OPEN ACCESS

ARTICLE

Journal homepage: www.the-innovation.org/energy-use

INNOVATION PRESS

Energy Use

Energy Use 2 (2026):100060

Contents lists available at INNOVATION PRESS

# Privacy-Preserving Federated Learning for VRF System Energy Use Prediction with SMOGN Augmentation And GA-Based Aggregation

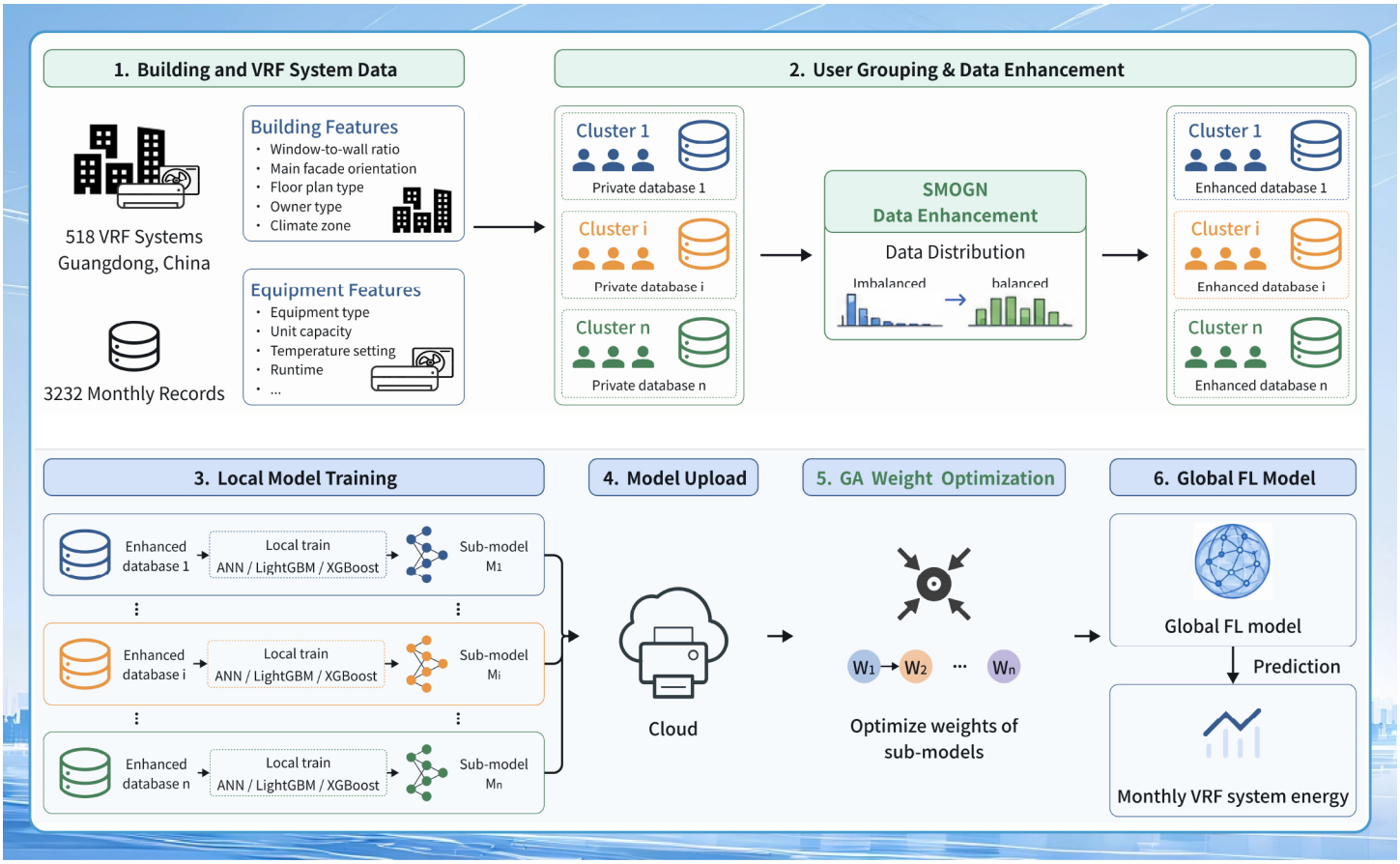
**Zhipeng Guo,[1,6] Bihai Wang,[2,6] Yu Jiang,[3] Leqi Zhu,[3] Yixing Chen,[4,5] and Guanjing Lin[2,*]**

*Correspondence: linguanjing@sz.tsinghua.edu.cn

Received: June 14, 2026; Accepted: June 23, 2026; Published Online: June 25, 2026; https://doi.org/10.59717/ipj.energy-use.2026.100060



## GRAPHICAL ABSTRACT




## PUBLIC SUMMARY

- Precise VRF energy prediction guides equipment matching and operation; traditional centralized forecasting risks user data privacy, while federated learning trains global models with local raw data retained.
- An improved federated learning framework combines client-side SMOGN data augmentation and GA weighted aggregation to solve VRF data imbalance and poor model fusion defects.
- Dataset of 518 Chinese VRF systems (3232 monthly records) is expanded to 4500 samples via SMOGN; three ML models are tested with building and equipment characteristic inputs.
- The proposed privacy-preserving FL framework reaches $R^2$=0.8390, approaching centralized benchmark $R^2$=0.8913, verifying its accuracy and scalability for VRF energy forecasting.

ARTICLE

OPEN ACCESS

Journal homepage: www.the-innovation.org/energy-use

Energy Use

Energy Use 2 (2026):100060

Contents lists available at INNOVATION PRESS

# Privacy-Preserving Federated Learning for VRF System Energy Use Prediction with SMOGN Augmentation And GA-Based Aggregation

**Zhipeng Guo,[1,6] Bihai Wang,[2,6] Yu Jiang,[3] Leqi Zhu,[3] Yixing Chen,[4,5] and Guanjing Lin[2,*]**

[1]Department of mechanical and automation engineering, Chinese University of Hong Kong, Hong Kong SAR, China

[2]Institute of Future Human Habitats, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China

[3]Midea Building Technologies, Midea Group, Foshan 528311, China

[4]College of Civil Engineering, Hunan University, Changsha 410082, China

[5]Key Laboratory of Building Safety and Energy Efficiency of Ministry of Education, Hunan University, Changsha 410082, China

[6]These authors contributed equally

*Correspondence: linguanjing@sz.tsinghua.edu.cn

Received: June 14, 2026; Accepted: June 23, 2026; Published Online: June 25, 2026; https://doi.org/10.59717/ipj.energy-use.2026.100060



Citation: Guo Z., Wang B., Jiang Y., et al. (2026). Privacy-Preserving Federated Learning for VRF System Energy Use Prediction with SMOGN Augmentation And GA-Based Aggregation. *Energy Use* **2**:100060.

Accurate prediction of energy use for Variable Refrigerant Flow (VRF) systems is critical for optimizing equipment selection and efficient operation. Traditional prediction methods typically depend on centralized data collection, raising privacy concerns. Federated Learning (FL) provides a privacy-preserving alternative by enabling multiple clients to collaboratively train a global model while keeping raw data localized. However, practical FL applications for VRF systems face challenges of data imbalance and suboptimal model aggregation. To address these issues, this paper proposes an enhanced FL framework integrating Synthetic Minority Oversampling Technique for Regression with Gaussian Noise (SMOGN) and Genetic Algorithm (GA) to improve prediction accuracy while preserving privacy. SMOGN is employed at client level to augment local datasets and mitigate imbalance. A GA-based performance-weighted ensemble aggregation strategy is introduced to optimize model fusion in FL, replacing conventional uniform averaging and enabling better utilization of heterogeneous local models. The framework was evaluated using data from 518 VRF systems in China (3,232 monthly records, expanded to 4,500 samples via SMOGN). Three machine learning models, i.e. Artificial Neural Networks, Light Gradient Boosting Machine, and eXtreme Gradient Boosting, were implemented under centralized and federated settings, using five building features and nine equipment features as inputs to predict monthly energy consumption. Results show that the proposed FL framework achieves an $R^2$ value of 0.8390, close to centralized model performance ($R^2$ = 0.8913), while preserving data privacy. These findings demonstrate the effectiveness of the proposed approach as a scalable and privacy-preserving solution for energy consumption prediction in VRF systems.

## INTRODUCTION

The building sector is a major contributor to global energy consumption and carbon emissions.[1] According to data from the International Energy Agency, energy use associated with building operations and construction accounts for 34% of global final energy consumption, with building operations alone responsible for 30% and contributing 28% of energy-related $CO_2$ emissions.[2] In China, the government has explicitly stipulated that by 2025, all new urban buildings must meet green standards, and the area of energy-efficient retrofits for existing buildings must exceed 350 million square meters.[3] This policy context imposes higher requirements for building energy efficiency optimization.

