Neural Networks 164 (2023) 489–496

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

An adaptive reinforcement learning-based multimodal data fusion framework for human–robot confrontation gaming

Wen Qi [a,b], Haoyu Fan [a], Hamid Reza Karimi [c,*], Hang Su [d]

[a] School of Future Technology, South China University of Technology, Guangzhou, 511436, China
[b] Pazhou Lab, Guangzhou, 510330, China
[c] Department of Mechanical Engineering, Politecnico di Milano, 20156 Milan, Italy
[d] Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, 20133, Italy

ARTICLE INFO

Article history:
Received 16 December 2022
Received in revised form 23 February 2023
Accepted 26 April 2023
Available online 6 May 2023

Keywords:
Reinforcement learning
Multimodal data fusion
Human–robot confrontation
Adaptive learning
Multiple sensors fusion
Hand Gesture Recognition

ABSTRACT

Playing games between humans and robots have become a widespread human–robot confrontation (HRC) application. Although many approaches were proposed to enhance the tracking accuracy by combining different information, the problems of the intelligence degree of the robot and the anti-interference ability of the motion capture system still need to be solved. In this paper, we present an adaptive reinforcement learning (RL) based multimodal data fusion (AdaRL-MDF) framework teaching the robot hand to play Rock–Paper–Scissors (RPS) game with humans. It includes an adaptive learning mechanism to update the ensemble classifier, an RL model providing intellectual wisdom to the robot, and a multimodal data fusion structure offering resistance to interference. The corresponding experiments prove the mentioned functions of the AdaRL-MDF model. The comparison accuracy and computational time show the high performance of the ensemble model by combining k-nearest neighbor (k-NN) and deep convolutional neural network (DCNN). In addition, the depth vision-based k-NN classifier obtains a 100% identification accuracy so that the predicted gestures can be regarded as the real value. The demonstration illustrates the real possibility of HRC application. The theory involved in this model provides the possibility of developing HRC intelligence.

1. Introduction

Over the last decade, Human–Robot Interaction (HRI) has become more prevalent in scientific research and applications, such as collaborative robots, human-like robots, and entertainment robots (Qiao, Zhong, Chen, & Wang, 2022; Qureshi, Nakamura, Yoshikawa, & Ishiguro, 2018). Especially, achieving HRI using hand gestures has gotten much attention due to its various applications, such as remote control, helping people with hearing impairments, and pick-and-place in the factory, where emotions and body postures cannot replace (Fiorini et al., 2021). Different Hand Gesture Recognition (HGR) sensors, such as Red, Green, and Blue (RGB) cameras, Leap Motion controllers (LMC), mechanic gloves, surface EMG (sEMG) sensors, and radars, have been popularly used and successfully applied in many research fields (Park et al., 2021). For example, HGR was used for sign language translation based on edge detection and cross-correlation, which implemented an automated translation system (Joshi, Sierra, & Arzuaga, 2017). For amputee patients, HGR can be utilized for exoskeleton controls to assist with daily living (Jiang, Kang, Song, Lo, & Shull, 2021). Recently, multi-model sensors have been utilized in HGR to improve recognizing accuracy. Meanwhile, more intelligent algorithms are proposed to solve complex HRI problems, such as reinforcement learning (RL).

However, most recent research needs to meet the requirements of HRI, which needs real-time interaction (Qiao, Chen, & Huang, 2021). They typically have problems with frequent interruptions and long response delays that lead to lousy performance (Skantze, 2021). To satisfy different users and robots in different environments, multi-modal sensors are extensively used (Matsufuji, Sato-Shimokawara, Yamaguchi, & Chen, 2019) in An Augmented Reality HRI multimodel system using a leap motion sensor controller to track the movement of the operator's hands and using a Kinect V2 camera to measure the corresponding motion velocities in 3D directions (Nishimura, Nakamura, & Ishiguro, 2020). Finally, Unreal Engine 4 creates an AR environment for the user to monitor the control process (Li, Fahmy, & Sienz, 2019). In addition, to make the robots more autonomous

* Corresponding author.
E-mail addresses: wenqi@scut.edu.cn (W. Qi), 201930360275@mail.scut.edu.cn (H. Fan), hamidreza.karimi@polimi.it (H.R. Karimi), hang.su@polimi.it (H. Su).
URLs: https://scholar.google.com/citations?user=ORBBQcAAAAAJ&hl=en (W. Qi), https://www.mecc.polimi.it/ricerca/personale-docente/prof-hamid-reza-karimi (H.R. Karimi).

https://doi.org/10.1016/j.neunet.2023.04.043

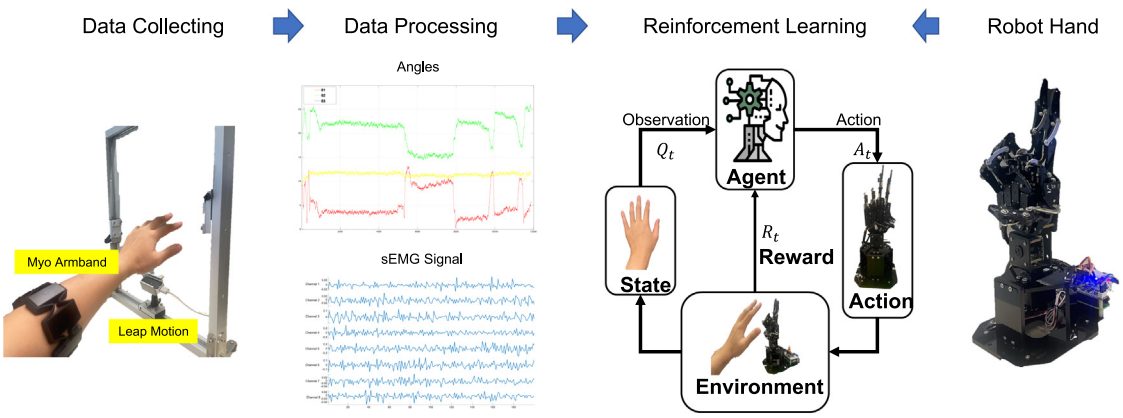

Fig. 1. The schematic diagram of multimodal data fusion based rock–paper–scissors game for human–robot interaction using reinforcement learning approach.

and innovative, unsupervised learning strategies like RL gradually replace traditional machine learning (Zhang et al., 2022). Although this field is explored at high speed, there are only some typical HGR systems to ensure stability and universality (Ito, Sueishi, Yamakawa, & Ishikawa, 2016).

