drawn from the cross-validation and test-set evaluations. XGBoost and Random Forest are the most sample-efficient and robust models, achieving strong generalization early and maintaining high accuracy throughout training. The Neural Network ultimately achieves competitive performance but requires substantially more data to reduce variance. Meanwhile, SVMs struggle to match the capacity and flexibility of ensemble methods. These findings further justify selecting XGBoost as the most reliable model for this classification task.

SVM

True \ Predicted	-1 (0)	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
-1 (0)	178	1	1	4	0	2	1	0
1 (1)	2	465	0	0	0	0	0	0
2 (2)	13	12	3712	0	56	0	0	0
3 (3)	0	0	0	128	0	0	0	0
4 (4)	2	0	0	0	361	0	0	0
5 (5)	6	0	0	1	0	331	3	0
6 (6)	1	0	0	0	0	1	269	0
7 (7)	0	0	0	0	0	0	0	1012

Random Forest

True \ Predicted	-1 (0)	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
-1 (0)	164	1	11	1	1	7	2	0
1 (1)	1	466	0	0	0	0	0	0
2 (2)	3	5	3779	0	6	0	0	0
3 (3)	0	0	0	128	0	0	0	0
4 (4)	0	0	8	0	355	0	0	0
5 (5)	4	0	0	1	0	333	3	0
6 (6)	0	0	0	0	0	1	270	0
7 (7)	0	0	0	0	0	0	0	1012

XGBoost

True \ Predicted	-1 (0)	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
-1 (0)	175	1	8	0	1	2	0	0
1 (1)	1	466	0	0	0	0	0	0
2 (2)	3	4	3781	0	5	0	0	0
3 (3)	0	0	0	128	0	0	0	0
4 (4)	0	0	7	0	356	0	0	0
5 (5)	4	0	0	1	0	333	3	0
6 (6)	0	0	0	0	0	1	270	0
7 (7)	0	0	0	0	0	0	0	1012

Neural Network

True \ Predicted	-1 (0)	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
-1 (0)	170	1	3	4	1	7	1	0
1 (1)	3	461	3	0	0	0	0	0
2 (2)	7	5	3778	0	3	0	0	0
3 (3)	0	0	0	128	0	0	0	0
4 (4)	0	0	14	0	349	0	0	0
5 (5)	3	0	0	1	0	337	0	0
6 (6)	1	0	0	0	0	4	266	0
7 (7)	0	0	0	0	0	0	0	1012

Figure 7. *Confusion matrices for classification models on the test set.*

Figure 9 summarizes the feature importance ranking produced by the XGBoost classifier. Total Efficiency and Electric Energy clearly dominate the model's decision process, indicating that the classification of operating states is primarily driven by the combined energy–efficiency behavior of the CHP unit. Intermediate contributions from Fuel Energy, Thermal Efficiency, and Thermal Heat Energy reflect additional refinement of class boundaries, while Electric Efficiency and Ambient Temperature play a smaller role. Overall, the ranking aligns with the system's physical characteristics and confirms that XGBoost relies on meaningful operational signals rather than on noise or spurious correlations.

