Human-Machine Collaborative Control for Smart Homes via Whale-Optimized Iterative Learning and Multimodal Fusion

Xuanzhang Zhu[1], Yafei Li[2*]
[1]Information and Network Center, Hunan University of Science and Engineering, Yongzhou, 425199, China
[2]School of Information Engineering, Hunan University of Science and Engineering, Yongzhou, 425199, China
E-mail: zxz@huse.edu.cn, lyf07200601@126.com
*Corresponding author

Keywords: multimodal fusion, smart home, human-machine collaboration, whale optimization algorithm, iterative learning control

Received: May 19, 2025

This study proposes a multimodal collaborative control method based on an improved whale optimization algorithm and iterative learning to address the issues of insufficient multimodal fusion and poor adaptability to dynamic environments in smart home human-machine collaborative control. Firstly, by introducing a dynamic learning gain mechanism to optimize the iterative learning control algorithm, the convergence speed and tracking accuracy of the system can be improved; Secondly, a feature level and decision level fusion strategy is adopted to achieve effective fusion of speech and gesture modalities; Finally, a complete smart home human-machine collaborative control system architecture is constructed. 1) In terms of control accuracy, the research method achieves average control accuracy of 0.9212 and 0.9053 in single-device and multi-device scenarios, respectively, significantly better than particle swarm optimization genetic algorithm (0.8751) and grey wolf optimization backpropagation network (0.8346). 2) In terms of error indicators, the maximum mean absolute error (0.167) and root mean square error (0.196) are reduced by more than 50% compared to particle swarm optimization genetic algorithm (0.373/0.338) and grey wolf optimization backpropagation network (0.337/0.324). 3) In terms of system performance, the accuracy recall curve area (0.9758) is improved by 5.45%-13.68% compared to the comparison methods, the system resource utilization rate is 0.054%-0.131%, and the average response time (10.31-24.12ms) is improved by more than 30% compared to particle swarm optimization genetic algorithm (18.89ms) and grey wolf optimization backpropagation network (16.21ms). The research provides a high-precision and low latency human-machine collaborative control solution for the field of smart homes.

Povzetek: Predlagan je multimodalni kontrolni sistem za pametne domove, ki združuje Whale Optimization Algorithm (WOA) z Iterative Learning Control (ILC) za izboljšanje kvalitete, odzivnosti in prilagodljivosti v dinamičnih okoljih.

1 Introduction

As Internet of Things (IoT) and artificial intelligence technology continuously develop, the intelligence level of home devices used in people's lives is becoming increasingly high [1]. Smart homes have brought great convenience to people's lives, reducing the pressure in their daily lives and deepening the dependence of different groups on smart homes [2]. With the increasing popularity of smart homes, people's control requirements for smart homes have become increasingly strict, and the convenience requirements for controlling smart homes have become higher. Therefore, how to design (HMCC) for smart homes has become a huge challenge at present. In traditional smart home control, machine learning and deep neural networks are commonly used to construct collaborative control models for smart homes. These traditional construction methods can effectively control smart homes and improve the convenience of smart home control for people [3]. The collaborative control methods of traditional smart homes have good control effects on a small number of smart homes, but have weak control effects on multiple types of home systems. Therefore, there is an urgent need for a method to achieve HMCC of smart home systems. Multimodal Fusion (MF) can unify and integrate data from different modalities, thereby improving the ability to process and understand information. It can be combined with speech and gestures for human-computer interaction [4]. Iterative Learning Control (ILC) can flexibly respond to dynamic changes in the system. When controlling the system, it requires fewer parameters to control the dynamic system and has a fast convergence speed. However, its collaborative control performance for the system is poor [5]. The Whale Optimization Algorithm (WOA), as a population-based intelligent optimization algorithm, is easy to operate and can greatly avoid falling into the trap of local optimal solutions, effectively compensating for the shortcomings of ILC [6-7]. In summary, the research problem lies in the poor performance of existing methods in multi-device collaborative control, especially their insufficient adaptability in dynamic environments. At the same time, it also lies in how to effectively integrate multimodal inputs such as voice and gestures to improve the accuracy and robustness of control. Therefore, the study introduces the WOA for improved ILC and combines MF for HMCC of smart homes, hoping to improve the HMCC capability of smart homes. The goal of the research is to solve the problem of insufficient control of traditional methods in multi-device and dynamic environments, and to apply it to resource constrained embedded devices. The proposed method provides a reference for domestic and foreign scholars to study the HMCC system of smart homes, and promotes the continuous progress and improvement of the HMCC system of smart homes.

2 Related work

With the popularization of smart homes, research on smart home systems has received widespread attention from scholars both domestically and internationally. Dong et al. developed a multimodal neural processing system based on memristor circuits to address the issues of high implementation costs and high power consumption in traditional smart home monitoring systems. By designing a multimodal sensory processing module, a memristor crossbar array was constructed. The results indicated that the occupancy rate of the constructed smart home monitoring system was 1.25% [8]. Wei et al. proposed a smart home energy management method based on deep reinforcement learning to address the difficulty in designing energy management strategies for smart home systems. Deep reinforcement learning algorithms were trained using approximate strategy optimization methods, and device action generation was designed using strategy networks that output discrete and continuous actions [9]. Ameer et al. developed a role-based attribute based access control model to address the challenges faced by smart homes in terms of access control. By combining a family centered approach with an attribute centered approach, an IoT access control model was explained, and a hybrid model was used to construct an access control system [10]. Perumal et al. developed an IoT-based smart home recognition system to address the lack of visualization capabilities in smart home recognition systems. By tracking activity data in smart home environments, real-time data collected by system sensors were transmitted to IoT edge servers [11].

In addition, research on MF has also received widespread attention from scholars both domestically and internationally. Zhou et al. developed a multi-task perception network based on hierarchical MF to improve the fusion and segmentation accuracy of multimodal features in assisted driving. An MF module was constructed to enhance feature fusion and an advanced semantic module was built for extracting semantic information. The experiment findings denoted that the accuracy of the proposed method reached 91.32% [12]. Chen et al. proposed an MF strategy based on graph neural networks to address the issue of heterogeneity between modalities in traditional multimodal detection methods for severe depression. Modal features were extracted by constructing a modal specific graph neural network architecture and a reconstruction network was utilized to determine individual modal fidelity [13]. Lu et al. developed an internal defect detection method based on MF convolutional neural network to address the issue of low accuracy in magnetic tile manufacturing. An end-to-end approach was utilized for network training and features were extracted from modal data. The findings showed that the internal defect detection accuracy of the proposed method reached 94.32% [14]. Fang et al. developed an MF model based on multi-attention mechanism to design an efficient and robust depression detection model. Long short-term memory networks were utilized for learning audio and visual features, and MF was utilized for feature delivery. The results indicates that the Root Mean Square Error (RMSE) of the detection model designed by the research was only 0.468 [15]. Dalila C et al. designed a multimodal feature fusion method based on artificial neural networks to achieve the best level of security for human recognition and identification. This method integrated biometric features such as facial recognition, voice recognition, and fingerprint recognition. The research results showed that compared with K-nearest neighbor classifiers and recent methods, the research method had superiority in recognition rate and equal error rate [16]. Following the above literature summary, Table 1 is compiled.

From the current research status of scholars at home and abroad, it can be seen that the efficiency and practicality of HMCC in the field of smart homes are relatively low. Therefore, the study introduces MF for HMCC in smart homes, hoping to improve the efficiency of intelligent collaborative control and enhance the convenience of using smart homes.

3 Design of a smart home control model that integrates multimodal and collaborative control

3.1 Construction of control model for smart home human machine system based on iterative control algorithm

With the continuous improvement of artificial intelligence technology, the technology applied to HMCC in smart homes has become increasingly mature. However, when using traditional control algorithms for smart home collaborative control, the control effect is poor. To address this issue, ILC algorithms are introduced for collaborative control of smart homes. ILC algorithm, as an algorithm that continuously executes the same instructions and steps, and controls through iteration of new and old variables, can make the system output approximate the ideal output. It only requires less computation and parameters for system control. Iterative control has both open-loop control $u^{O}(t,k)$ and closed-loop control $u^{C}(t,k)$, and its mathematical expression is shown in equation (1).

Table 1: Summary of literature results.