Within building operational energy consumption, Heating, Ventilation, and Air Conditioning (HVAC) systems are the primary consumers, typically accounting for 40%-50% of a building's total energy use.[4,5] Among HVAC systems, Variable Refrigerant Flow (VRF) systems have become a mainstream choice for multi-split air conditioning systems due to their efficient load regulation capability, flexible control systems, and relatively low maintenance costs, with their market share continually rising, particularly in China and Japan.[6]

As the application scale of VRF systems expands, accurately predicting their energy consumption becomes crucial for optimizing new system selection, improving operational efficiency, and enabling energy-saving control. In recent years, data-driven methods have been widely applied in this domain.[7] For instance, Hsu et al. (2025) compared the performance of Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) networks for VRF system energy prediction, finding LSTM advantageous for handling temporal data.[8] In each season, it selected one month for training, one month for validation, and one month for testing. Oh et al. (2024) proposed an eXtreme Gradient Boosting (XGB)-based regression model that predicts the instantaneous power input of a VRF system under current operating conditions (e.g., indoor/outdoor temperature, humidity, and set-point values) using only manufacturer performance data and limited experimental measurements. The model achieved high accuracy ($R^2$ > 0.9, RMSE < 0.2 kW), significantly outperforming traditional catalog-based approaches.[9] Although these studies have achieved favorable results in terms of modeling accuracy, they still face two key challenges in practical applications: insufficient data volume and heterogeneity and data privacy issues.

Training high-precision prediction models usually requires large-scale, high-quality data. However, in practice, the data available from a single building is often limited. Furthermore, significant differences in thermal physical properties and operational patterns among different building groups lead to highly heterogeneous data distributions.[10] Traditional centralized modeling methods struggle to adapt to this diversity, limiting the model's generalization capability.

Data privacy is another challenge. Existing data-driven models typically rely on centralized data collection and sharing, requiring building operators to upload sensitive information—including user behavior patterns (e.g., occupancy heat disturbance profiles, equipment start-stop schedules), equipment operational parameters (e.g., runtime, temperature setpoints), and building characteristics (e.g., orientation, window-to-wall ratio)—to third-party

servers.[11,12] This approach not only potentially violates the "data minimization" principle of regulations like the European Union's General Data Protection Regulation but also risks leaking commercial secrets, thereby hindering cross-institutional data collaboration.

Federated Learning (FL) is an emerging privacy-preserving technique that enables model training on decentralized data without centralizing it. In the typical FL framework, an optimal model is cooperatively trained through distributed model training and parameter aggregation coordinated by a central server. The entire training process does not involve any exchange of raw data. Therefore, FL, as a scalable distributed machine learning paradigm, can not only effectively avoid transmitting privacy-sensitive data and protect data subjects' privacy rights but also leverage the local computing and storage resources of all participants to collaboratively build models.[13-15] However, although FL has found numerous applications in mobile devices, healthcare, finance, and other fields, its feasibility and effectiveness in building energy consumption prediction, particularly for VRF systems, have not been fully validated.

To address this, this paper proposes a FL framework integrating Synthetic Minority Oversampling Technique for Regression with Gaussian Noise (SMOGN) data augmentation and genetic algorithm (GA) optimization to enhance the accuracy and generalization capability of VRF system monthly energy consumption prediction while safeguarding data privacy. Specifically, the main contributions of this paper include:

- A privacy-preserving and enhanced FL framework is developed for monthly energy consumption prediction of VRF systems. To address the challenges of data scarcity and sample imbalance commonly encountered in building energy datasets, the Synthetic Minority Over-sampling Technique for Regression with Gaussian Noise (SMOGN) is incorporated into the local training process. Furthermore, a genetic algorithm (GA)-based performance-weighted aggregation strategy is proposed to replace the conventional Federated Averaging (FedAvg) scheme. By optimizing aggregation weights according to the cross-client generalization performance of local models, the proposed method enables more effective knowledge fusion among heterogeneous users, resulting in improved prediction accuracy and robustness of the global model.
- The proposed framework is validated using a real-world dataset comprising 3,232 monthly operational records collected from 518 VRF systems in Guangdong Province, China. The results demonstrate that the framework can achieve accurate monthly energy use prediction while preserving data privacy, highlighting its potential for large-scale deployment across distributed building energy management scenarios.
- An evaluation of three representative machine learning algorithms—Artificial Neural Network (ANN), Light Gradient Boosting Machine (LightGBM), and eXtreme Gradient Boosting (XGBoost) is conducted within the proposed FL framework. Using input variables that encompass building characteristics, VRF operational information, and temporal factors, the study provides insights into the prediction performance, robustness, and suitability of different learning algorithms for federated building energy modeling applications.

## Energy consumption prediction for VRF systems

Current modeling methods for predicting HVAC system energy consumption can be broadly categorized into three types: white-box, grey-box, and black-box models.[16] White-box models are physics-based, using physical equations to analyze and predict energy consumption patterns of HVAC systems through a transparent and quantifiable process. For example, Pachano et al. (2025) built a detailed physical model based on the EnergyPlus platform,[17] encompassing building envelopes, HVAC systems, and their control strategies, and used a genetic algorithm to calibrate model parameters. The results showed that this white-box model met the accuracy requirements during both training and validation phases, demonstrating good consistency and stability, and could serve as a reliable benchmark for black-box models. Grey-box models combine physical mechanisms with data-driven methods, often integrating data analysis techniques within a physical framework to enhance prediction accuracy and adaptability. For instance, Yue et al. (2023) proposed a grey-box model integrating physical equations and machine learning methods for VRF system energy consumption prediction, balancing model interpretability and prediction accuracy.[18] In recent years, data-driven methods have been increasingly applied to HVAC system energy consumption prediction.[19] Black-box models rely entirely on data-driven approaches, establishing mapping relationships by training on input and output samples without requiring physical knowledge. These models include ANN,[20] Support Vector Machines, and Long Short-Term Memory (LSTM) networks, and so on. Such models excel at handling nonlinear problems and have achieved good results in energy prediction tasks. For example, Fan et al. (2022) proposed enhancing the predictive capability of black-box models through transfer learning,[21] demonstrating the superior performance of LGB and XGB based black-box models in building energy prediction, especially after employing a source building data weighting strategy, which reduced CV-RMSE by up to 54%. Mustaffa et al. (2025) proposed combining XGB with eight metaheuristic algorithms (e.g., GA, Particle Swarm Optimization) to systematically optimize model hyperparameters.[22] Based on real campus building data, the study predicted cooling and heating loads separately; results indicated that the GA-XGBoost model performed optimally on metrics like RMSE and MAE.

Although the aforementioned models continue to improve in accuracy, most studies still rely on centralized data training paradigms. However, in practical applications, cross-institutional and cross-building data sharing is often difficult to achieve due to the need to avoid privacy data leakage. This practical issue limits the applicability of traditional data-driven methods and presents a development opportunity for distributed, privacy-preserving modeling methods like federated learning.

## A brief overview of federated learning

Federated Learning is a distributed machine learning paradigm that allows multiple participants (e.g., mobile devices, edge nodes, or organizations) to independently train models using their local data without sharing raw data. Only model parameters or gradient updates are uploaded to a central server for aggregation, thereby collaboratively constructing a global model. This concept was first proposed by McMahan et al. (2016) from Google AI to address privacy protection issues on mobile devices.[23] Federated learning presents several key advantages over traditional centralized learning. Firstly, it ensures data remains on local devices, which effectively safeguards user privacy. Additionally, by leveraging distributed computing resources, it alleviates the computational burden on a central server. Finally, its decentralized nature makes it particularly suitable for data silo scenarios common in sensitive domains like healthcare, finance, and building management.

Recently, many researchers have also introduced FL into the field of building energy prediction. Gao et al. (2021) proposed a decentralized FL framework for residential building load forecasting.[24] They designed a mechanism for local training and gradient broadcasting, avoiding reliance on cloud aggregation. This method achieved about 97% prediction accuracy while preserving privacy. Wang et al. (2023) constructed a clustered FL model for predicting HVAC system capacity requirements,[25] validating the feasibility of FL for building system modeling. Lu et al. (2023) proposed a Federated Reinforcement Learning (FedRL) framework,[26] introducing Q-learning (a model-free reinforcement learning algorithm) into the federated aggregation process to dynamically assign weights to local models, overcoming the performance degradation potential of "simple averaging" in traditional Federated Averaging (FedAvg) methods. Tang et al. (2023) proposed a dynamic clustering FL algorithm combined with shared random mask technology to encrypt sensitive data,[27] suitable for few-shot building energy prediction scenarios, and capable of improving model generalization under data scarcity.