In our previous work, we proposed different HGR and HRI architectures by designing multimodal data fusion (Qi, Wang, Su, & Aliverti, 2022), multiple sensors fusion-based data collection platform (Qi, Ovur, Li, Marzullo, & Song, 2021; Su et al., 2022), control theory-driven remote robot arm motions (Qi & Su, 2022; Su, Qi, Chen, & Zhang, 2022). Although our previous contributions had solved some mentioned problems, the labeling burden, low identification accuracy, and intelligence level of the HRI system still affect human–robot confrontation (HRC) research.

This paper demonstrates an adaptive RL-based multimodal data fusion framework (AdaRL-MDF) to boost identification accuracy and teach the robot to play games, as shown in Fig. 1. It includes an offline training module to capture multimodal sensors fusion, an updating module to retrain the old ensemble classifier, and an RL model for teaching the robot hand playing the Rock–Paper–Scissors (RPS) game with humans. We collect depth vision data and sEMG signals to build the ensemble classifier; the experimental results prove an excellent performance of the presented AdaRL-MDF model. The following statements highlight the main contributions of this paper.

1. It provides a new multimodal fusion-based ensemble classifier combining depth data and sEMG signals.
2. A RL model is built to teach the robot to play the RPS game with humans.
3. A hierarchical trigger mechanism is designed to control the updated procedure.

The rest of the paper is organized into the following six sections. The state-of-the-art HRI and HGR techniques are discussed in Section 2. Section 3 explains the proposed AdaRL-MDF framework by illustrating each module. The designed hardware system is shown in Section 4. Section 5 prove the excellent performance of the AdaRL-MDF modal by the designed experiments and results. The last section conducts a conclusion and also explains our future work.

2. Related works

The HRI application scenarios have become broader and more varied as the research goes into great depth (Onnasch & Roesler, 2021)—many methods to achieve HRI, such as emotion, voice, and hand gestures. Hand gesture was a meaningful way for humans to convey information and express intuitive intention, which had a high degree of differentiation, efficient information transmission, and robust flexibility (Wang et al., 2022). These benefits made HGR a research hotspot in HRI (Guo, Lu, & Yao, 2021). A dynamic HGR framework was proposed to improve the performance of each model by using a multimodal training or unimodal testing scheme, where the fused modalities consist of RGB and Optical flow (Abavisani, Joze, & Patel, 2019).

A multi-features sensor device was designed to obtain capacitance values on fingers and processed based on Error Correction Output Code Support Vector Machines (ECOC-SVM) and k-Nearest Neighbor (k-NN) (Wong, Juwono, & Khoo, 2021). In addition, a novel Few-Shot learning-Hand Gesture Recognition (FS-HGR) architecture using an sEMG sensor could work on a few training observations with less training time and mitigate the variability of sEMG signals (Rahimian et al., 2021). In addition, the optimal control problems restrict the performance of general nonlinear systems, so it needs to build a novel stability analysis for tracking control (Wang, Ha, & Zhao, 2022). Although many multiple sensors system and novel methods were proposed to achieve high recognition accuracy, the following three drawbacks remain the setback for the performance.

1. Time-consuming: Most models used supervised learning algorithms and public datasets. But in real applications, collecting, labeling, and training data would take a lot of time.
2. Limited categories: These classification methods only could classify finite gestures.
3. Sensor restriction: Single type of sensor always had limitations. Wearable sensors were unstable and must keep worn, such as capacitance glove (Wong et al., 2021) and Myo band Rahimian et al. (2021). Besides, vision sensors were easily influenced by the environment (Abavisani et al., 2019).

The RPS game was a typical HRI task achieved by hand gestures. It requested a combination between high-speed active hand tracking and fast sign recognition in a dynamic environment (Ito et al., 2016). A lot of work about HRI using RPS games has been done. An RPS interaction experiment used a machine vision sensor to recognize hand gestures by threshold (Yoon & Chi, 2006). However, the accuracy would become unstable if the environment changed, such as the camera angle and light. Moreover, it mainly focused on classification and did not consider the specific playing situation and rule.

To perfect this experiment, a better vision classification algorithm, HSV color space, was used for hand motion recognition (Ahn, Sa, Lee, & Choi, 2011). Using a four-fingered robot hand in this system, the game's setup was exhaustively considered.