Author	Research contents	Research findings	Limitation
Dong et al.	Design a multimodal perception processing module and construct a cross array of memristors	System occupancy rate of 1.25%, supporting multimodal sensing	Dependent on dedicated hardware (memristors), low deployment flexibility
Wei et al.	Train deep reinforcement learning algorithms using approximate strategy optimization methods and design strategy networks to generate device actions	Dynamically adjust equipment energy consumption	High training computation cost, control delay (200-500ms)
Ameer et al.	Build a hybrid model by combining the family centered method and attribute centered method	Support fine-grained permission management	Not involving actual control performance, only focusing on the safety aspect
Perumal et al.	By tracking activity data in smart home environments, real-time data is transmitted to IoT edge servers	Recognition accuracy 82.3%, delay 100-300ms	Only supports sensor modes, with a single interaction method
Chen et al.	Building multimodal fusion modules and advanced semantic modules	Fusion accuracy 91.32%	High computational complexity, unverified applicability in smart homes
Zhou et al.	Constructing a modal specific graph neural network architecture and utilizing reconstruction networks to determine individual modal fidelity	The accuracy rate of depression detection is 88.6%	Relying on high-dimensional physiological signals makes it difficult to migrate to smart home scenarios
Lu et al.	Using end-to-end methods for network training to extract features from modal data	Defect detection accuracy rate 94.32%	Industrial scenario specific, real-time performance not optimized
Fang et al.	Using Long Short Term Memory Networks to Learn Audio and Visual Features, Multimodal Fusion for Feature Transfer	RMSE=0.468	High computational cost, delay not reported
Dalila C et al.	Integrate multiple biometric features such as facial recognition, speech recognition, and fingerprint recognition	Recognition rate better than KNN	Not involving control tasks, only applicable to identity verification scenarios

$$\begin{aligned} u^{O}(t,k) &= u(t,k-1)+L(k)e(t,k-1) \\ u^{C}(t,k) &= u(t,k-1)+L(k)e(t,k) \end{aligned} \tag{1}$$

In equation (1), t and k respectively represent the time and number of control iterations, and $e(t,k)$ represents the iteration error. $u(t,k)$ and $L(k)$ respectively represent the control input and learning gain matrix at the k iteration. ILC is mainly used for system control by calculating trajectory updates in the time domain of the system, as expressed in equation (2).

$$u_{i+1} = L_u u_i + L_e e_i \tag{2}$$

In equation (2), u represents the vector of the timing signal within a certain period. L_u and L_e are the proportional gain matrix of the control input and the gain matrix of the error signal, respectively. u_i and e_i are the control input and tracking error at the i iteration. The principle structure of ILC is shown in Figure 1.

From Figure 1, iterative learning is used for control, with the system providing the desired input. When the controller receives the output signal, the ILC algorithm repeats the output iteration and uses the alternation of new and old data to eliminate errors. The input signal after repeated iterations is transmitted to the controlled object and output. The error correction of ILC algorithm can be expressed in mathematical form, as shown in equation (3).

$$u_{i+1} = L_u u_i + L_e e_i \tag{3}$$

In equation (3), K_D represents the learning gain of the parameter during iterative learning, and $\dot{e}_i$ represents the derivative of the difference between the true output value and the ideal output value of the system. The mathematical expression of the error iteration update rule of the ILC algorithm during error iteration update is shown in equation (4).

$$u_{k+1}\left[n\right] = u_k\left[n\right] + K_P e_k\left[n\right] + K_I \Delta t \sum_{m=0}^{n} e_k\left[m\right] \tag{4}$$

In equation (4), a fixed sampling interval Δt is used to discretize the time into $t = n \times \Delta t\left(n = 0,1,...,N\right)$; $u_k\left[n\right]$ corresponds to the control input of the n time step in the k iteration; $e_k\left[n\right]$ represents the error signal between the expected output and the actual output, while $\sum_{m=0}^{n} e_k\left[m\right]$ is the discrete approximation of the integral term. K_P and K_I are proportional gain and integral gain, respectively. ILC has a small control error when controlling smart home systems, but its collaborative control effect on smart homes is poor [17]. To compensate for the poor performance of ILC in system collaborative control, the WOA is introduced to improve the ILC algorithm. The

WOA, as a population-based intelligent optimization algorithm that mimics the hunting of humpback whale populations in nature, is easy to operate and less prone to falling into local optimal solution traps. The implementation process of the WOA is shown in Figure 2.

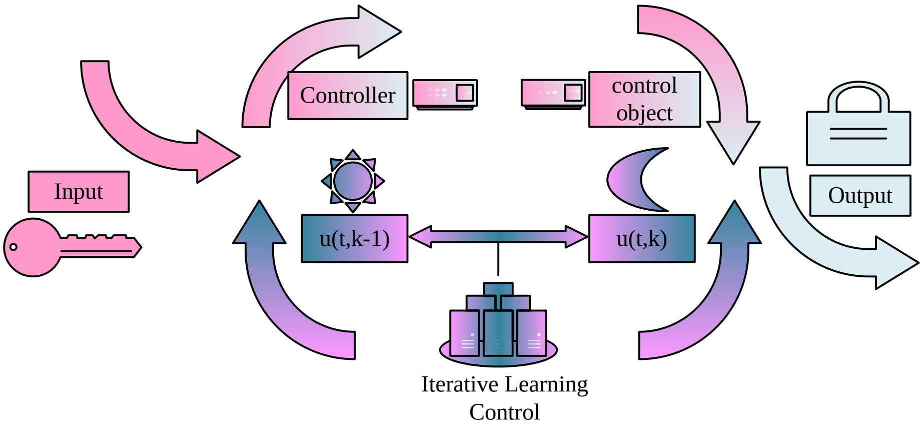


Figure 1: The basic principle and structure of ILC.

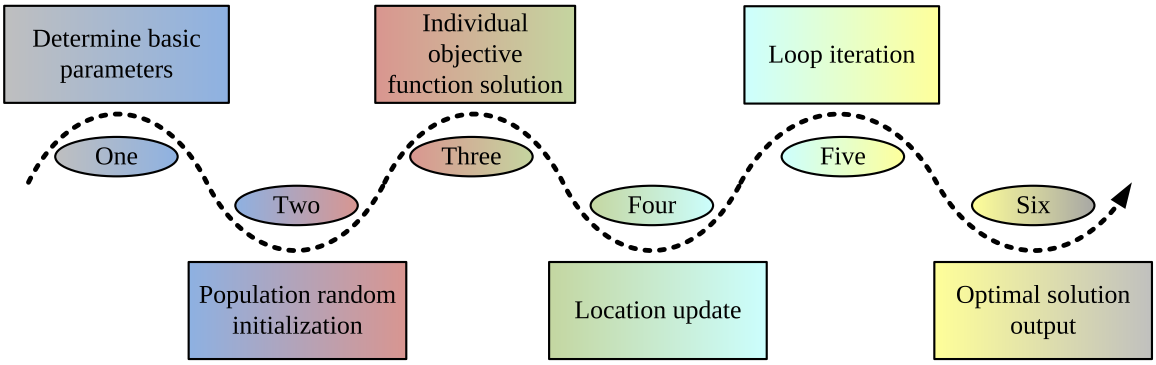

Determine basic parameters
Individual objective function solution
Loop iteration
One
Two
Three
Four
Five
Six
Population random initialization
Location update
Optimal solution output

Figure 2: Implementation process of WOA.

In Figure 2, the WOA first sets the initial input parameters and determines the population range and iteration times during optimization. Subsequently, the population is initialized and the objective function is solved for the independent individuals of the population. Finally, the population location is updated and the results are output. The WOA for population initialization can be expressed mathematically, as shown in equation (5).

$$X_{i,j}(initial) = Lb_j + rand_{i,j}(0,1) \cdot (Ub_j - Lb_j) \quad (5)$$

In equation (5), Lb means the lower bound position of the search space in the WOA, Ub means the upper bound position of the search space in the algorithm, and $rand_{i,j}(0,1)$ represents a random number with a value range of $rand_{i,j}(0,1) \in [0,1]$. The process of updating individual positions using the WOA can be expressed mathematically, as shown in equation (6).

$$X(t+1) = X(t) - \vec{A} \cdot \left| \vec{C} \cdot \vec{X}^*(t) - \vec{X}(t) \right| \quad (6)$$

In equation (6), $\vec{X}(t)$ means the vector of the position of the individual whale, $\vec{X}^*(t)$ means the position vector found by the leader whale of the whale population, and $\vec{A}$ and $\vec{C}$ denote the coefficient vectors. The WOA simulates the spiral movement of whales during hunting to complete the pursuit of prey, and its mathematical expression is shown in equation (7).

$$X(t+1) = D'e^{bl}\cos(2\pi l) + X^*(t) \quad (7)$$

In equation (7), D' means the distance between the individual whale and its prey, and b represents the constant coefficient of the whale's spiral motion. The ILC algorithm improved by the WOA can comprehensively control the system and effectively process the system control data. Therefore, the study utilized the WOA-ILC algorithm to construct a smart home human-machine control system and established a smart home human-machine control model. The specific structure is shown in Figure 3.