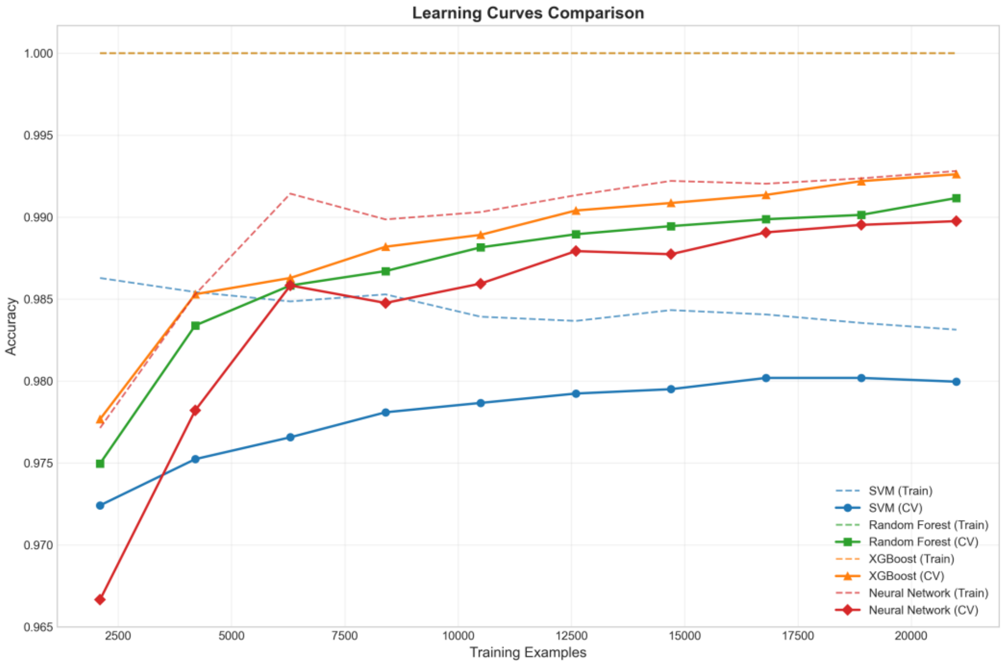

Figure 8. *Learning curves for classification models.*

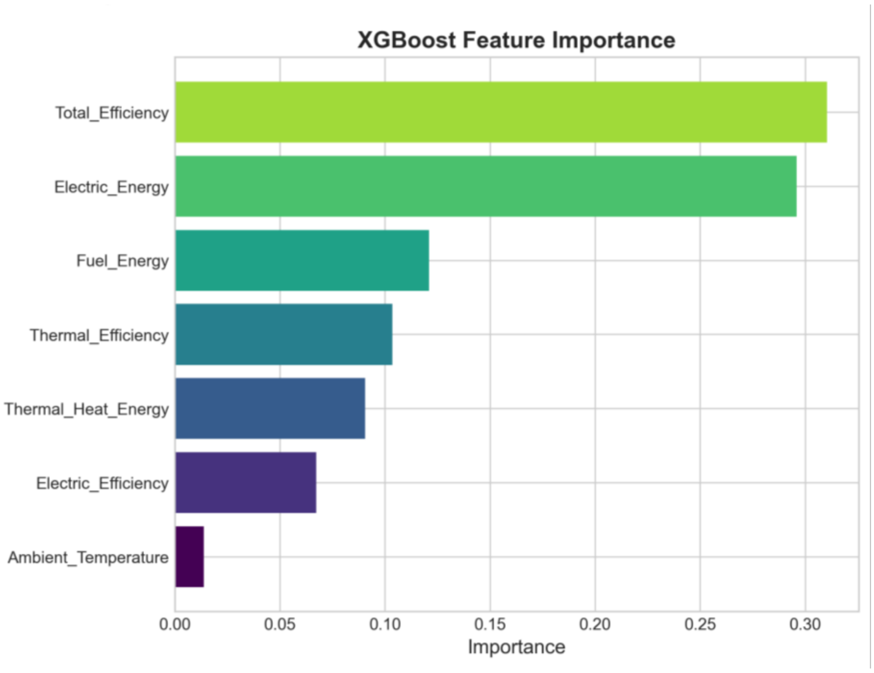

***Figure 9**. Feature importance ranking from the XGBoost classifier.*

3.2.4 Model deployment on unseen data

The trained XGBoost pipeline was applied to a new, unseen dataset using a reproducible inference script that enforces feature-schema validation, identical preprocessing, and label decoding. The pipeline outputs predicted classes and class probabilities.

Figure 10. presents the class distribution predicted by XGBoost on the unseen dataset. The model assigns most samples to Classes 2 and 4, while Classes -1, 1, 5, 6, and 7 appear far less frequently. This indicates that the unseen data is similarly imbalanced to the training set and that XGBoost correctly identifies the dominant operational states while still recognizing the less frequent ones.

To evaluate how the XGBoost classifier redistributed samples across operational clusters, we compared the original (ground-truth) cluster assignments with the new labels predicted on the unseen dataset. For each feature and each cluster, paired boxplots were generated to visualize distributional changes, and descriptive statistics (mean, standard deviation, and sample size) were computed.