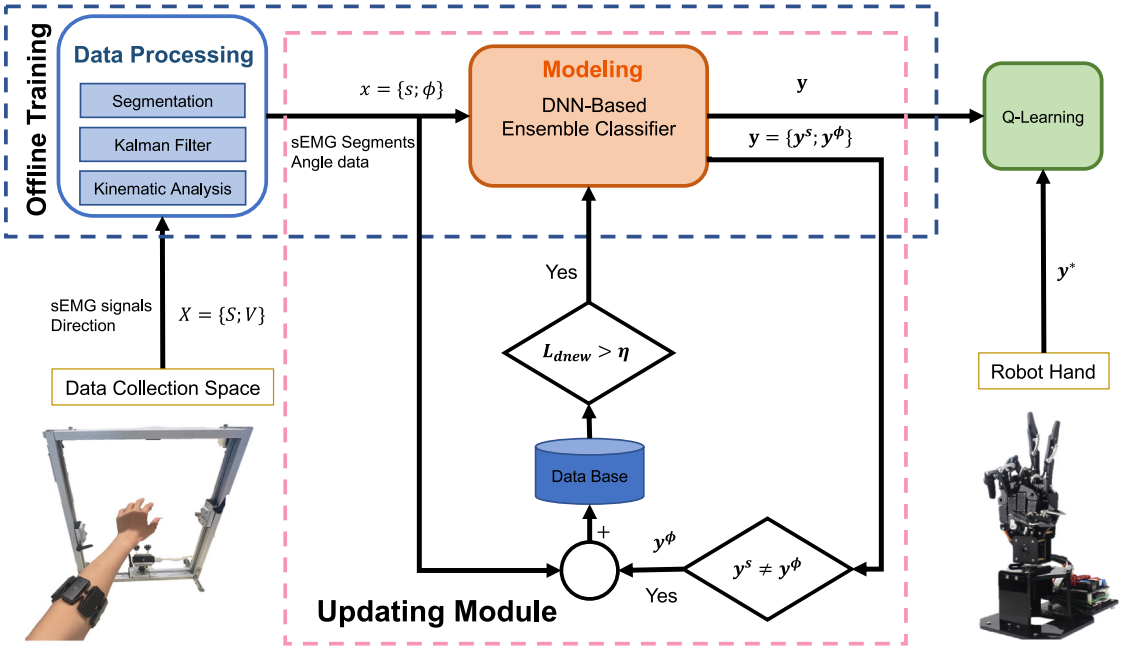

Fig. 2. The adaptive RL-based multimodal data fusion (AdaRL-MDF) model pipeline.

Nevertheless, the limitation of the camera still existed (Chen, Zhang, Karimi, Deng, & Yin, 2022). One of the methods that could improve the stability of the vision-based recognition method was using light and mirrors to reflect hand motion signals to the camera so that it could be suitable to acquire non-blurry images, which achieves hand gesture classification in a high speed (Ito et al., 2016). The algorithm was based on the length of fingertips. However, misclassification would occur if the light was influenced in the room. The Leap Motion devices were used to build the platform, which performs better than the usual camera (Ahmadi, Pour, Siamy, Taheri, & Meghdari, 2019). There were two playing strategies to compare the results of prediction. The recognition results could reach 93%. But in this research, the game strategy needed to find and generalize more participants to make strong claims.

Brock, Chulani, Merino, Szapiro, and Gomez (2020) presented an RPS interaction system with a social robot using leap motion in the recognition system. Machine learning was used to train multiple dimensions finger joint data from LMCs. This experiment was run in a high-speed response environment, and the results showed that popularly used classifiers perform well in shape recognition but poorly in movement segmentation. Instead of concentrating on different collection environments, an experiment mainly focused on implementing upgraded Markov chain (Jiang, Wu, & Karimi, 2022) mode and artificial NN on human behavior patterns during RPS game (Wang, Huang, Li, Evans, & He, 2020).

3. Methodology

Fig. 2 illustrates the whole structure of the proposed AdaRL-MDF framework, namely offline training, updating module, and RL learning model. The former aims to build an ensemble classifier by combining the collected dataset. The updating module updates the previous classifier when a change is triggered. The RL mechanism seeks to give the robot hand wisdom. The raw data is processed by Kalman filter, kinematic analysis, and segmentation steps to eliminate noises and improve stability. We also adopt both Sarsa and Q-learning to build the RL model to evaluate its feasibility. The following sections have described the details of each module.

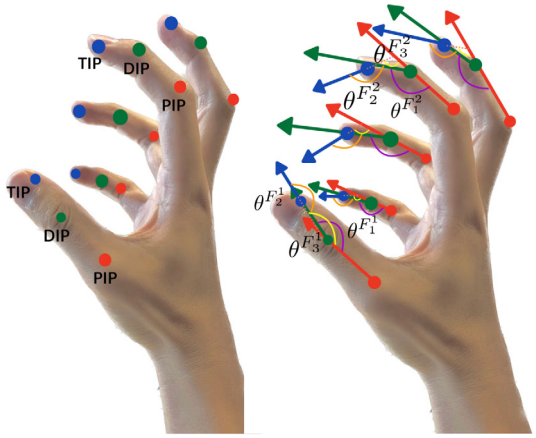

Fig. 3. The positions and angles of finger joints.

3.1. Data preprocessing

The multimodal data collection module uses LMC and Myo Armband to capture depth vision data and sEMG signals. Fig. 3 shows the position and angles of five finger joints, where DIP is the distal interphalangeal point, PIP is the proximal interphalangeal point, and TIP is the interphalangeal tip point. They are computed based on the Kalman filter and angle calculation.

The Kalman filter aims to estimate $\theta \in \mathbb{R}$, minimizing the squared error. The θ is the state that can be governed by Kong, Payne, Council, and Johnson (2021)

$$\theta_t = A_t\theta_{t-1} + B_t u_t + \omega_{t-1}. \tag{1}$$

The matrices A_t and B_t are related to the angle θ at time $t-1$ and the optional input $u \in \mathbb{R}$ at time t, respectively. ω_{t-1} means the process noise. When adding a driving function or process noise, the current angle θ can be calculated.

The measurement angle $z \in \mathbb{R}$ is $z_t = H_t\theta_t + v_t$ at time t. Where H represents the observation matrix. The independent

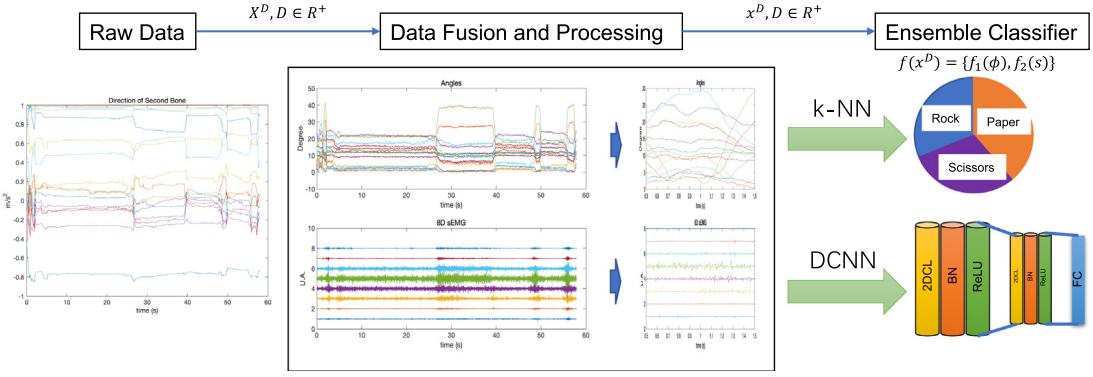

Fig. 4. The data stream to establish the ensemble classifier combining with k-NN and DCNN model.