Figure 3 shows the specific implementation architecture of WOA-ILC in smart homes, whose core is to convert algorithm outputs into physical operation instructions for heterogeneous devices through a unified control bus. This architecture mainly consists of three parts: data acquisition module, controller based on WOA-ILC algorithm, and smart home devices. The data acquisition module is responsible for collecting various data in the smart home environment, such as indoor temperature, humidity, light intensity, etc., and sending the collected data to the WOA-ILC controller. The controller based on the WOA-ILC algorithm calculates the optimal control input and uses the optimized control signal to regulate the operating status of smart home devices.

Smart home devices adjust their operating status based on the output signal of the WOA-ILC controller to achieve control over the smart home environment. The controller is the core component of smart home devices, responsible for centralized management and coordination of various smart terminal devices, achieving automation and intelligent control of the home environment. Specifically, the integration mechanism of WOA and ILC is as follows: first, the parameters of the ILC algorithm and the WOA are initialized. Secondly, the WOA dynamically adjusts the learning gain in the ILC algorithm, namely K_P, K_I, and K_D, through its population search capability. In each iteration, the WOA updates the learning gain based on the current population position and objective function. Then the control input changes in the ILC algorithm are updated and finally the above process is repeated to guide the realization of a predetermined number of iterations or error convergence to a satisfactory value. Based on the above content, it can be seen that the design of the WOA-ILC algorithm mainly focuses on unimodal control performance, such as temperature control and humidity control. When voice commands and gesture commands are input simultaneously, the single error signal of this method cannot distinguish the modal source, resulting in control conflicts.

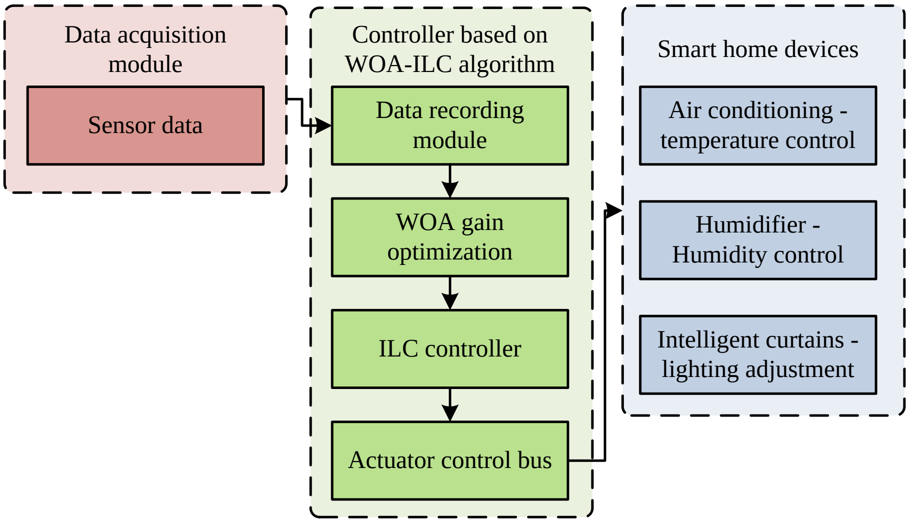

Data acquisition module
Sensor data
Controller based on WOA-ILC algorithm
Data recording module
WOA gain optimization
ILC controller
Actuator control bus
Smart home devices
Air conditioning - temperature control
Humidifier - Humidity control
Intelligent curtains - lighting adjustment

Figure 3: Smart home human-machine control system model.

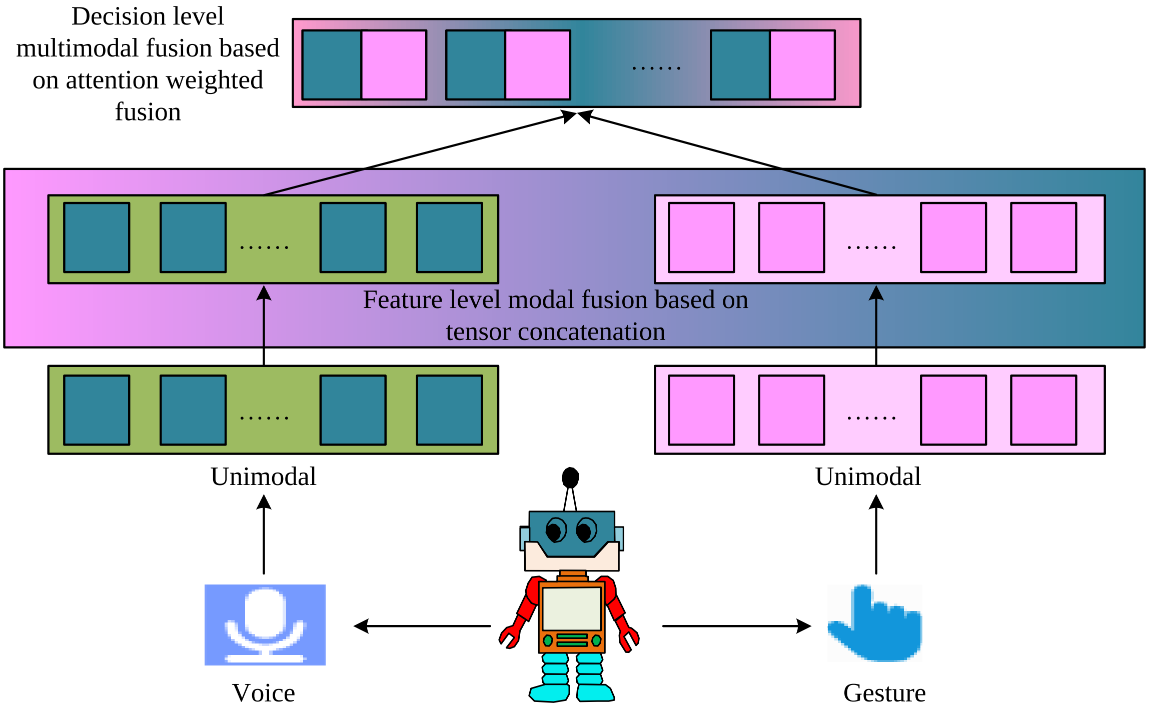

Decision level multimodal fusion based on attention weighted fusion
……
Feature level modal fusion based on tensor concatenation
……
……
Unimodal
Unimodal
Voice
Gesture

Figure 4: MF structure diagram.

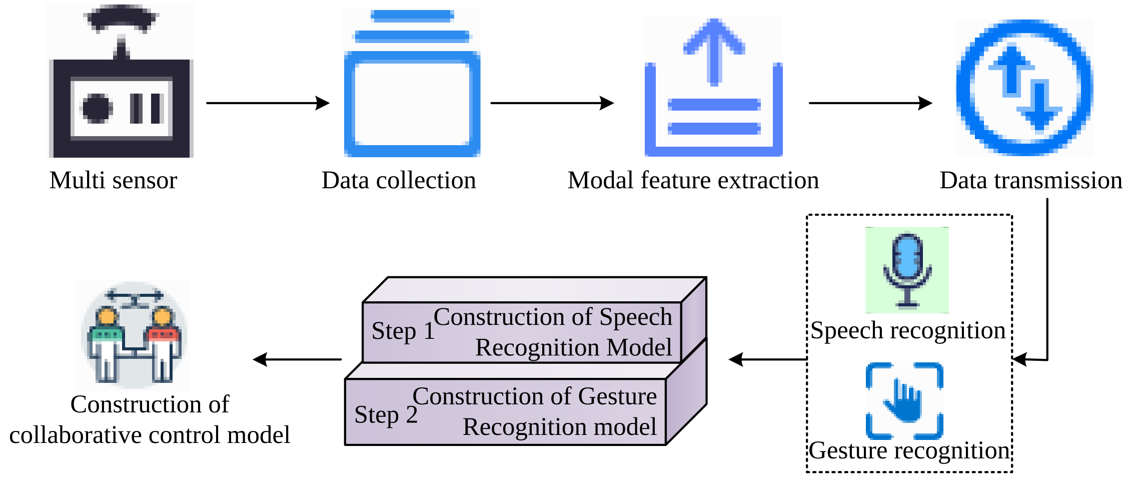

Multi sensor
Data collection
Modal feature extraction
Data transmission
Speech recognition
Gesture recognition
Step 1 Construction of Speech Recognition Model
Step 2 Construction of Gesture Recognition model
Construction of collaborative control model

Figure 5: Optimization process of multimodal smart home collaborative control model.