In FL, genetic algorithms are often combined to optimize model aggregation strategies, improve communication efficiency, enhance security and robustness, and address specific domain problems. For example, Chaudhary et al. (2025) proposed a Clustered FL method based on a real-coded genetic algorithm (ReGen-CFL), enhancing model performance and communication efficiency by optimizing hyperparameters (e.g., learning rate) for each cluster.[28] Wu et al. (2025) designed the Federated Client Selection based on Genetic Algorithm (FedCSGA),[29] using GA to optimize the client selection strategy, maximizing the number of participating clients under training time constraints, and improving model accuracy, effectively addressing system and data heterogeneity. Zhai et al. (2023) proposed a genetic algorithm-

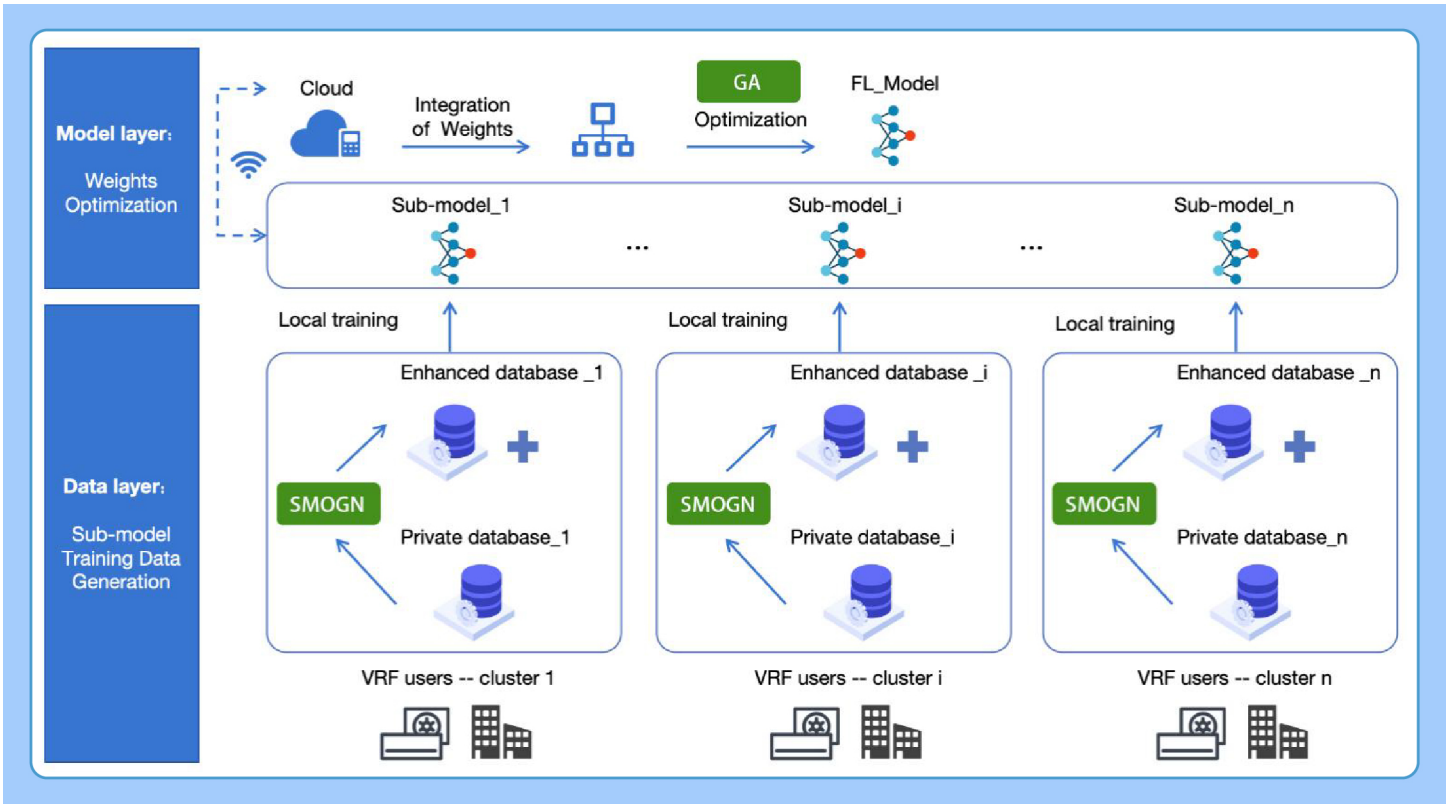



**Figure 1. Federated-learning-based VRF system energy prediction model architecture**

based malicious data detection model to defend against model poisoning attacks; this model improved training accuracy without compromising information loss.[30]

Overall, FL offers a new privacy-safe, distributed, and collaborative approach for building energy prediction, and combining it with optimization methods like GA further enhances its usability and efficiency in practical complex environments. With advancements in computing architectures and privacy-enhancing technologies, federated learning is poised to play a more significant role in the fields of smart building and energy management.

### SMOGN for data augmentation

In machine learning practice, data imbalance is a common issue affecting model performance. SMOGN is a preprocessing method for handling imbalanced regression problems, proposed by Branco et al.[31] in 2017. It combines random under-sampling with two over-sampling strategies: SMOTE for regression (SmoteR) and the introduction of Gaussian noise. Its primary aim is to improve data distribution by generating synthetic samples, enabling learning algorithms to focus better on sparse yet important data, thereby enhancing model predictive performance on imbalanced datasets.

Recently, SMOGN has demonstrated good application results in various fields. Rad et al. (2025) used the SMOGN algorithm to process imbalanced datasets,[32] improving the distribution of target values and enhancing model predictive ability for minority classes by over-sampling minority classes, under-sampling majority classes, and introducing Gaussian noise. Kocoglu et al. (2024) combined SMOGN with the Tree-structured Parzen Estimator optimization algorithm,[33] proposing a novel data-driven framework for handling imbalanced regression problems in shale gas production forecasting, achieving better prediction accuracy compared to traditional methods. Studies have proven that SMOGN can effectively enhance model generalization capability in scenarios with small samples or uneven data distributions.

In this study, the monthly energy consumption data of VRF systems exhibit a significant right-skewed distribution and long-tail outliers, and the values of the target variable vary greatly. Therefore, this paper introduces the SMOGN algorithm to enhance local data, aiming to improve the predictive performance of the FL model in data-scarce or imbalanced scenarios.

## MATERIALS AND METHODS

### Research framework

This paper proposes an enhanced FL framework integrating the SMOGN and GA to improve VRF system monthly energy prediction accuracy while preserving data privacy, as shown in Figure 1. Under this architecture, different VRF user groups optimize their local prediction models using SMOGN data augmentation. The models are then uploaded to the cloud. The aggregation parameters of each model in the FL process are optimized via GA to train the global federated model, to achieve better prediction accuracy. More details of the proposed method are described in next section.

To validate the effectiveness of the proposed "Federated Learning Framework Integrating SMOGN Data Augmentation and GA Optimization" for VRF system monthly energy consumption prediction, this paper designs a systematic experimental process. The overall research framework is shown in Figure 2, including the following four steps:

• Data Collection and User Grouping. This study collected actual operational data from 518 VRF systems in Guangdong Province, China, totaling 3232 monthly records. To simulate the "multi-user, multi-distribution" scenario in FL, we used the K-prototypes algorithm to cluster the data into 3 user groups based on building characteristics and operational patterns, representing different usage types and energy consumption patterns.

• Machine Learning and Model Training. We designed three experimental scenarios: centralized model training, distributed model training, and federated model training.

• Ablation Study. Ablation studies are widely used in machine learning to evaluate the contribution of specific components or features to the overall performance of a model, thereby enhancing the model's interpretability and efficiency. By systematically removing or replacing different parts of the model, researchers can gain a deep understanding of the role and importance of these components in the model's decision-making process.[34,35] This paper designs ablation experiments to progressively verify the contribution of modules such as SMOGN enhancement and genetic algorithm optimization to the overall performance, clarifying the practical utility of key technologies.

• Results Analysis. This paper uses the Coefficient of Determination ($R^2$) and Symmetric Mean Absolute Percentage Error (sMAPE) as evaluation metrics to compare the predictive performance under different modeling scenarios and algorithms (ANN, XGB, LGB).