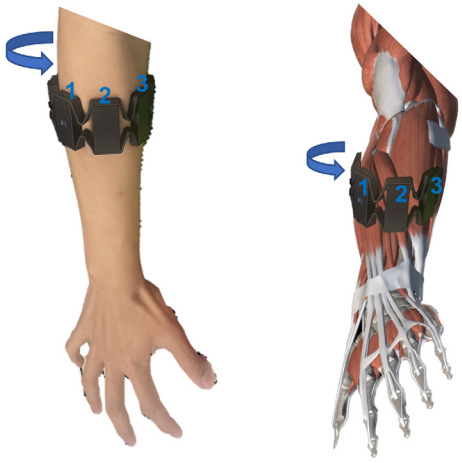

Fig. 5. Baseline setting on the forearm muscles by wearing the Myo Armband.

noise v_t is a random variable in measurement because it is a normal probability distribution as

$$p(v) \sim (0, \mathbb{R}). \tag{2}$$

Hence, the Kalman filter is summarized into time update

$$\hat{\theta}_t^- = A\hat{\theta}_{t-1} + Bu_{t-1}$$

$$P_t^- = AP_{t-1}A^T + Q, \tag{3}$$

and measurement update steps

$$K = P_t^- H^T (HP_t^- H^T + R)^{-1}$$

$$\hat{\theta}_t = \hat{\theta}_t^- + K(\hat{z}_t - H\hat{\theta}_t^-)$$

$$P_t = (I - KH)P_t^- \tag{4}$$

where P means the error covariance and Q is variance of processing noisy. In the second equations group, K notes the Kalman gain matrix, and $\hat{z}_t$ are the real (noisy) measurements at each time step t. The $\hat{\theta}_t^-$ notes the "super minus", a priori state estimate at time t.

By collecting finger joints direction data, θ of adjacent joint on each finger can be calculated by Fu et al. (2022):

$$\theta^{F_j^i} = \cos^{-1} \frac{\vec{V}_1 \cdot \vec{V}_2}{|\vec{V}_1||\vec{V}_2|} \tag{5}$$

where $\vec{V}_{i(i=1,2)}$ is the adjacent direction vector of the finger joint, see Fig. 3. Finally, the raw data $X = \{S; V\}$ has been processed by Kalman filter, kinematic analysis, and segmentation steps. The segments $x^* = \{s; \phi\}$ are regarded as the inputs for ensemble classifier building.

3.2. Offline modeling

Fig. 4 describes the whole procedure of offline modeling. It aims to establish the gesture recognition model based on the collected datasets. To enhance the classification accuracy, we build the designed DCNN model consisting of two DCNN-based modules.

The acquired inputs $x^* = \{s; \phi\}$ is used to build the ensemble classifier $f(x^*, \Theta) = \{f(s, \Theta_s); f(\phi, \Theta_\phi)\}$ (shown in Fig. 4). We adopt the k-NN method and a designed DCNN model to train the ensemble classifier. The former can save computational time by analyzing each finger's angle, while the latter aims to identify the gestures using sEMG signals.

The 8D sEMG signals $s \in R^8$ are captured by wearing Myo Armband, which can accurately locate each muscle on the forearm. The baseline setting and anatomical details on the forearm muscles are illustrated in Figs. 5 and 6. They capture the primary activities of surface muscles by using the device.

3.3. Updating module

The updating module is to improve the recognition accuracy by modifying some parameters of the DCNN classifiers. Hence, The updating module includes two threshold-based discriminant modules. In Algorithm 1, the first step is to find the predicted abnormal gestures by comparing the y_t^ϕ and y_t^s. Because we regard the y_t^ϕ as the real gestures, the inconsistent results will be selected to save into the training dataset. The second step tries to retrain the DCNN model $\hat{f}(x^s)$ and combine the previous angle-based classifier with building the new ensemble classifier $\hat{f}(x^D) = \{\hat{f}(x^s); \hat{f}(x^\phi)\}$.

To achieve the updating mechanism, it needs to build two steps hierarchical trigger modules. The former is to select the predicted results which are not equal to the real gesture, while the latter tries to set up a threshold η to verify if the length of the new dataset is more extended than to be saved into the training dataset. In detail, the classifier will be updated when there are too many categories of errors.

3.4. Reinforcement learning model

An RL model is adopted to teach the robot hand-playing RPS gaming with humans. In this paper, we adopt both Q-learning and Sarsa to train the Q matrix, which can be regarded as the intelligent brain of the robot hand. It aims to build a learning map

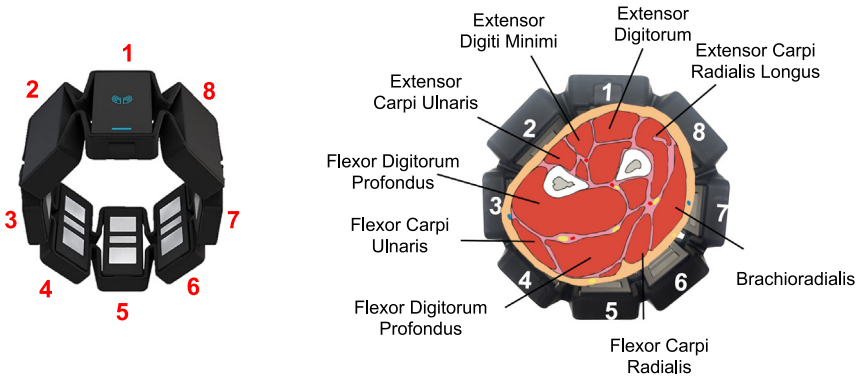

Fig. 6. Anatomical details of forearm muscles.