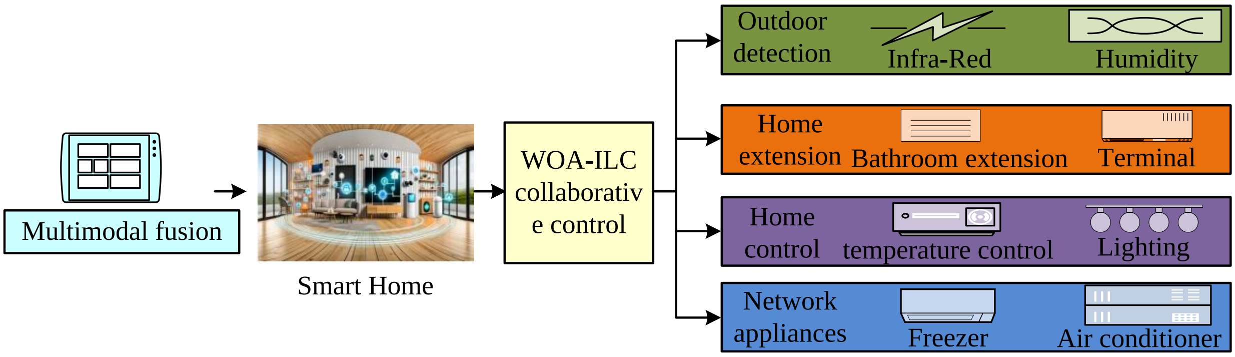


Figure 6: A human-machine collaborative control model combining MF.

3.2 Optimization of smart home collaborative control model combining MF

The smart home collaborative control model constructed using the WOA-ILC algorithm can ensure the intelligent control of the home system and is easy to operate. However, the control model constructed by the WOA-ILC algorithm has poor control effect on multiple modalities. Therefore, the study introduced MF to optimize the smart home collaborative control model. MF technology can fuse data through multiple sensors and eliminate the repulsion between different data modalities, thereby achieving data fusion. Its structure is shown in Figure 4.

In Figure 4, MF combines different single modalities and integrates data between them, mainly focusing on the fusion of features between speech and text data. After the feature extraction of speech and text data is completed, feature level fusion can be performed through tensor concatenation operation in equation (8), and then decision level MF can be performed using attention weighted fusion in equation (11). When performing MF, the weight tensor is used to linearly expand the modality, and the mathematical representation of vector generation is obtained using the input tensor, as shown in equation (8).

$$Z = \bigotimes_{m=1}^{M} z_m \tag{8}$$

In equation (8), Z represents the input tensor generated by MF. z_m represents the feature tensor of the m mode, and M represents the total number of modes. MF decomposes data weights into different modal factors, and its mathematical expression is shown in equation (9).

$$W = \sum_{k=1}^{R} \bigotimes_{m=1}^{M} (\phi_{m,k} \cdot W_{m,k}) \tag{9}$$

In equation (9), m represents the modal factor and R represents the minimum tensor rank. $\phi_{m,k}$ represents the contribution weight of the m modality in the k factor, and $W_{m,k}$ represents the k factor weight of the m modality. The optimization process of the smart home collaborative control model using multimodality is shown in Figure 5.

In Figure 5, MF technology mainly collects data through multiple data sensors and extracts features from different data modalities. After the feature extraction of different data modalities is completed, it can be processed through a collaborative control model, which can comprehensively process multiple inputs and output control decisions. In the smart home control model, data collection and transmission of speech recognition and gesture recognition are carried out through multiple sensors, and the data information of speech and gesture is preprocessed to reduce the loss of speech and gesture information, ensuring the integrity of speech and gesture information [18-20]. When collecting speech data information, a speech recognition model is first established to preprocess the language information to improve the decoding efficiency of the decoder for language information. For language data collection in speech information, the probability of word sequence occurrence is calculated, and the calculation of word generation probability can be expressed mathematically, as shown in equation (10).

$$p(W) = p(w_1^K) = \prod_{k=1}^{K} p(w_k \left| w_1^{k-1} \right.) \tag{10}$$

In equation (10), K means the total number of words contained in the recognized speech information, and w_k represents the k th word string in the number of recognized words. The construction of a collaborative control model for smart home human-machine system using WOA-ILC combined with MF is shown in Figure 6.

As shown in Figure 6, firstly, the collected multimodal data is sent to the feature extraction module. Secondly, key features are extracted and transmitted to the MF module. The fused feature vectors are used for subsequent control or decision-making. Based on the fused feature data, the optimal control input is calculated using the WOA-ILC algorithm, and the optimized control signal is sent to the actuator. Finally, the operating status of the smart home device is adjusted based on the output signal of the controller. The MF for gesture and speech recognition can be expressed mathematically, as shown in equation (11).

$$h = Attention\left(Z_{voice}, Z_{gesture}\right) = \sum_{i=1}^{N} \alpha_i \cdot \left(W_v z_i^v \| W_g z_i^g\right) \tag{11}$$

In equation (11), Z_{voice} and $Z_{gesture}$ are the feature sequences of the speech modality and gesture modality, W_v and W_g are the projection matrices of the corresponding modalities, α_i represents the attention weight of the time step, h is the fused feature representation, and $\|$ represents vector concatenation. The data feature extraction of voice information and gesture recognition information can be represented by mathematical expressions, as shown in equation (12).

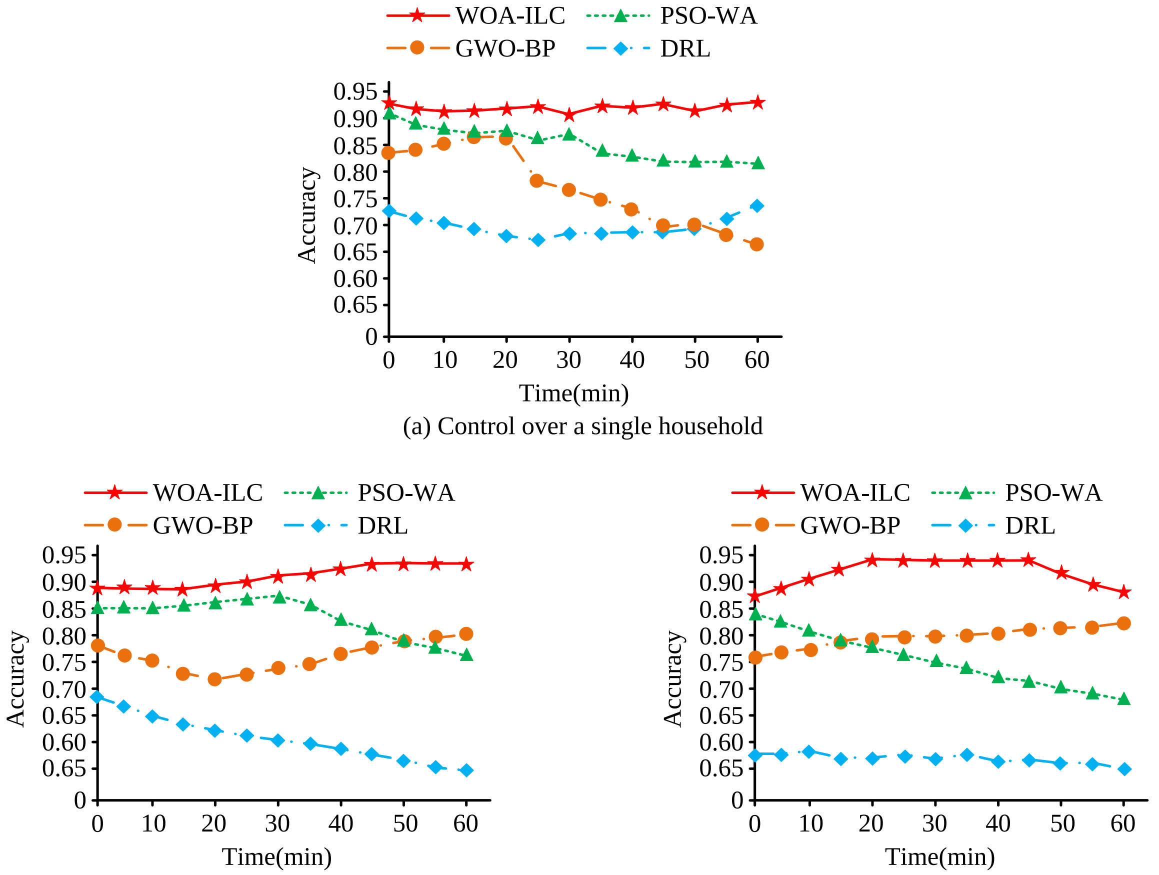
$$R_{ab}^{(n)} = \frac{\sum_{k=1}^{n} D_k^{(t)} S_k^{(t)} C_k^{(t)} Z_k^{(t)}}{\sum_{k=1}^{n} a_{rk}^{(1)} b_{rk}^{(1)} z_{rk}^{(1)}} \tag{12}$$