Overall, XGBoost preserved the structure of the major clusters, especially clusters 1, 2, and 5, with changes in mean values remaining within ±3% for most features. These clusters, therefore, show high stability and robustness between the original and predicted labeling.

More substantial changes were observed in clusters 4 and 6, where XGBoost reassigned a portion of observations, resulting in noticeable differences in several energy-related variables. In particular:

- Cluster 4 shows consistent downward shifts in Electric Energy, Thermal Heat Energy, and Fuel Energy (−8% to −14%), along with moderate increases in Thermal Efficiency and Total Efficiency (+4% to +7%), indicating a movement of lower-energy, higher-efficiency points into this cluster.
- Despite these shifts, the predicted label distributions remain physically interpretable and consistent with XGBoost's optimized decision boundaries. The boxplot comparisons provide valuable diagnostic insight into how the classifier reorganizes the operational states in the unseen dataset. Examples of the generated comparison plots are shown in Figure 11.

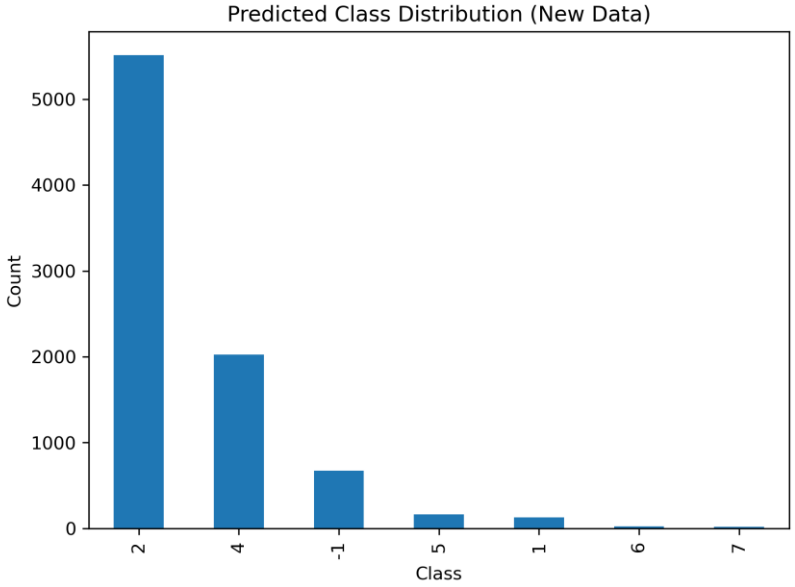

Figure 10. *The class distribution predicted by XGBoost on the unseen dataset.*

The results of this study demonstrate the effectiveness of the XGBoost model in understanding new operational data while maintaining the original group structure. Using a strict set of steps to prepare the data, fine-tuning the settings, and thoroughly testing the results enabled the classifier to accurately recreate the most common operating conditions and provide valuable insights into changes in system behavior.

4. Conclusion

This work introduced a data-driven framework for analyzing the operational behavior of a CHP system and identifying anomalies through a combination of multistage clustering and supervised classification. The initial clustering process produced six interpretable operational groups that captured the unit's main thermodynamic and electrical regimes, forming a solid foundation for developing a reliable classifier.

Among the tested models, XGBoost delivered the best performance and demonstrated strong generalization to unseen data. Its predictions preserved the dominant structure of the original clusters and correctly identified less frequent modes, confirming its suitability for real-world deployment. The comparison between original and predicted clusters showed that most groups, especially Groups 1, 2, and 5, remained highly stable. In contrast, meaningful shifts in Groups 4 and 6 revealed how the system adapts under varying operating conditions.

Overall, the results show that combining clustering-based knowledge discovery with a high-performance classifier provides a practical and interpretable solution for real-time CHP monitoring and anomaly detection.