Algorithm 1: Updating Module.

Input: Predicted gestures sequence $y_t = \{y_t^{\phi}; y_t^{s}\}$; old model $f(x^D)$; threshold η.
Output: Updated model $\hat{f}(x^D)$.

```
if y_t^φ ≠ y_t^s then
    Update the training dataset {(s_old, y_old^s); (s_new, y_new^s)};
    Compute the length of new data L_dnew = length(y_new^s)
    if L_dnew > η then
        only retrain the DCNN model f̂(x^s);
        Combine the new ensemble classifier
        f̂(x^D) = {f̂(x^s); f(x^φ)}
    end
end
```

and a training mechanism to teach the hand robot to play games with humans. The learning map can be regarded as a recognition classifier, while the training procedure is to build the classifier. The RL technology provides a method to avoid the problem of insufficient training caused by fewer datasets. It can use fewer inputs to establish the HGR model.

The Q-learning is an off-policy and value-based control method in reinforcement learning that uses Bellman optimal equations and the ε-greed policy to update the action selection, which separates learning policy from the deferral policy (Jang, Kim, Harerimana, & Kim, 2019). The specific updating processing is as follow (Spano et al., 2019):

$$Q_{new}(s_t, a_t) = (1-\alpha)Q(s_t, a_t) + \alpha(r_t + \gamma max_a Q(s_{t+1}, a)) \quad (6)$$

The step size α, ε in ε-greed policy and reward matrix R should be confirmed before updating. It means that the newest knowledge can replace the older. Meanwhile, the long-run rewards have a more prominent role than the immediate rewards. In addition, the agent chooses the state s and action a at time t and $t+1$.

The Sarsa approach is another RL method, which means state–action–reward–next state–next action. It has the same four attributes: states S, actions A, discount γ, and step size α. However, the difference between the Sarsa method by comparing with Q-learning is that the Sarsa method adopts the current step and policy to compute the next step as (Mohan, Sharma, & Narayan, 2021):

$$Q_{new}(s_t, a_t) = (1-\alpha)Q(s_t, a_t) + \alpha(r_t + \gamma Q(s_{t+1}, a_{t+1})) \quad (7)$$

The next state and action (s_{t+1} and a_{t+1}) are selected by ε-greedy.

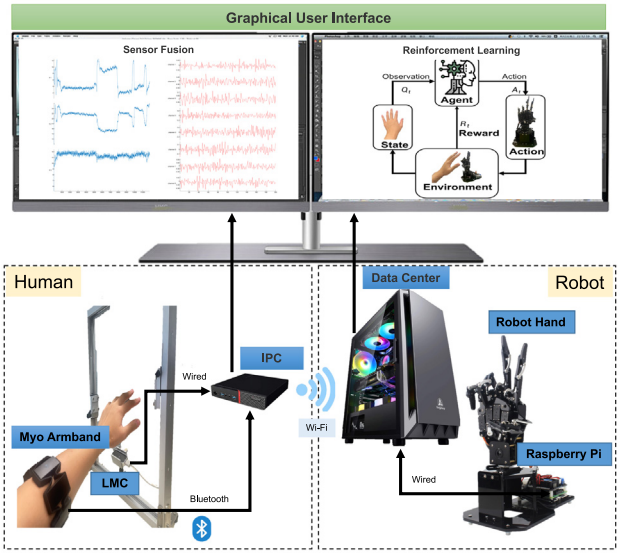

Fig. 7. Overview of the designed human–robot confrontation system.

According to their different updating ways, it is faster for Sarsa to get ideal performance than Q-Learning, but the Sarsa algorithm is easily stuck in the local minimum (Wang, Li, & Lin, 2013).

4. System description

Fig. 7 displays the designed hardware system consisting of data collection and robot interaction modules. The former aims to collect 8D sEMG signals and depth vision simultaneously, while the latter makes the robot hand play a game with the human. The multimodal data collection module aims to capture both sEMG and direction data from Myo Armband and Leap Motion Controller. The robot hand response module is to play the game with a human when a gesture is triggered. To achieve this goal, the whole platform is assembled as follows.

Data Collection Module: Leap Motion Controller (Leap Motion Inc, California, United States) capture depth data of fingers and connect to the host computer by using a USB cable. Myo Armband (Thalmic Labs, Kitchener, Canada) collect eight channels of sEMG signal of muscles of the forearm with Bluetooth. 200 Hz Synchronous data transfer and processing are ensured by means of timestamps at MATLAB on an IPC with a 15-6500T(2.5 GHz) processor and 8 GB of RAM.

Robot Interaction Module: a data processing center is used to identify hand gestures by connecting the IPC by Wi-Fi. It