In equation (12), $R_{ab}^{(t)}$ represents the correlation between voice nodes and gesture nodes at different times, and n represents the number of intermediate nodes between gesture nodes and voice nodes. $D_k^{(t)}$, $S_k^{(t)}$, $C_k^{(t)}$, and $Z_k^{(t)}$ represent the data matrix, state matrix, context matrix, and feature vector of the k feature at time t. $a_{rk}^{(1)}$ and $b_{rk}^{(1)}$ are the weights of the r speech node and gesture node on the k feature, respectively. $z_{rk}^{(1)}$ represents the feature value of the r node on the k feature.

4 Empirical analysis of integrating multimodal smart home control models

4.1 Performance validation of improved iterative control model

To validate the performance of the improved iterative control model, model control experiments were conducted in Matlab software using Windows 11 system, Intel i5-12600KF processor, and NVIDIA GTX1070 graphics card model. Different types of smart homes, including smart air conditioners, smart refrigerators, and smart washing machines, were collected from the network, and the collected data was divided into training data and testing data. The voice commands were sourced from home environment recordings with a sample size of 8000, including background noise. The gesture trajectories were sourced from infrared depth cameras with a sample size of 6500, including 10 control gestures. The device status was sourced from smart home device logs with a sample size of 12000, featuring real-time recording of temperature, humidity, and power consumption. The dataset was divided into a training set and a testing set according to 7:3. The collection and use of personal data complied with privacy protection regulations and ethical requirements. The experimental parameters are set as follows: the population size, helix coefficient, and maximum iteration number of the WOA algorithm were set to 50, 1.0, and 200, respectively. The initial learning gain, differential gain, integral gain, and sampling interval in the ILC algorithm were 0.5, 0.05, 0.01, and 10ms, respectively. The backbone network of the speech recognition model adopted a two-layer bidirectional long short-term memory network and a 1-layer convolutional neural network, with the former having a hidden layer of 256 and the latter having 3 * 3, 64 channels. The training was 50 cycles (early stop), Batch=32, and the optimizer was AdamW. The architecture of the gesture recognition model is as follows: with a 1D convolutional neural network as the backbone network, Kernel=5, Stride=2, 64 → 128 → 256 channels, combined with the max pooling layer. The optimizer selected stochastic gradient descent. To verify the effectiveness of the WOA-ILC model, it was experimentally compared with the Particle Swarm Optimization-Genetic Algorithm (PSO-WA) model, the Grey Wolf Optimization-Back Propagation Neural Network (GWO-BP) model, and the Deep Reinforcement Learning (DRL) model. The DRL strategy network is a 3-layer fully connected network, and the value network adopts a symmetrical structure with the strategy network. The training cycle is 50000 steps (early stop strategy), and its end-to-end optimization characteristics are suitable as a performance upper limit reference. The input layer, hidden layer, and output layer of GWO-BP



(a) Control over a single household

(b) Control two types of home furnishings

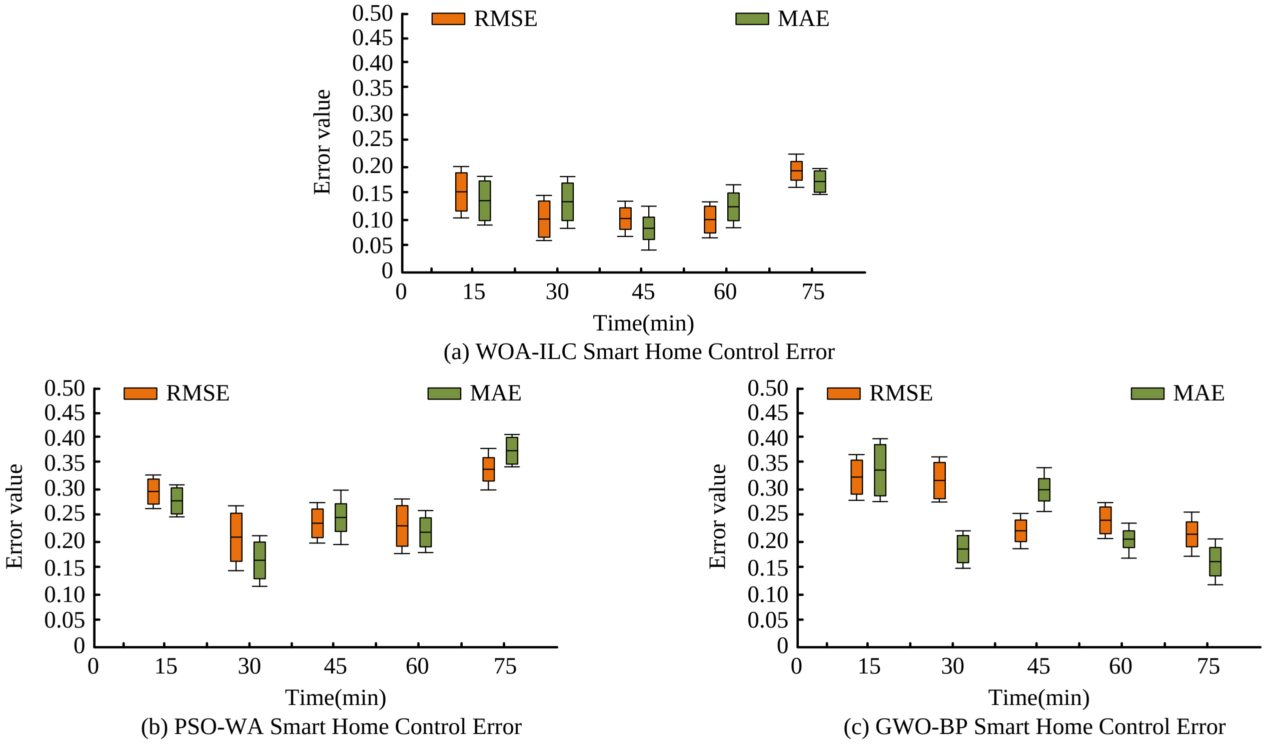
(c) Control three types of home furnishings

Figure 7: Comparison of control accuracy of different models.

were 8, 24, and 4, respectively. The population size and maximum iteration times were 50 and 200, respectively. Its interpretability and stability reflected the advantages and disadvantages of traditional methods. The crossover probability and mutation probability of the PSO-WA model were 0.85 and 0.02, respectively, making it suitable as a representative of metaheuristic algorithms. The comparison results are shown in Figure 7.

Figures 7 (a)-7 (c) correspond to the control accuracy tests of different models under single-device control, multi-device collaborative control, and noisy environments. According to Figure 7 (a), WOA-ILC had the highest control accuracy for smart homes, with an average control accuracy of 0.9212 and a maximum control accuracy of 0.9387. Moreover, the control accuracy for smart homes was not affected by changes in time span. The highest control accuracy of the PSO-WA model reached 0.8751, which was 0.0636 lower than the WOA-ILC model. From Figure 7 (b), the WOA-ILC model continuously improved its control accuracy for smart homes over time, with the highest control accuracy reaching 0.9412 and the average control accuracy being 0.9053. For the GWO-BP model, it could not perform well in controlling smart homes at the beginning, and over time, the control accuracy of smart homes slowly increased. Compared with the WOA-ILC model, the highest control progress was reduced by 0.1045. From Figure 7 (c), the control accuracy of the WOA-ILC model slowly increased with time and maintained a certain balance. Its average control accuracy was 0.8976, which was 0.2742 higher than that of the DRL model. The above outcomes denoted that the WOA-ILC model had higher precision in smart home control compared to other control models, and could perform high-precision control of smart homes well. The above results may be due to the fact that the WOA can dynamically adjust the search range and direction during the iteration process, which can effectively avoid getting stuck in local optima. To verify the control accuracy error of the WOA-ILC model on smart homes, RMSE and Mean Absolute Error (MAE) were used as experimental indicators. Different models were compared using the same dataset, and the experimental outcomes are indicated in Figure 8.