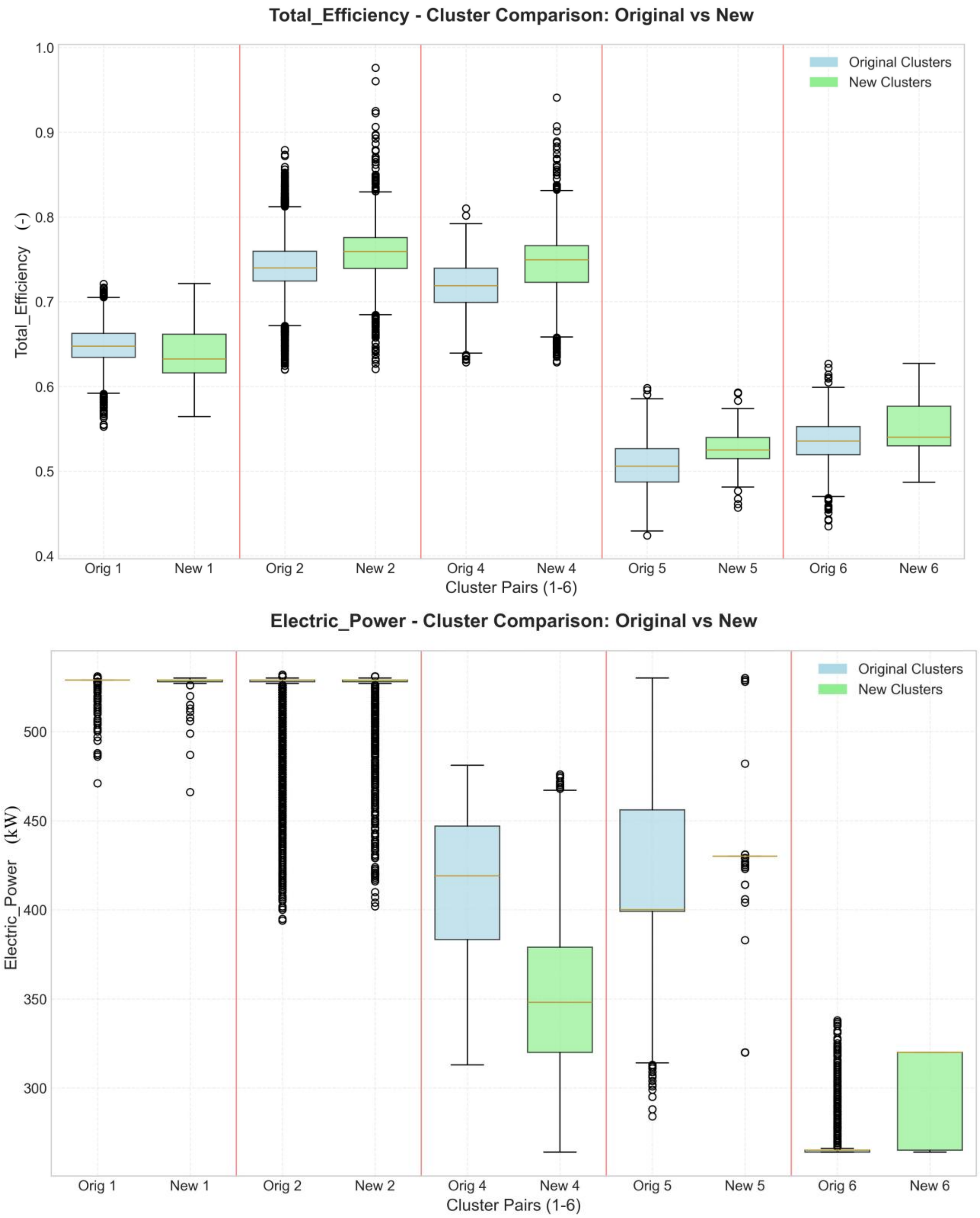

Figure 11. *Boxplot for cluster comparison: a) total efficiency, b) electric power.*

Nomenclature

Acronyms & Abbreviations

AI	Artificial Intelligence
BMS	Building Management System
CV	Cross-Validation
CHP	Combined Heat and Power
DBSCAN	Density-Based Spatial Clustering of Applications with Noise

FDD	Fault Detection and Diagnosis
HVAC	Heating, Ventilation, and Air Conditioning
HC	Hierarchical (Agglomerative) Clustering
K	Number of classes
MinPts	Minimum number of points
ML	Machine Learning
MLP	Multilayer Perceptron (neural network)
RF	Random Forest
SVM	Support Vector Machine
T	Ambient temperature [°C (degree Celsius)]
XGBoost	Extreme Gradient Boosting

Greek symbols

ε	the radius

References

[1] Safari, A., & Oshnoei, A. (2025). An Overview of Recent AI Applications in Combined Heat and Power Systems. Energies, 18(11), 2891. https://doi.org/10.3390/en18112891.