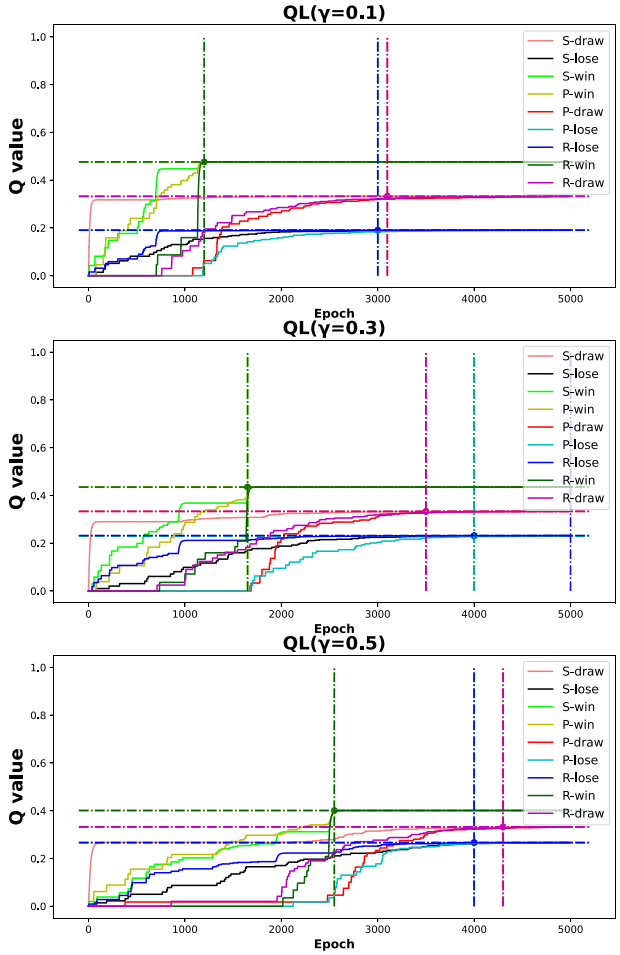

Fig. 8. The quantitative analysis of the changes of Q values (Q-Learning) when $\gamma = 0.1$, $\gamma = 0.3$, and $\gamma = 0.5$.

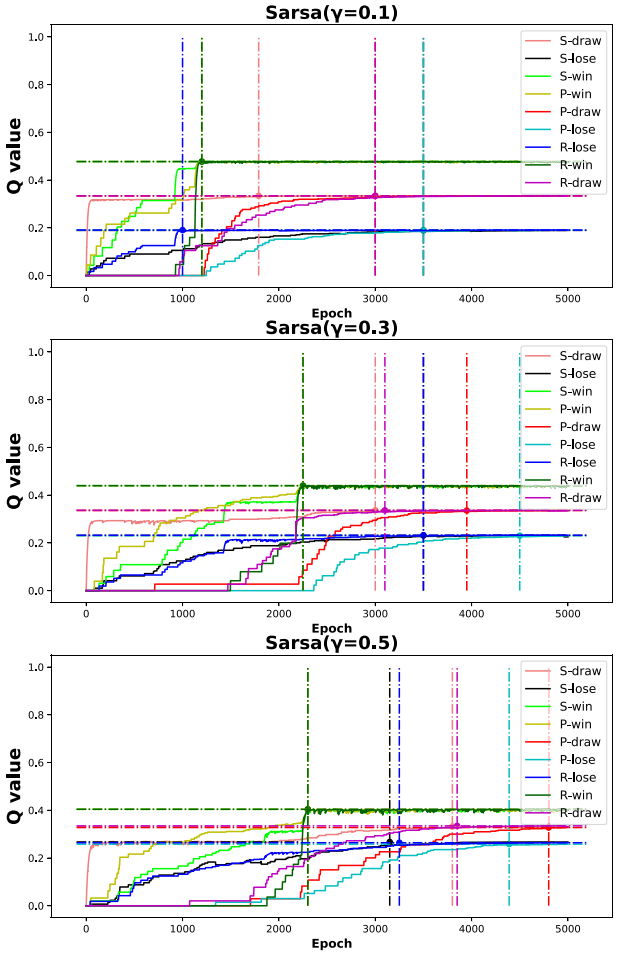

Fig. 9. The quantitative analysis of the changes of Q values (Sarsa) when $\gamma = 0.1$, $\gamma = 0.3$, and $\gamma = 0.5$.

Table 1
The comparison accuracy and computational time among the ensemble, k-NN, and DCNN classifier.

Method	Accuracy (%)	Computational time (t)	
		Training	Testing
Ensemble	100%	123.25 s	0.42 s
k-NN	100%	1.44 s	0.05 s
DCNN	92.22%	121.48 s	0.47 s

Table 2
The comparison accuracy and computational time based on updating strategy.

Method	Accuracy (%)	Computational time (t)	
		Training	Testing
Ensemble	100%	145.46 s	0.36 s
k-NN	100%	0	0.36 s
DCNN	100%	145.46 s	0.36 s

also provides a Reinforcement Learning environment with a 15-12400F (2.5 GHz) processor and 32 GB of RAM. A UhandPi Robot hand (Hiwonder, Shenzhen, China) is the interaction object which can be triggered on a given gesture by serial. It has five finger degrees of freedom in total using an Anti-blocking steering gear controlled by a 4B Raspberry pie.

5. Experiments and results

We designed several experiments to evaluate the performance of each AdaRL-MDF model, namely offline modeling, updating module, and RL model. The first experiment aims to prove the accuracy of the k-NN and DCNN classifiers by comparing the identification rate among k-NN, DCNN, and the ensemble model. The second one is to evaluate the updating capability by setting interference. The last aims to test the performance of three different RL-based RPS models.

We collected both sEMG signals and depth vision data from 10 subjects (3 females and 7 males), and each person held the gesture for one minute. Due to the sampling rate being 200 Hz, we get 12000 samples from each subject making one gesture. The detection length and overlap are 100 Hz and 50 samples. Finally, it acquires 723 segments (241 for each gesture). The experiment is implemented in Python with i5 Core, 16.0-GB RAM, and a 2.80-GHz CPU hardware platform.

5.1. Ensemble updating model performance

Table 1 shows the comparison results of accuracy and computational time among the ensemble, k-NN, and DCNN classifiers. It aims to prove that the ensemble classifier can get higher accuracy and lower computational time than only adopting the k-NN method or DCNN model. We use the training dataset from the ten subjects to build the three classifiers. The testing dataset is acquired from two new subjects. The results in Table 2 show that the DCNN classifier cannot always identify the three gestures,

Table 3
The comparison percentage (%) of between Q-learning and Sarsa based RL model ($\gamma = 0.1$).