Figures 8 (a)-8 (c) correspond to the control error results of the WOA-ILC model, PSO-WA model, and GWO-BP model. From Figure 8 (a), the highest MAE value of the WOA-ILC model was 0.167, and the highest RMSE was 0.196. At the beginning of smart home control, the error value generated by controlling the home was relatively small, and over time, the control error of the smart home gradually decreased. When the control time reached 45 minutes, the MAE value for smart home control was the lowest, which was 0.094, and the lowest RMSE was 0.112. According to Figure 8 (b), the highest MAE of the PSO-WA model was 0.373, and the highest RMSE was 0.338. Overall, as time goes

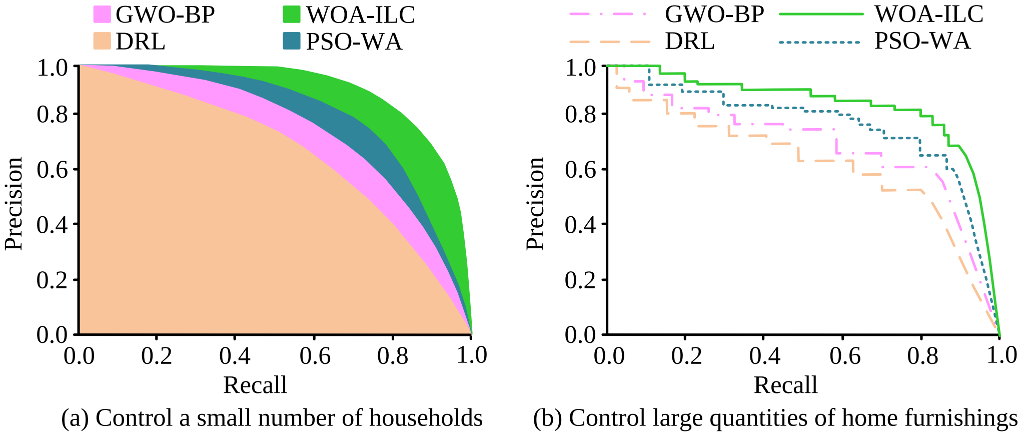


(a) WOA-ILC Smart Home Control Error

(b) PSO-WA Smart Home Control Error

(c) GWO-BP Smart Home Control Error

Figure 8: Comparison of control errors between different models.



(a) Control a small number of households

(b) Control large quantities of home furnishings

Figure 9: Comparison of PR curves for different control models.

on, the error value of the PSO-WA control model fluctuated to some extent and reached its minimum value at 30 minutes before increasing again. When the control time was 30 minutes, the RMSE reached a minimum of 0.216, which was 0.104 higher than the WOA-ILC model. The minimum MAE was 0.151, which was 0.057 higher than the WOA-ILC model. From Figure 8 (c), the control error of the GWO-BP model gradually decreased with time. The highest MAE was 0.337, which was 0.17 higher than the WOA-ILC model. The highest RMSE was 0.324, which was 0.128 higher than the WOA-ILC model. The above results indicated that the WOA-ILC model had a small control error for smart homes and performed well in controlling smart homes. This is because WOA can dynamically adjust the learning gain based on the current population position and objective function, so that it can better approximate the ideal output in each iteration. This dynamic adjustment mechanism significantly improves the convergence speed and control accuracy of the system. To further validate the control performance of the control model, an experimental comparison was conducted on its precision-recall (PR) curve, and the results are shown in Figure 9.

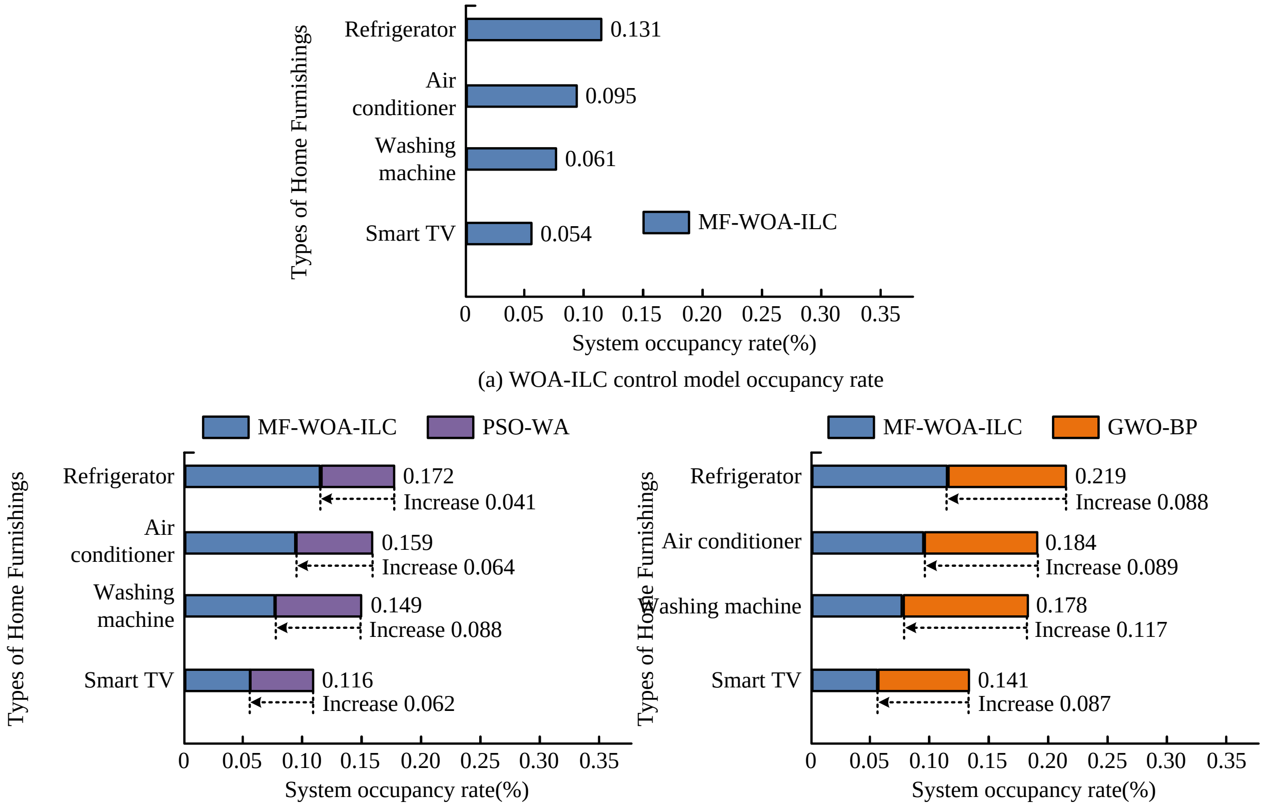
Comparison of PR curves for different methods in the cases of few-batch and multi-batch corresponds to Figure 9 (a) and Figure 9 (b). According to Figure 9 (a), when the WOA-ILC model was used to control small batch smart homes, the PR curve area was 0.9758, which was higher than the other three control models. The offline area of the PR curve of the PSO-WA control model was 0.9213, which was 0.0545 lower than that of the WOA-ILC model. Compared with the DRL model, the PR curve area of the WOA-ILC model increased by 0.2471. From Figure 9 (b), when performing multi-batch

Table 2: Statistical test results.

Comparison group	MF-WOA-ILC vs PSO-GA	MF-WOA-ILC vs DRL	MF-WOA-ILC vs GWO-BP
Experimental group mean	0.9387	0.9387	0.9387
Control group mean	0.8751	0.8234	0.8346
t-values	7.82	9.15	8.37
p-value	<0.001	<0.001	<0.001
Cohen's d	1.24	1.53	1.41

Table 3: Actual deployment of equipment.