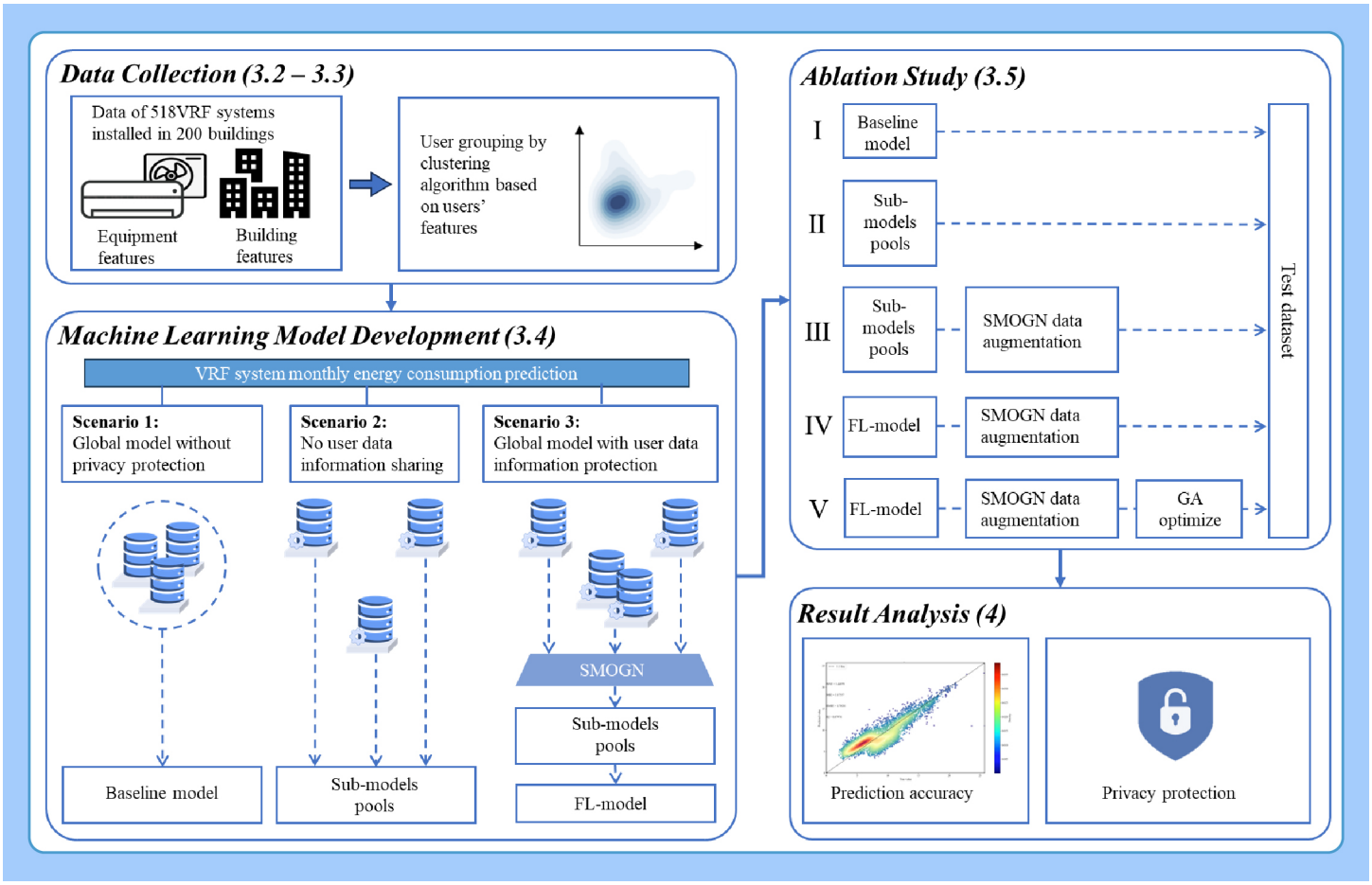



**Figure 2. Research framework of this study**

## Data collection and augmentation

This study collected actual energy consumption data from 518 VRF systems in Guangdong Province, China, totaling 3232 monthly operational records. As shown in Table S1 and Table S2, the dataset includes five building features (building owner type, building floor plan geometry type, building main facade orientation, building window-to-wall ratio, building climate zone), nine equipment features (operating month, equipment type, indoor/outdoor unit capacity ratio, total indoor unit horsepower, total outdoor unit horsepower, monthly average part-load ratio, monthly average indoor air temperature setpoint, monthly weekday average runtime, monthly weekend average runtime), and one target variable (monthly total energy consumption).

The building-related features are derived from textual descriptions of 518 multi-split VRF systems across 200 buildings, provided by after-sales manufacture personnel. We developed a data-linking program that employs fuzzy text retrieval to call an open-source map API and obtain street-view, satellite, and 3-D model data for each building. From these data the building owner type is inferred and stored in the variable bldg_type; the building footprint is manually outlined and its plan geometry classified, assigned to bldg_plan; building window-to-wall ratios are divided into high, medium, or low grades and recorded in bldg_WWR; a Python script computes the angle between the main façade line and true north, with the result saved as bldg_forward; geographic location is finally used to identify the climate zone, stored in bldg_loc. All equipment-related features are retrieved from the VRF system operation management platform.

Analysis of this data reveals that the numerical variables Monthly Total Energy Consumption, Total Indoor Unit Horsepower, and Total Outdoor Unit Horsepower exhibit significant right-skewed distributions and long-tail outlier characteristics. As shown in Figure 3, taking Monthly Total Energy Consumption as an example, its mean is about 2.17 times the median; the standard deviation exceeds the mean, indicating high data dispersion; the maximum value is 11.68 times the third quartile, indicating significant extreme values. Such distribution characteristics are unfavorable for machine learning model training, easily causing the model to overfit the high-frequency range and underfit the high energy consumption range.

To mitigate the impact of extreme values, we applied a natural logarithm plus one transformation to the severely skewed numerical variables, as per Equation (1):

$$x' = \ln(1 + x) \tag{1}$$

Where $x$ represents the original feature value, and $x'$ represents the transformed feature value. Adding 1 avoids the undefined log(0) when x=0. The log transformation can improve the symmetry of the data distribution by compressing large values, smoothing the distribution, and mitigating the influence of outliers. It enhances the stability and predictive performance of linear regression models without losing information from zero values.[36]

Subsequently, we performed Min-Max normalization on all numerical variables, scaling them to the [0, 1] interval to eliminate scale differences. Categorical variables were encoded using One-Hot Encoding to ensure uniform feature processing by the model. All transformation parameters (e.g., encoders, normalizers) were saved as global configuration files to ensure consistency in the feature space between the training and test sets, and to facilitate subsequent inverse transformations for restoring the original scale during error calculation.

To further alleviate data scarcity and uneven distribution issues, this paper introduces the SMOGN algorithm (code provided in the Appendix) to augment the training set. SMOGN generates synthetic samples in sparse regions of the target variable by combining SmoteR (regression interpolation) and Gaussian noise injection, thereby enhancing the model's ability to identify rare energy consumption patterns. The process first identifies sparse regions based on the density of the target variable (Monthly Total Energy Consumption); then performs random under-sampling on dense regions; and after that performs over-sampling on sparse regions to generate new samples; finally, the original training set is expanded from 2585 records to

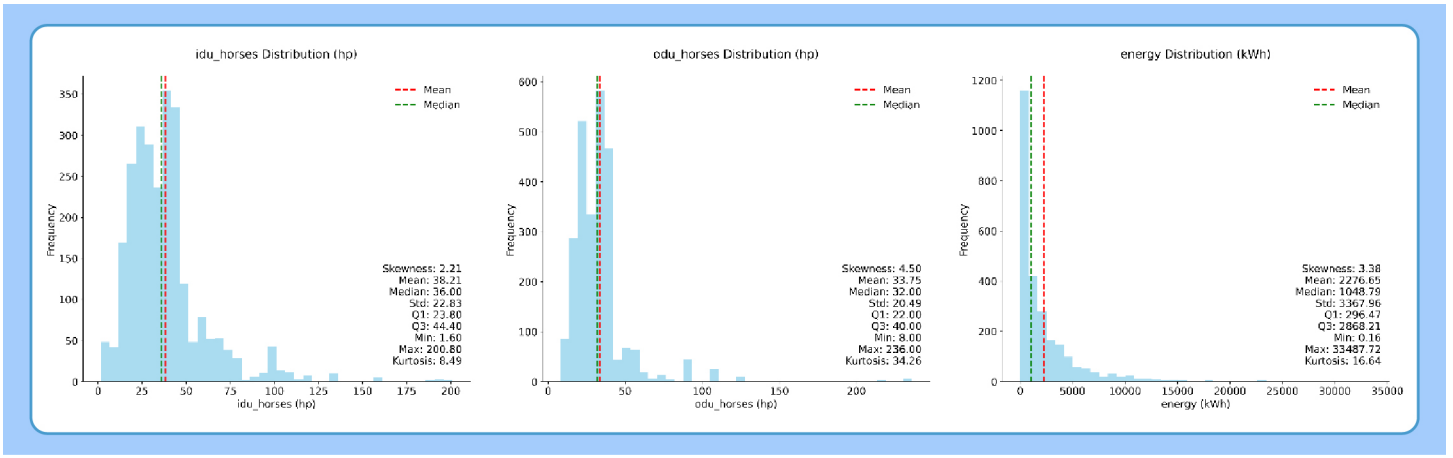



**Figure 3. Right-Skewed Distributions of total indoor/outdoor unit horsepower and monthly total energy consumption**

approximately 4500 records, an enhancement ratio of about 74%. The augmented data distribution is more balanced, with the sample density in the high energy consumption range significantly increased, providing more comprehensive coverage for subsequent model training.

Taking the target variable Monthly Total Energy Consumption as an example, its original data showed a significant right-skewed distribution and long-tail characteristics, which are unfavorable for machine learning model training and generalization. After log transformation and SMOGN data augmentation processing, the data distribution tends towards symmetry, the density of tail samples is significantly increased, and the impact of extreme values is effectively mitigated, as shown in Figure 4.

## User grouping and clustering

To simulate the process of federated learning aggregating distributed models, we first split the original data into a base training set (train_data_0, 2585 records) and a base test set (test_data_0, 647 records) in an 8:2 ratio. Then, the base training set was clustered based on user characteristics to represent different user groups.