Epoch	Method	3:2:1			3:1:0			3:0:0		
		Win	Draw	Lose	Win	Draw	Lose	Win	Draw	Lose
1000	QL	24.84	38.20	36.97	65.33	34.67	0	100	0	0
	Sarsa	25.83	41.00	33.17	61.50	38.50	0	100	0	0
2000	QL	36.17	31.83	32.00	57.50	42.50	0	100	0	0
	Sarsa	28.83	36.33	34.84	64.17	35.83	0	100	0	0
3000	QL	36.67	33.17	30.16	60.83	39.17	0	100	0	0
	Sarsa	41.17	27.50	31.33	59.17	40.83	0	100	0	0
4000	QL	38.83	34.17	27.00	59.00	41.00	0	100	0	0
	Sarsa	39.67	32.33	28.00	58.33	41.67	0	100	0	0
5000	QL	42.83	32.33	24.84	59.67	40.33	0	100	0	0
	Sarsa	42.83	31.83	25.34	59.33	40.67	0	100	0	0

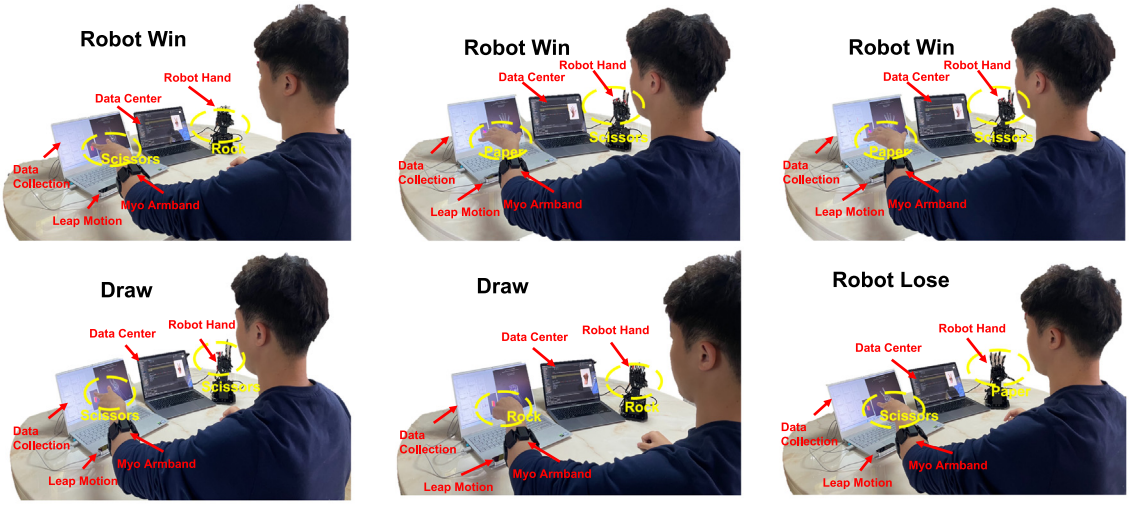

Fig. 10. The examples of human–robot confrontation gaming.

while the k-NN model can keep a higher accuracy even if a new dataset is obtained.

5.2. RL model

We use two methods to build the RL model: Q-learning and Sarsa. Three different Q values tables are also designed according to other incentive mechanisms. We set the win, draw, and loss ratios as '3 : 2 : 1', '3 : 1 : 0', and '3 : 0 : 0'. Table 3 shows the comparison results between Q-learning and Sarsa. However, it must find the best iterations by comparing these two methods.

Figs. 8 and 9 demonstrate the comparison results with different parameters γ, which means the discount factor of the learning. It is relevant to the weight of rewards earned earlier and received later. Therefore, we choose a small γ value to balance the values of all actions and observe learning processing rather than just finding the optimal policy. It can train the Q table more quickly when $\gamma = 0.1$ than the other two values. Meanwhile, the ratio will be close to the final value after 5000 epochs.

Fig. 10 displays several scenes when a human plays the RPS game with a robot hand.

6. Conclusion

We presented a new adaptive RL-based multimodal data fusion framework (AdaRL-MDF) for human–robot confrontation gaming, including an offline training step, an updating module, and an RL model. It builds an ensemble classifier to avoid some interference, such as occlusion, artifact, and insufficient sampling. The updating model remedies the problem of classification accuracy caused by a small sample array. Finally, the RL model can teach the robot to play games with humans by establishing the Q-learning or Sarsa model. Although the designed experiments show the excellent performance of the presented AdaRL-MDF model, some limitations also exist. For example, the activity space limits people's actions, and the current multimodal data fusion modal has low identification numbers. The multiple LMCs fusion system and multimodal network should be considered in our future work to manage the existing problems.

CRediT authorship contribution statement

Wen Qi: Conceptualization, Methodology, Software, Data curation, Formal analysis, Writing – Original draft preparation, Funding acquisition. **Haoyu Fan:** Software, Writing – Original draft preparation. **Hamid Reza Karimi:** Methodology, Visualization, Investigation, Resources, Supervision, Writing – reviewing, and editing. **Hang Su:** Methodology, Writing — reviewing and editing, Validation, Formal analysis, Software.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the Guangdong Provincial Key Laboratory of Human Digital Twin (2022B1212010004) and in part by the Italian Ministry of Education, University and Research through the Project "Department of Excellence LIS4.0-Lightweight and Smart Structures for Industry 4.0".

References

Abavisani, M., Joze, H. R. V., & Patel, V. M. (2019). Improving the performance of unimodal dynamic hand-gesture recognition with multimodal training. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1165–1174).

Ahmadi, E., Pour, A. G., Siamy, A., Taheri, A., & Meghdari, A. (2019). Playing rock-paper-scissors with RASA: A case study on intention prediction in human-robot interactive games. In *International conference on social robotics* (pp. 347–357). Springer.

Ahn, H. S., Sa, I.-K., Lee, D.-W., & Choi, D. (2011). A playmate robot system for playing the rock-paper-scissors game with humans. *Artificial Life and Robotics*, *16*(2), 142–146.

Brock, H., Chulani, J. P., Merino, L., Szapiro, D., & Gomez, R. (2020). Developing a lightweight rock-paper-scissors framework for human-robot collaborative gaming. *IEEE Access*, *8*, 202958–202968.

Chen, Y., Zhang, D., Karimi, H. R., Deng, C., & Yin, W. (2022). A new deep learning framework based on blood pressure range constraint for continuous cuffless BP estimation. *Neural Networks*, *152*, 181–190.