Types of home furnishings	Number of data	Model	Brand
Smart refrigerator	50	XQS70-128	Meiling refrigerator
Air conditioning	50	KFR-50GW/N1A1	Xiaomi air conditioner
Washing machine	50	EG100MATESL6	Haier washing machine
Smart TV	50	KD-75X9000F	Sony TV



(a) WOA-ILC control model occupancy rate

(b) PSO-WA control model occupancy rate

(c) GWO-BP control model occupancy rate

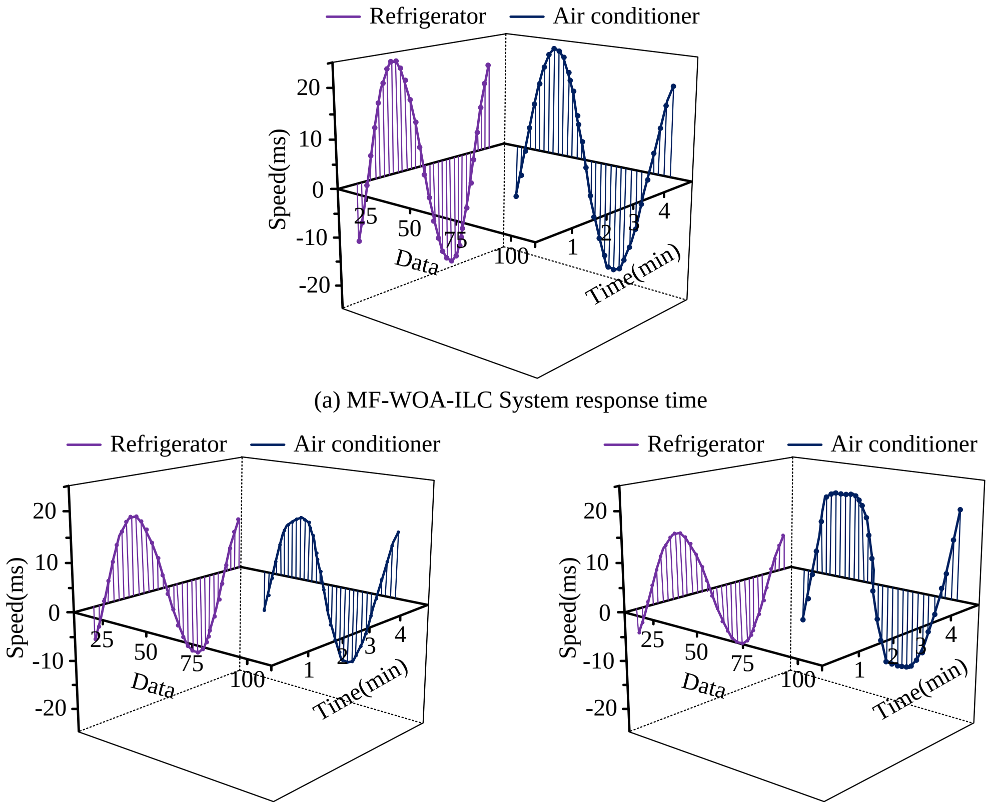
Figure 10: Comparison of system occupancy between different control models.

smart home control, the PR curve area of the WOA-ILC model decreased, but it was still significantly higher than the other three models at 0.9645, which was 0.0753 higher than the PSO-WA model and 0.1368 higher than the GWO-BP model. To further quantify the performance of the comparative methods, statistical tests were conducted, and the results are shown in Table 2.

According to Table 2, the experimental group showed a statistically significant advantage (p<0.001) in all comparisons, and an effect size exceeding 1.2 indicated that the differences in research methods were of practical significance. The F-value of the three groups' repeated cross validation was 46.37, (p<0.001).

4.2 Practical effect analysis of MF smart home collaborative control model

To analyze the practical effect of WOA-ILC smart home control combined with MF, different types of smart home control modes, including speech recognition and gesture recognition, were collected from different households for experimental verification. 200 sets of data were collected from the network. Due to the low complexity of research methods, there is no need for a large amount of training data to learn model parameters. In this case, less training data may already be sufficient, while more testing data can better evaluate the model's generalization ability. At the same time, to ensure that the testing set has sufficient sample size to evaluate the model's performance, 50 sets



(a) MF-WOA-ILC System response time

(b) PSO-WA System response time

(c) GWO-BP System response time

Figure 11: Comparison of response times of different models.

of data were used as the training set and 150 sets of data were used as the testing set. The actual deployed equipment is shown in Table 3.

To verify the system occupancy rate of the WOA-ILC control model combined with MF, different home control system occupancy rate experiments were compared with the PSO-WA model and GWO-BP model. The experimental results are shown in Figure 10.

According to Figure 10 (a), when the MF-WOA-ILC control model was used for intelligent refrigerator control, the system occupancy rate was 0.131%, the system occupancy rate for smart air conditioner control was 0.095%, the system occupancy rate for smart washing machine control was 0.061, and the system occupancy rate for smart air conditioner was 0.054. From Figure 10 (b), when the PSO-WA model was used for intelligent refrigerator control, the system occupancy rate reached 0.72%, which was 0.041% higher than the occupancy rate of the MF-WOA-ILC model. The system occupancy rate during smart TV control was 0.116, which was 0.088 higher than the MF-WOA-ILC model. The system occupancy rates of smart air conditioner and smart washing machine reached 0.159% and 0.149% respectively. From Figure 10 (c), the GWO-BP model achieved occupancy rates of 0.184%, 0.178%, 0.219%, and 0.141% for the four types of smart home control systems, respectively. Compared with the MF-WOA-ILC control model, the system occupancy rate for smart air conditioner control increased by 0.088%, and the system occupancy rate for smart washing machine control increased by 0.117. This is because MF technology can more comprehensively reflect the user's intention and environmental status by fusing data from different modalities, thereby improving the control accuracy of the system. Simultaneously, through tensor concatenation and attention weighted fusion, the features of speech and gesture modalities are effectively integrated. This can preserve more useful information, reduce information loss, and thus improve the quality of data representation. To verify the system response time of the MF-WOA-ILC HMCC model, it was experimentally compared with the PSO-WA model and GWO-BP model, and the results are shown in Figure 11.

In Figure 11 (a), the MF-WOA-ILC control model required less response time and had a fast response speed. Under 25 sets of data conditions, the average response speed was 10.31ms. As the data increased, the response time decreased and the response speed slowly accelerated. When the number of data reached 50, the average response speed reached a maximum of 24.12ms. The PSO-WA model necessitated a longer system reaction time than the suggested model, as seen in Figure 11(b). Its average response speed, when tested on 25 data sets, was 5.35 ms, 4.96 ms faster than the MF-WOA-ILC model. Its average response time, when 50 sets of data were included, was 18.89 ms, 5.23 ms faster than the MF-WOA-ILC model. The GWO-BP model had average response speeds of 4.48 ms and 16.21 ms, respectively, when there were 25 and 50 sets of data. This was 5.83 ms and 7.91 ms faster than the MF-WOA-ILC model, as shown in Figure 11 (c). To test the effectiveness of different parts of the MF-WOA-ILC model in practical

Table 4: The effectiveness of different parts of the MF-WOA-ILC model in practical applications.

Evaluation	Single voice	Single gesture	MF-WOA-ILC model
Inference time/ms	6.2	3.5	2.1
Memory usage/%	58.3	42.1	16.4
CPU utilization rate/%	12.7	8.9	5.2

applications, the study introduced inference time, memory usage, and CPU utilization for evaluation. The results are shown in Table 4.

Table 4 shows that the speech recognition module had the highest time consumption, at 6.2ms, mainly due to its bidirectional LSTM network needing to process temporal signals. The gesture recognition module has reduced its time consumption by 44%. After being processed by MF technology, the time consumption was only 2.1ms. In the fusion stage, WOA's population parallel computing was used to achieve an ultra-low occupancy of 5.2%. In addition, the research method could increase market share by 8% in terms of energy-saving benefits, maintenance costs, and annual single household benefits retained by users, which were 320-450%, 150-200%, and 12%-15%, respectively.