Given that the data in this study contains both numerical and categorical variables, traditional clustering algorithms like K-means or K-modes, which are only suitable for single-type data, cannot directly handle the mixed attribute structure. The K-prototypes algorithm combines K-means and K-modes, enabling unified clustering of mixed-type data. It offers advantages such as simple structure, computational efficiency, and no need for variable type conversion,[37,38] so we adopted the K-prototypes algorithm.

In this dataset, a building may have one or more VRF systems, and the system ID represents different users. We aggregated the data with the same system ID in train_data0: for numerical data, we took the median; for categorical data, we took the mode. This resulted in 515 aggregated records. We then used the K-prototypes algorithm to cluster these into 3 groups. Based on the system IDs within each cluster, we restored the aggregated data back to the original record level, finally obtaining three subsets: Cluster_1 (744 records), Cluster_2 (470 records), and Cluster_3 (1371 records). These three subsets represent three user groups.

To facilitate subsequent model training, we further randomly split each of these three subsets into training and test sets in an 8:2 ratio. This yielded: for User Group 1, training set train_data_1 (595 records) and test set test_data_1 (149 records); for User Group 2, training set train_data_2 (376 records) and test set test_data_2 (94 records); and for User Group 3, training set train_data_3 (1096 records) and test set test_data_3 (275 records).

Within the FL framework, each user group independently trains models locally and uploads parameters to the central server for aggregation. No raw data is shared throughout the process, ensuring privacy and security.

## Model development

To systematically evaluate the effectiveness of the proposed method, this paper designed three scenarios: centralized model training, distributed model training, and federated model training, simulating different data sharing and collaboration conditions in reality. The predictive performance of three mainstream machine learning algorithms (ANN, XGB, LGB) was compared under each scenario.

- Centralized Model Training, simulating the scenario where user data is shared. The base training set train_data0 is used for model training to obtain the baseline model M0, which is then evaluated using the base test set test_data0.
- Distributed Model Training, simulating the scenario where cross-institutional collaboration is impossible due to privacy concerns. The training sets of the three user groups obtained above are used for model training separately, resulting in sub-models $M_1$, $M_2$, $M_3$. These models are evaluated using their respective test sets.
- Federated Model Training, investigating whether FL can achieve prediction accuracy comparable to centralized models while avoiding privacy issues. In this scenario, we designed two aggregation methods for comparison. The first is Simple Averaging Ensemble, where the aggregation weight for each sub-model is set to 1/3, resulting in the simple weighted ensemble model MF-ave, as shown in Equation (2).

$$\hat{y}FL(x) = \sum_{i=1}^{3} \frac{1}{3} \cdot \hat{y}_i(x) \tag{2}$$

Where $\hat{y}FL(x)$ represents the predicted output of the federated aggregation model for the input feature vector x, $\hat{y}_i(x)$ represents the predicted output of the i-th sub-model for input x, and $\frac{1}{3}$ represents the weight of each sub-model.

The second method is the novel GA-optimized Multi-model Performance Weighted Ensemble, as detailed in Figure 5. Vector A represents the energy consumption prediction results of all sub-models for each user group. Vector W represents the weight $\omega_1$, $\omega_2$ … $\omega_n$ of each sub-model in the federated model. Vector C represents the weighted cross-model energy consumption prediction result. To optimize the weight vector W, for each sub-model Mj corresponding to weight component $\omega_n$, we calculate its prediction error (R2) across all user groups and use their mean (Equation (3)) as part of the optimization objective.

$$R^2_{mean}(W) = \frac{1}{n}\sum_{j=1}^{n}\left(1 - \frac{\sum_{i=1}^{n}\left(C_j^{(i)} - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2}\right) \tag{3}$$

Where $C_j^{(i)}$ is the predicted value of sub-model Mj for user group Cluster_j. $\hat{y}_i$ is the true historical energy consumption value for user group Cluster_j. $\overline{y}$ represents the average of the historical energy consumption actual values for user group Cluster_j.

To encourage diversity in the weight distribution, an entropy regularization term is introduced as Equation (4). The final objective function is defined in Equation (5) as minimizing the loss function incorporating the regularization term.

$$H(W) = -\sum_{i=1}^{n} \omega_i \ln(\omega_i) \tag{4}$$

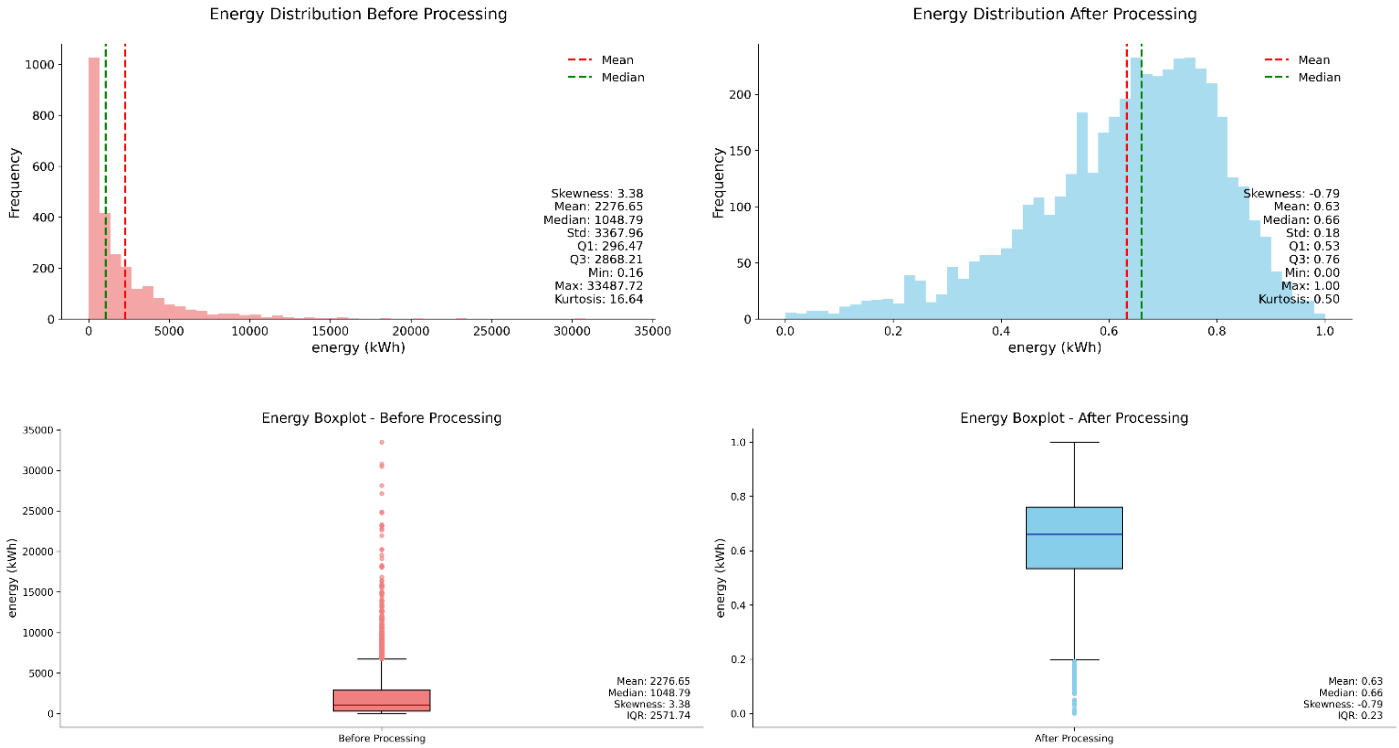



**Figure 4. Monthly total energy consumption distribution comparison histogram & monthly total energy consumption distribution comparison boxplot.**