[2] Aghazadeh Ardebili, A., Hasidi, O., Bendaouia, A. et al. Enhancing resilience in complex energy systems through real-time anomaly detection: a systematic literature review. Energy Inform 7, 96 (2024). https://doi.org/10.1186/s42162-024-00401-8.

[3] Rahal, M., Ahmed, B.S., Renström, R. et al. Data-driven heat pump management: combining machine learning with anomaly detection for residential hot water systems. Neural Comput & Applic 37, 16203–16229 (2025). https://doi.org/10.1007/s00521-025-11318-y.

[4] Li D., Qi Z., Zhou Y., Elchalakani M., Machine learning applications in building energy systems: review and prospects. Buildings 2025;15(4):648. doi:10.3390/buildings15040648.

[5] Wang X., Wang H., Bhandari B., Cheng L., AI-empowered methods for smart energy consumption: a review of load forecasting, anomaly detection, and demand response. Int J Precis Eng Manuf-Green Technol 2024; 11:963-993. doi:10.1007/s40684-023-00537-0.

[6] Semmelmann L., Henni S., Weinhardt C., Load forecasting for energy communities: a novel LSTM-XGBoost hybrid model based on smart meter data. Energy Informatics 2022;5(24):1-21. doi:10.1186/s42162-022-00212-9.

[7] Song K-M., Kim T-G., Cho S-M., Song K-B., Yoon S-G., XGBoost-based very short-term load forecasting using day-ahead load forecasting results. Electronics 2025;14(18):3747. doi:10.3390/electronics14183747.

[8] Huang G., Wan H., Enhancing building energy forecasting with advanced ML techniques: an in-depth study of model performance, parameter optimization, and predictive accuracy. J Mech Sci Technol 2025; 39:6347-6362. doi:10.1007/s12206-025-0959-x.

[9] Jian Bi, Hua Wang, Enbo Yan, Chuan Wang, Ke Yan, Liangliang Jiang, Bin Yang, AI in HVAC fault detection and diagnosis: A systematic review, Energy Reviews, Volume 3, Issue 2,2024,100071, ISSN 2772- 9702.

[10] Matetić, I., Štajduhar, I., Wolf, I., & Ljubic, S. (2023). A Review of Data-Driven Approaches and Techniques for Fault Detection and Diagnosis in HVAC Systems. *Sensors*, *23*(1), 1. https://doi.org/10.3390/s23010001

[11] Ke Yan, Chaowen Zhong, Zhiwei Ji, Jing Huang, Semi-supervised learning for early detection and diagnosis of various air handling unit faults, Energy and Buildings, Volume 181,2018, Pages 75-83, ISSN 0378-7788.

[12] Hao Yang, Jianfeng Xu, Hao Wu, Yuzhi Liu, Liangzhou Chen, Fault diagnosis of air conditioning units using XGBoost ensemble and evidence distance fusion, Journal of Building Engineering, Volume 124,2026,116085, ISSN 2352-7102.

[13] Economopoulou E.-E., Vonitsanos G., Mylonas P., Kanavos A., Efficient energy disaggregation using DBSCAN: a novel approach for enhanced energy management. In: AIAI 2024 IFIP WG 12.5 Workshops. IFIP AICT, vol. 715. Springer; 2024. p.236-249. doi:10.1007/978-3-031-63227-3_16.

[14] Habib U., Hayat K., Zucker G., Complex building's energy system operation patterns analysis using bag-of-words representation with hierarchical clustering. Complex Adaptive Systems Modeling 2016; 4:8. doi:10.1186/s40294-016-0020-0.