4.3 Discussion

The MF-WOA-ILC model was designed to address the challenges of HMCC in smart home environments. The experimental results showed that, firstly in terms of accuracy, the research model achieved high control accuracy in smart home control, with an average control accuracy of 0.9212 and a maximum control accuracy of 0.9387. Meanwhile, its maximum RMSE was 0.196 and maximum mean absolute error was 0.167, significantly better than the RMSE of 0.468 in reference [15]. This indicates that the research method can achieve more accurate HMCC when dealing with smart home control tasks in complex dynamic environments.

Secondly, in terms of controlling latency, the MF-WOA-ILC model had an average response time of 10.31ms under 25 data conditions and a response time of 24.12ms under 50 data conditions. This fast response capability enabled it to better adapt to tasks in smart home environments that require high real-time performance, such as real-time control and feedback of smart devices. Most existing literature has not quantitatively analyzed control delay.

In terms of system occupancy rate, the system occupancy rate in reference [8] was 1.25%. The MF-WOA-ILC model had significantly lower system occupancy rates in different smart home device controls, which were 0.131% for smart refrigerators, 0.095% for smart air conditioners, 0.061% for smart washing machines, and 0.054% for smart TVs. This indicates that the method proposed in this article has significant advantages in resource utilization efficiency, can effectively reduce system resource consumption, and is suitable for application on embedded devices with limited resources.

Finally, in terms of the scope of multimodal integration, the research results of references [12] - [16] focused on a single application scenario and did not involve the application scenario of collaborative control. However, the MF-WOA-ILC model could significantly improve the flexibility and adaptability of smart home control. In summary, the research methods have significant advantages in various aspects, providing new ideas and technical support for the research and application in this field.

5 Conclusion

Aiming at the problem that traditional smart home control methods cannot achieve good HMCC, a method was proposed to improve the construction of a smart home HMCC model by combining ILC with MF. ILC was improved to determine the input and output of the control system, and errors in home data were iteratively reduced, and finally MF was combined to fuse speech and gesture modalities for multimodal HMCC. The experimental results showed that the MF-WOA-ILC model had a maximum MAE of 0.167 and a maximum RMSE of 0.196 when performing smart home collaborative control, with a control accuracy of 0.9387 and an average control accuracy of 0.9212. In practical applications, the proposed control model had a system occupancy rate of 0.131%, demonstrating high accuracy and good control performance. In summary, the proposed model could effectively carry out HMCC of smart homes, with high control accuracy. However, there are still certain limitations in the research, as the research method only integrates speech and gesture modalities, which limits the system's adaptability in multiple scenarios. Therefore, in future research, facial expressions, EEG signals, eye tracking and other modalities can be introduced to enhance the robustness of the system.

References

[1] Guanglong Du, Linlin Zhang, Kang Su, Xueqian Wang, Shaohua Teng, and Peter X. Liu. A multimodal fusion fatigue driving detection method based on heart rate and PERCLOS. IEEE Transactions on Intelligent Transportation Systems, 23(11):21810-21820, 2022. https://doi.org/10.1109/TITS.2022.3176973

[2] Long Jin, Xin Zheng, and Xin Luo. Neural dynamics for distributed collaborative control of manipulators with time delays. IEEE/CAA Journal of Automatica Sinica, 9(5):854-863, 2022. https://doi.org/10.1109/JAS.2022.105446

[3] Yong Xu, and Zhengguang Wu. Data-based collaborative learning for multiagent systems under distributed denial-of-service attacks. IEEE Transactions on Cognitive and Developmental Systems, 16(1):75-85, 2022. https://doi.org/10.1109/TCDS.2022.3172937

[4] Genfeng Liu, and Zhongsheng Hou. Adaptive iterative learning fault-tolerant control for state constrained nonlinear systems with randomly varying iteration lengths. IEEE Transactions on Neural Networks and Learning Systems, 35(2):1735-1749, 2022. https://doi.org/10.1109/TNNLS.2022.3185080

[5] Ahmed S. Elkorany, Mohamed Marey, Khaled M. Almustafa, and Zeinab F. Elsharkawy. Breast cancer diagnosis using support vector machines optimized by whale optimization and dragonfly algorithms. IEEE Access, 10(3):69688-69699, 2022. https://doi.org/10.1109/ACCESS.2022.3186021

[6] Yu Liu, Yu Shi, Fuhao Mu, Juan Cheng, Chang Li, and Xun Chen. Multimodal MRI volumetric data fusion with convolutional neural networks. IEEE Transactions on Instrumentation and Measurement, 71:1-15, 2022. https://doi.org/10.1109/TIM.2022.3184360

[7] Mahlous A. Threat model and risk management for a smart home IoT system. Informatica, 47(1):51-63, 2023. https://doi.org/10.31449/inf.v47i1.4526

[8] Zhekang Dong, Xiaoyue Ji, Guangdong Zhou, Mingyu Gao, and Donglian Qi. Multimodal neuromorphic sensory-processing system with memristor circuits for smart home applications. IEEE Transactions on Industry Applications, 59(1):47-58, 2022. https://doi.org/10.1109/TIA.2022.3188749

[9] Guixi Wei, Ming Chi, Zhiwei Liu, Mingfeng Ge, Chaojie Li, and Xianggang Liu. Deep reinforcement learning for real-time energy management in smart home. IEEE Systems Journal, 17(2):2489-2499, 2023. https://doi.org/10.1109/JSYST.2023.3247592

[10] Safwa Ameer, James Benson, and Ravi Sandhu. Hybrid approaches (ABAC and RBAC) toward secure access control in smart home IoT. IEEE Transactions on Dependable and Secure Computing, 20(5):4032-4051, 2022. https://doi.org/10.1109/TDSC.2022.3216297

[11] Thinagaran Perumal, E. Ramanujam, Sukhavasi Suman, Abhishek Sharma, and Harshit Singhal. Internet of Things centric-based multiactivity recognition in smart home environment. IEEE Internet of Things Journal, 10(2):1724-1732, 2022. https://doi.org/10.1109/JIOT.2022.3209970

[12] Wujie Zhou, Shaohua Dong, Jingsheng Lei, and Lu Yu. MTANet: Multitask-aware network with hierarchical multimodal fusion for RGB-T urban scene understanding. IEEE Transactions on Intelligent Vehicles, 8(1):48-58, 2022. https://doi.org/10.1109/TIV.2022.3164899

[13] Tao Chen, Richang Hong, Yanrong Guo, Shijie Hao, and Bin Hu. MS^2-GNN: Exploring GNN-based multimodal fusion network for depression detection. IEEE Transactions on Cybernetics, 53(12):7749-7759, 2022. https://doi.org/10.1109/TCYB.2022.3197127

[14] Houhong Lu, Yangyang Zhu, Ming Yin, Guofu Yin, and Luofeng Xie. Multimodal fusion convolutional neural network with cross-attention mechanism for internal defect detection of magnetic tile. IEEE Access, 10(4):60876-60886, 2022. https://doi.org/10.1109/ACCESS.2022.3180725

[15] Fang M, Peng S, Liang Y, Hung C C, and Liu S. A multimodal fusion model with multi-level attention mechanism for depression detection. Biomedical Signal Processing and Control, 82(3):561-573, 2023. https://doi.org/10.1016/j.bspc.2022.104561

[16] Cherifi Dalila, El Affifi Omar Badis, Boushaba Saddek, and Amine Naït-ali. Feature Level Fusion of Face and voice Biometrics systems using Artificial Neural Network for personal recognition. Informatica, 44(1):85-96, 2020. https://doi.org/10.31449/inf.v44i1.2596

[17] Chi R, Lv Y, and Huang B. Distributed iterative learning temperature control for multi-zone HVAC system - ScienceDirect. Journal of the Franklin Institute, 357(2):810-831, 2020.

[18] Hsien-Pin Hsu, and Chia-Nan Wang. Hybridizing whale optimization algorithm with particle swarm optimization for scheduling a dual-command storage/retrieval machine. IEEE Access, 11(4):21264-21282, 2023. https://doi.org/10.1109/ACCESS.2023.3246518

[19] Ankita Gandhi, Kinjal Adhvaryu, Soujanya Poria, Erik Cambria, and Amir Hussain. Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. Information Fusion, 91(5):424-444, 2023. https://doi.org/10.1016/j.inffus.2022.09.025

[20] Can Cui, Haichun Yang, Yaohong Wang, Shilin Zhao, Zuhayr Asad, Lori A Coburn, Keith T Wilson, Bennett A Landman, and Yuankai Huo. Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: A review. Progress in Biomedical Engineering, 5(2):201-213, 2023. https://doi.org/10.1088/2516-1091/acc2fe