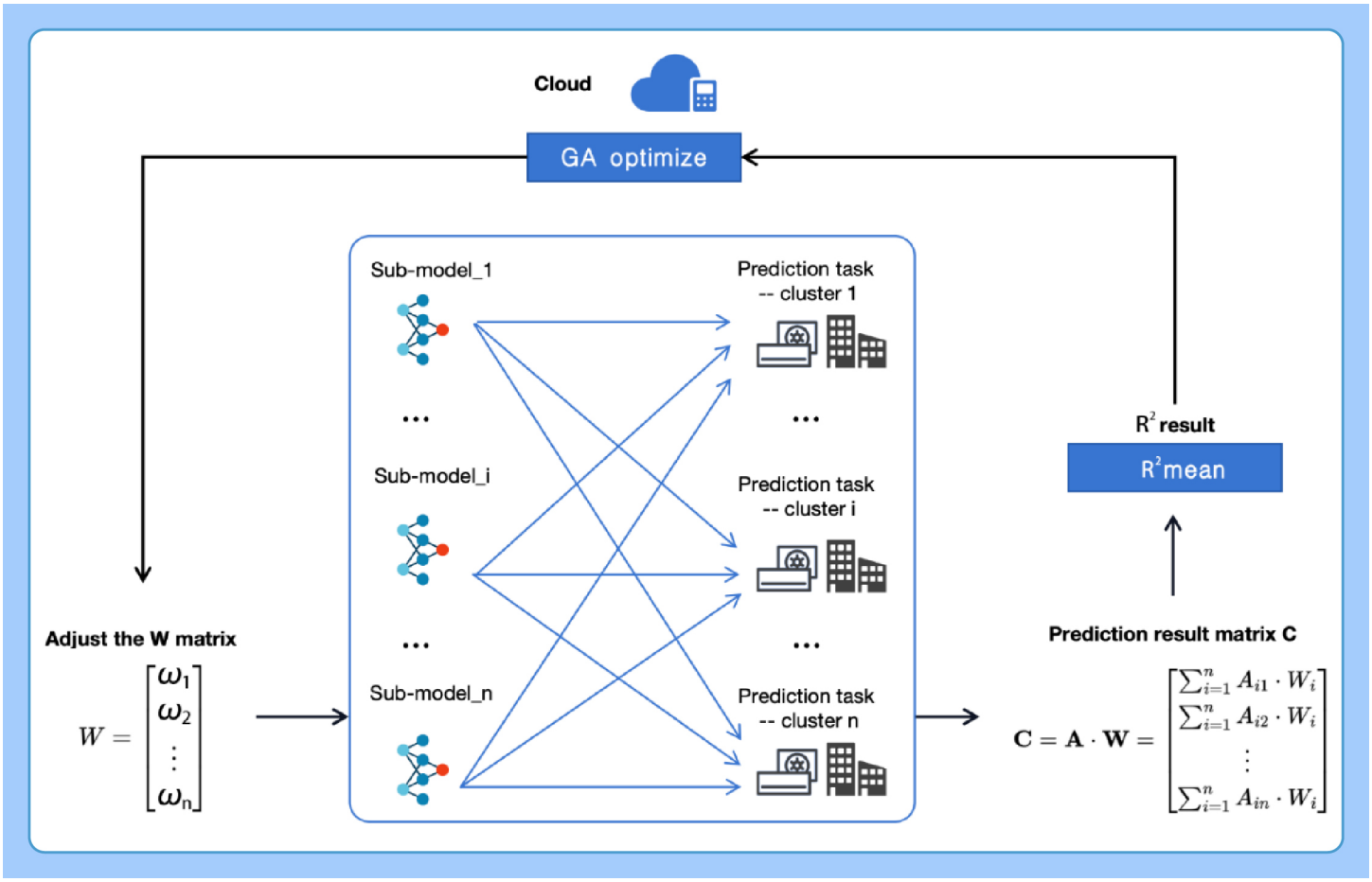



**Figure 5. Optimization of energy consumption prediction model based on genetic algorithm**

$$L(W) = R^2_{mean}(W) - \lambda H(W) \quad (5)$$

Where λ is the weight coefficient for entropy regularization, which can be

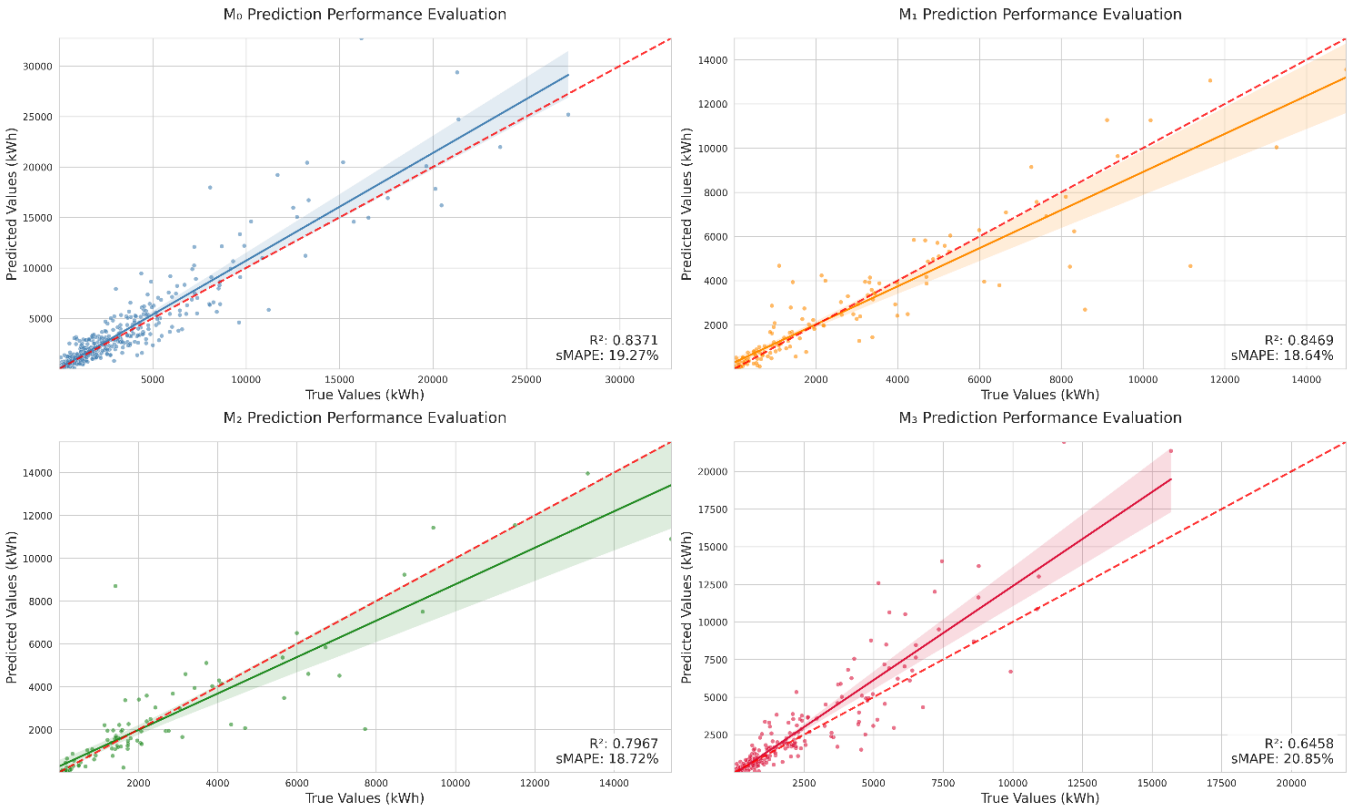



**Figure 6. Comparison of ANN model prediction performance evaluation for the testing data before data augmentation**

adjusted based on the actual situation.

We used GA (code provided in the appendix) to continuously adjust the weight matrix W to minimize the objective function, as shown in Equation (6). After obtaining the optimized weights $\omega_1$, $\omega_2$ and $\omega_3$ for each sub-model via GA, the sub-models are aggregated to form model MF-ga. Finally, the base test set test_data0 is used to evaluate both model MF-ave and MF-ga.

$$W^* = argmin_B L(W) \quad (6)$$

### Ablation study design

To verify the contribution of modules such as SMOGN enhancement and GA optimization to the overall performance, this paper designed the following ablation experiments.

• Under Scenario 1 (shared user data), trained the baseline model $M_0$ with the base training set train_data0, and accessed its prediction accuracy using the base test set test_data0.

• Under Scenario 2 (no shared user data), trained sub-models $M_1$, $M_2$, $M_3$ with training set train_data_1, training set train_data_2, and training set train_data_3, respectively. Then evaluated the prediction performance of distributed models on the testing data within their respective user groups.

• To evaluate the performance improvement brought by the SMOGN data augmentation method under Scenario 2 (no shared user data), we firstly trained the sub-models $M_{1\_enhanced}$, $M_{2\_enhanced}$, $M_{3\_enhanced}$ using augmented training datasets. In this process, the training set of each sub-model was uniformly expanded to 1500 records. Subsequently, the prediction performance of these augmented sub-models was evaluated with the same user-group testing data as in experiment 2.

• Performed a simple average ensemble of $M_{1\_enhanced}$, $M_{2\_enhanced}$, $M_{3\_enhanced}$ and evaluated the prediction performance of conventional federated model MF-ave without GA optimization using the base test set test_data0.

• Performed the GA-optimized multi-model performance weighted ensemble of $M_{1\_enhanced}$, $M_{2\_enhanced}$, $M_{3\_enhanced}$ and evaluated the prediction performance of novel federated model MF-ga using the base test set test_data0.

During the training of all the above models, the global categorical feature encoder was loaded to unify the feature scale. During the evaluation of all models, the global normalization file was loaded to reverse the normalization of the data, and then the target variable (Monthly Total Energy Consumption) underwent an inverse log transformation using Equation (7). This ensures that error metrics calculated on the original scale have practical physical meaning.

$$x = exp(x') - 1 \quad (7)$$

Where $x'$ represents the transformed feature value, $x$ represents the value restored to the original scale, and $exp(x')$ is the natural exponential function.

## RESULTS

### Prediction performance of baseline model and distributed models

After training and performance evaluation, the predictive performance of the baseline model $M_0$ and the distributed sub-models $M_1$, $M_2$, $M_3$ is shown in Table S3. The prediction performance evaluation charts of models trained using the ANN algorithm are shown in Figure 6 to demonstrate the model performance before data augmentation.

The data in the table show that the centralized baseline model $M_0$ does not always outperform the distributed sub-models. For example, with the ANN algorithm, $M_1$ ($R^2$ = 0.8469) > $M_0$ ($R^2$ = 0.8371); with the LGB algorithm, $M_3$ ($R^2$ = 0.8920) > $M_0$ ($R^2$ = 0.8913). This phenomenon confirms the heterogeneity of data distribution in the real world. For Cluster_1 and Cluster_3, their local data possess unique patterns, and models trained based on this can more accurately capture the specific laws of that group. The centralized baseline model $M_0$, seeking a global optimum across all data, may smooth out these local specificities, resulting in a slight performance compromise on specific groups. Furthermore, the performance of different sub-models varies significantly. Even for the same algorithm (e.g., XGB), the $R^2$ difference across different user groups can exceed 0.15 ($M_2$ $R^2$ = 0.7494 vs. $M_3$ $R^2$ = 0.9033). This performance fluctuation further emphasizes the significant impact of data heterogeneity on model generalization ability and poses a challenge for subsequent federated aggregation algorithms.

### Prediction performance of distributed sub-models after SMOGN data augmentation

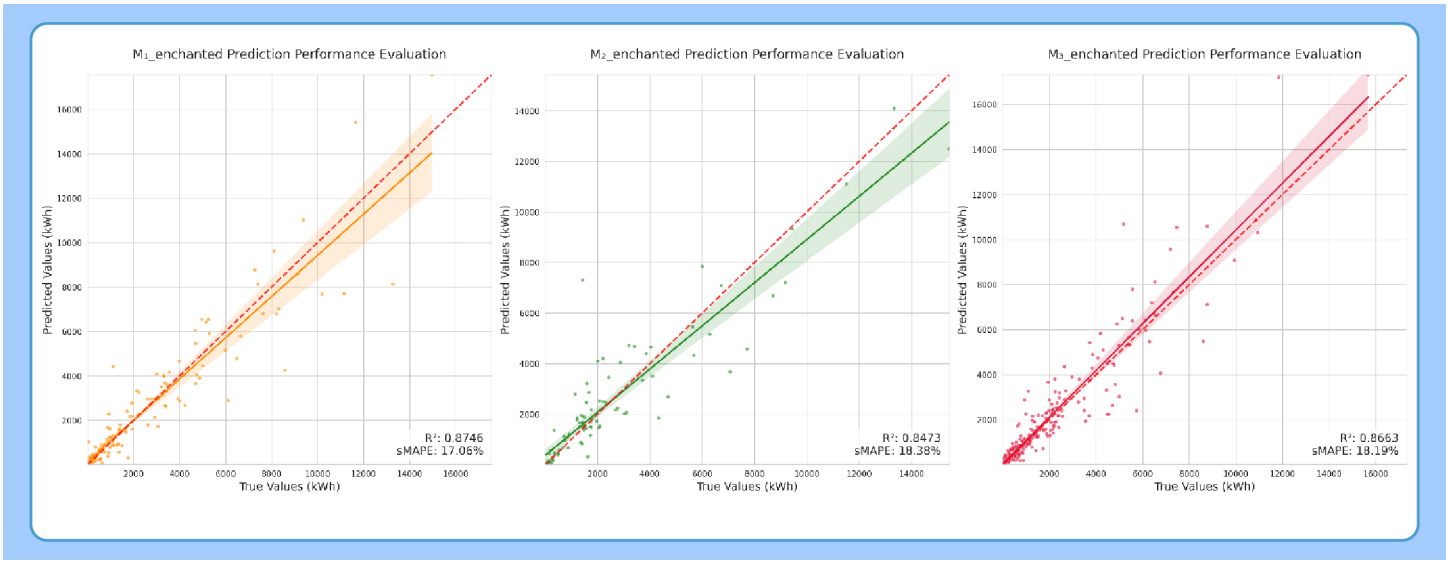



**Figure 7. Comparison chart of ANN model prediction performance evaluation for the testing data after data augmentation**

After SMOGN data augmentation, the predictive performance and relevant information of the sub-models $M_{1\_enhanced}$, $M_{2\_enhanced}$, $M_{3\_enhanced}$ are shown in Table S4. Bayesian optimization algorithm was used to tune the hyperparameters. The prediction performance evaluation charts of models trained using the ANN algorithm is shown in Figure 7 to demonstrate the distributed sub-models' performance after data augmentation.

Table S4 show that after SMOGN data augmentation and hyperparameter optimization, the performance of all local sub-models reached a high and balanced level. Compared with $R^2$ of M1, M2, and $M_3$ in Table S3, the performance improvement for each sub-model is significant, especially for originally weaker models. Taking the ANN model as an example, the $R^2$ of $M_3$_enchanted increased substantially from 0.6458 before augmentation to 0.8797. This indicates that SMOGN effectively compensated for the lack of original data in this group by generating representative synthetic samples, allowing the model to learn more comprehensive data patterns. After data augmentation, the $R^2$ metrics of the three sub-models under the same algorithm are all stable above 0.85, and the performance differences between different user groups are significantly reduced. SMOGN data augmentation not only increased data volume but also improved the quality and distribution of each client's local data. It enabled producing high-performance and balanced sub-models, providing a solid foundation for a high-quality federated global model.

### Prediction performance of FL model with simple averaging ensemble and with multi-model weighted ensemble optimized by GA

The augmented sub-models $M_{1\_enhanced}$, $M_{2\_enhanced}$, $M_{3\_enhanced}$ were aggregated using simple averaging to obtain MF-ave, and were aggregated using GA-optimized multi-model performance weighted ensemble to obtain MF-ga. Their predictive performance was compared with the traditional centralized baseline model $M_0$, as shown in Figure 8.

As shown in Figure 8, on the core accuracy metric $R^2$, the MF-ga model consistently outperforms the MF-ave model across all three learners (ANN, LGB, XGB). This improvement is particularly pronounced for the ANN model ($R^2$ increased from 0.2530 to 0.5135), demonstrating that the GA algorithm can effectively integrate the strengths of each local model by optimizing the aggregation weights, thereby significantly mitigating the performance degradation caused by the simple averaging method in non-IID data scenarios. However, the optimization effect on the sMAPE metric is mixed (e.g., decreased for XGB, slightly increased for LGB and ANN). This suggests that the primary advantage of the GA optimization strategy lies in improving the model's coefficient of determination ($R^2$), i.e., better capturing the overall variance in the data, while the optimization on the symmetric mean absolute percentage error varies depending on the model.

In comparison with the centralized baseline model $M_0$, the MF-ga model based on LGB performed most outstandingly. Its $R^2$ value reached 0.8390, closest to the centralized model's 0.8913, with a significantly reduced gap. Considering that the FL framework completely avoids sharing raw data, adhering to strict privacy protection principles, while the performance upper bound of the centralized model is achieved at the cost of sacrificing this principle, the accuracy level achieved by the LGB-based MF-ga model can be considered comparable to the centralized model's performance under privacy protection constraints, possessing high practical value for engineering applications.

## CONCLUSION

To overcome data privacy concerns, data imbalance, and client heterogeneity in monthly VRF system energy consumption prediction, this study proposes and validates a novel federated learning framework that combines SMOGN-based data augmentation with GA-optimized model aggregation. It empirically demonstrated the effectiveness of the proposed "SMOGN + GA + FL" framework with actual operational data from 518 VRF systems in Guangdong Province, China. Experiments showed that under strict privacy protection constraints (without sharing any raw data), the global federated learning model trained by this framework (especially the GA-optimized LightGBM model) achieves prediction accuracy comparable to the centralized model that requires pooling all data. This provides a practical solution for achieving privacy-safe collaborative energy management in the smart buildings. Within the proposed framework, the SMOGN algorithm, by adaptively augmenting data based on the scarcity situation of different clients, significantly improved the performance of each local model and makes them more balanced. This step laid a solid foundation for subsequent high-quality federated aggregation, proving the importance of data augmentation for FL in data-heterogeneous scenarios. In addition, in our FL experiments, the proposed GA-optimized ensemble strategy yielded a superior $R^2$ compared to the conventional simple average aggregation method. The GA-optimized LightGBM model demonstrated the highest prediction accuracy in most cases, making it the preferred algorithm for building both local and global FL models.

The future work of this study are as follows: (1) further explore more complex federated learning aggregation architectures, such as dynamic clustering or meta-learning-based aggregation methods, to maintain excellent performance even under more extreme data heterogeneity; (2) apply this framework to more challenging time-series prediction scenarios and explore how to effectively incorporate temporal features to enhance the prediction capability for dynamic energy consumption of VRF systems; (3) validate the framework on larger-scale, more diverse building datasets to further examine its universality and robustness.

## REFERENCES

1. Lin G., Casillas A., Sheng M., et al. (2023). Performance Evaluation of an Occupancy-Based HVAC Control System. *Energies* **16**:7088. DOI:10.3390/en16207088


**Figure 8. Comparison of the performance of centralized baseline model, simple weighted federated model, and genetic algorithm-optimized weighted ensemble federated model for the testing data**


2. United Nations Environment Programme. (2024). 2023 Global Status Report for Buildings and Construction: Beyond Foundations – Mainstreaming Sustainable Solutions to Cut Emissions from the Buildings Sector. *UNEP*. DOI:10.59117/20.500.11822/45095.
3. National Development and Reform Commission, Ministry of Housing and Urban-Rural Development. (2022). 14th Five-Year Plan for Building Energy Efficiency and Green Building Development. *China Architecture & Building Press*.
4. Shin M., Kim S., Kim Y., et al. (2024). Development of an HVAC system control method using deep reinforcement learning. *Build. Environ.* **248**:111069. DOI:10.1016/j.buildenv.2024.111069
5. Hagström F., Garg V., Oliveira F., et al. (2025). Employing federated learning for training autonomous HVAC systems. *Energy Build.* **340**:115761. DOI:10.1016/j.enbuild.2025.115761
6. Enteria N., Sawachi T., Saito K., et al. (2023). Variable refrigerant flow systems: advances and applications. *Springer Nature Singapore*. DOI:10.1007/978-981-99-4783-2.
7. Yesilyurt H., Dokuz Y., Dokuz A.S., et al. (2024). Data-driven energy consumption prediction of a university office building. *Energy* **310**:133242. DOI:10.1016/j.energy.2024.133242
8. Hsu P.C., Gao L., Hwang Y., et al. (2025). Comparative study of LSTM and ANN models for power consumption prediction of variable refrigerant flow (VRF) systems. *Int. J. Refrig.* **169**:55–68. DOI:10.1016/j.ijrefrig.2025.02.014
9. Oh K. and Kim E.J. (2024). Predicting the energy consumption of a VRF heat pump using manufacturer performance data. *Energy Build.* **303**:113798. DOI:10.1016/j.enbuild.2024.113798
10. Wang R., Bai L., Rayhana R., et al. (2024). Personalized federated learning for buildings energy consumption forecasting. *Energy Build.* **323**:114762. DOI:10.1016/j.enbuild.2024.114762
11. Liu H., Liu Y., Huang H., et al. (2024). Energy consumption dynamic prediction for HVAC systems based on feature clustering deconstruction. *Build. Simul.* **17**:1439–1460. DOI:10.1007/s12273-024-00892-7
12. Ali M., Kumar A. and Choi B.J. (2025). Privacy preserving federated learning for energy disaggregation of smart homes. *IET Cyber-Phys. Syst.: Theory Appl.* **10**. DOI:10.1049/cps2.70013.
13. Mulik A., Mehta D., Mirgh R., et al. (2023). A federated learning approach to predict energy consumption. *ICCCIS*:187–192. DOI:10.1109/ICCCIS60361.2023.10344237.
14. Venkataramanan V., Kaza S., Annaswamy A.M., et al. (2023). DER forecast using privacy-preserving federated learning. *IEEE Internet Things J.* **10**:2046–2055. DOI:10.1109/JIOT.2022.32280
15. Khalil M., Esseghir M., Merghem-Boulahia L., et al. (2021). Federated learning for energy-efficient thermal comfort service. *IEEE GLOBECOM*:1–6. DOI:10.1109/GLOBECOM46548.2021.9685634.
16. Zhou X., Wang N., Zou J., et al. (2024). Analysis and prediction of energy consumption in office buildings with variable refrigerant flow systems. *J. Build. Eng.* **97**:110936. DOI:10.1016/j.jobe.2024.110936
17. Pachano J.E., Nuevo-Gallardo C. and Fernández Bandera C. (2025). An empirical comparison of a calibrated white-box versus multiple LSTM black-box building energy models. *Energy Build.* **333**:115485. DOI:10.1016/j.enbuild.2025.115485
18. Yue B., Wei Z., Zheng C., et al. (2023). Power consumption prediction of variable refrigerant flow system through data-physics hybrid approach. *Energy* **278**:127826. DOI:10.1016/j.energy.2023.127826
19. Liu Q., Yan Y., Jin Y., et al. (2024). Secure federated evolutionary optimization-a

survey. *Engineering* **34**:23–42. DOI:10.1016/j.eng.2024.02.004
20. Li A., Xiao F., Fan C., et al. (2021). Development of an ANN-based building energy model using transfer learning. *Build. Simul.* **14**:89–101. DOI:10.1007/s12273-020-00677-9
21. Fan C., Lei Y., Sun Y., et al. (2022). Data-centric or algorithm-centric: Exploiting the performance of transfer learning for improving building energy predictions in data-scarce context. *Energy* **240**:122775. DOI:10.1016/j.energy.2021.122775
22. Mustaffa Z., Sulaiman M.H. (2025). Advanced forecasting of building energy loads with XGBoost and metaheuristic algorithms. *Energy Storage Sav.* DOI:10.1016/j.enss.2025.1000305.
23. McMahan H.B., Moore E., Ramage D., et al. (2016). Communication-efficient learning of deep networks from decentralized data. arXiv:1602.05629. DOI:10.48550/arXiv.1602.05629.
24. Gao J., Wang W., Liu Z., et al. (2021). Decentralized federated learning framework for the neighborhood: a case study on residential building load forecasting. *19th ACM Conf. Embed. Net. Sens. Sys.*:453–459. DOI:10.1145/3460312.3482513.
25. Wang Z., Yu P. and Zhang H. (2023). Privacy-preserving regulation capacity evaluation for HVAC systems. *IEEE Trans. Smart Grid* **14**:3535–3549. DOI:10.1109/TSG.2023.32422
26. Lu Y., Cui L., Wang Y., et al. (2023). Residential energy consumption forecasting based on federated reinforcement learning. *Comput. Model. Eng. Sci.* **137**:717–732. DOI:10.32604/cmes.2023.02876
27. Tang L., Xie H., Wang X., et al. (2023). Privacy-preserving knowledge sharing for few-shot building energy prediction. *Appl. Energy* **337**:120860. DOI:10.1016/j.apenergy.2023.120860
28. Chaudhary R.K., Kumar R. and Saxena N. (2025). Ameliorating clustered federated learning using real-coded genetic algorithm. *Cluster Comput.* **28**:415. DOI:10.1007/s10586-024-04867-1
29. Wu J., Ji H., Yi J., et al. (2025). Optimizing client selection in federated learning base on genetic algorithm. *Cluster Comput.* **28**:400. DOI:10.1007/s10586-024-04869-z
30. Zhai R., Chen X., Pei L., et al. (2023). A federated learning framework against data poisoning attacks. *Electronics* **12**:560. DOI:10.3390/electronics12030560
31. Branco P., Torgo L., Ribeiro R.P. (2017). SMOGN: a pre-processing approach for imbalanced regression. *Int. Workshop Learn. Imba. Dom. PMLR*:36–50. DOI:10.48550/arXiv.1704.03466.
32. Rad M., Rafiei A., Grunwell J., et al. (2025). Tackling the small imbalanced horizontal dataset regressions by stability selection and SMOGN. *Int. J. Med. Inform.* **196**:105809. DOI:10.1016/j.ijmedinf.2024.105809
33. Kocoglu Y., Gorell S.B., Emadi H., et al. (2024). Enhancing shale gas EUR predictions with TPE optimized SMOGN. *Gas Sci. Eng.* **131**:205475. DOI:10.1016/j.gsxe.2024.205475
34. Meyes R., Lu M., Puiseau C.W., et al. (2019). Ablation studies in artificial neural networks. arXiv:1909.00608. DOI:10.48550/arXiv.1909.00608.
35. Zhao Y., Chen W., Xu Z., et al. (2025). AbGen: evaluating large language models in ablation study design. arXiv:2504.09641. DOI:10.48550/arXiv.2504.09641.
36. Kamalov F., Moussa S., Reyes J.A. (2023). Data transformation in machine learning. *2023 Int. Conf. Innov. Intel. Inform. Comput. Tech.*:115–120. DOI:10.1109/3ICT58783.2023.10330167.
37. Ji J., Pang W., Zhou C., et al. (2012). A fuzzy k-prototype clustering algorithm for mixed numeric and categorical data. *Knowl.-Based Syst.* **30**:129–135. DOI:10.1016/j.knosys.2012.01.015
38. Ji J., Bai T., Zhou C., et al. (2013). An improved k-prototypes clustering algorithm for mixed numeric and categorical data. *Neurocomputing* **120**:590–596. DOI:10.1016/j.neucom.2013.04.011

## FUNDING AND ACKNOWLEDGMENTS

This work was supported by scientific research start-up funds grant QD2024005C from Tsinghua Shenzhen International Graduate School, Tsinghua University. The data used in this study were provided by Midea Group, with permission granted for their use in this research.

## AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Zhipeng Guo], [Bihai Wang], [Yu Jiang] and [Leqi Zhu]. The first draft of the manuscript was written by [Zhipeng Guo†], [Bihai Wang†] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. † These authors contributed equally to this work.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## DATA AND CODE AVAILABILITY

The data and code that support the findings of this study are not publicly available due to data privacy and confidentiality restrictions.

## SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.59717/ipj.energy-use.2026.100060