Fiorini, L., Loizzo, F. G. C., Sorrentino, A., Kim, J., Rovini, E., Di Nuovo, A., et al. (2021). Daily gesture recognition during human-robot interaction combining vision and wearable systems. *IEEE Sensors Journal*, *21*(20), 23568–23577.

Fu, J., Poletti, M., Liu, Q., Iovene, E., Su, H., Ferrigno, G., et al. (2022). Teleoperation control of an underactuated bionic hand: Comparison between wearable and vision-tracking-based methods. *Robotica*, *11*(3), 61.

Guo, L., Lu, Z., & Yao, L. (2021). Human-machine interaction sensing technology based on hand gesture recognition: A review. *IEEE Transactions on Human-Machine Systems*.

Ito, K., Sueishi, T., Yamakawa, Y., & Ishikawa, M. (2016). Tracking and recognition of a human hand in dynamic motion for Janken (rock-paper-scissors) robot. In *2016 IEEE international conference on automation science and engineering* (pp. 891–896). IEEE.

Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019). Q-learning algorithms: A comprehensive classification and applications. *IEEE Access*, *7*, 133653–133667.

Jiang, S., Kang, P., Song, X., Lo, B. P., & Shull, P. B. (2021). Emerging wearable interfaces and algorithms for hand gesture recognition: A survey. *IEEE Reviews in Biomedical Engineering*, *15*, 85–102.

Jiang, B., Wu, Z., & Karimi, H. R. (2022). A distributed dynamic event-triggered mechanism to HMM-based observer design for H ∞ sliding mode control of Markov jump systems. *Automatica*, *142*, Article 110357.

Joshi, A., Sierra, H., & Arzuaga, E. (2017). American sign language translation using edge detection and cross correlation. In *2017 IEEE Colombian Conference on Communications and Computing* (pp. 1–6). IEEE.

Kong, N. J., Payne, J. J., Council, G., & Johnson, A. M. (2021). The Salted Kalman Filter: Kalman filtering on hybrid dynamical systems. *Automatica*, *131*, Article 109752.

Li, C., Fahmy, A., & Sienz, J. (2019). An augmented reality based human-robot interaction interface using Kalman filter sensor fusion. *Sensors*, *19*(20), 4586.

Matsufuji, A., Sato-Shimokawara, E., Yamaguchi, T., & Chen, L.-H. (2019). Adaptive multi model architecture by using similarity between trained user and new user. In *2019 International Conference on Technologies and Applications of Artificial Intelligence* (pp. 1–6). IEEE.

Mohan, P., Sharma, L., & Narayan, P. (2021). Optimal path finding using iterative SARSA. In *2021 5th international conference on intelligent computing and control systems* (pp. 811–817). IEEE.

Nishimura, Y., Nakamura, Y., & Ishiguro, H. (2020). Human interaction behavior modeling using Generative Adversarial Networks. *Neural Networks*, *132*, 521–531.

Onnasch, L., & Roesler, E. (2021). A taxonomy to structure and analyze human–robot interaction. *International Journal of Social Robotics*, *13*(4), 833–849.

Park, K.-B., Choi, S. H., Lee, J. Y., Ghasemi, Y., Mohammed, M., & Jeong, H. (2021). Hands-free human–robot interaction using multimodal gestures and deep learning in wearable mixed reality. *IEEE Access*, *9*, 55448–55464.

Qi, W., Ovur, S. E., Li, Z., Marzullo, A., & Song, R. (2021). Multi-sensor guided hand gesture recognition for a teleoperated robot using a recurrent neural network. *IEEE Robotics and Automation Letters*, *6*(3), 6039–6045.

Qi, W., & Su, H. (2022). A cybertwin based multimodal network for ecg patterns monitoring using deep learning. *IEEE Transactions on Industrial Informatics*.

Qi, W., Wang, N., Su, H., & Aliverti, A. (2022). DCNN based human activity recognition framework with depth vision guiding. *Neurocomputing*, *486*, 261–271.

Qiao, H., Chen, J., & Huang, X. (2021). A survey of brain-inspired intelligent robots: Integration of vision, decision, motion control, and musculoskeletal systems. *IEEE Transactions on Cybernetics*.

Qiao, H., Zhong, S., Chen, Z., & Wang, H. (2022). Improving performance of robots using human-inspired approaches: a survey. *Science China. Information Sciences*, *65*(12), 1–31.

Qureshi, A. H., Nakamura, Y., Yoshikawa, Y., & Ishiguro, H. (2018). Intrinsically motivated reinforcement learning for human–robot interaction in the real-world. *Neural Networks*, *107*, 23–33.

Rahimian, E., Zabihi, S., Asif, A., Farina, D., Atashzar, S. F., & Mohammadi, A. (2021). Fs-hgr: Few-shot learning for hand gesture recognition via electromyography. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 1004–1015.

Skantze, G. (2021). Turn-taking in conversational systems and human-robot interaction: a review. *Computer Speech and Language*, *67*, Article 101178.

Spano, S., Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Giardino, D., Matta, M., et al. (2019). An efficient hardware implementation of reinforcement learning: The q-learning algorithm. *Ieee Access*, *7*, 186340–186351.

Su, H., Qi, W., Chen, J., & Zhang, D. (2022). Fuzzy approximation-based task-space control of robot manipulators with remote center of motion constraint. *IEEE Transactions on Fuzzy Systems*, *30*(6), 1564–1573.

Su, H., Qi, W., Schmirander, Y., Ovur, S. E., Cai, S., & Xiong, X. (2022). A human activity-aware shared control solution for medical human–robot interaction. *Assembly Automation*, (ahead-of-print).

Wang, D., Ha, M., & Zhao, M. (2022). The intelligent critic framework for advanced optimal control. *Artificial Intelligence Review*, 1–22.

Wang, L., Huang, W., Li, Y., Evans, J., & He, S. (2020). Multi-AI competing and winning against humans in iterated Rock-Paper-Scissors game. *Scientific Reports*, *10*(1), 1–8.

Wang, X., Karimi, H. R., Shen, M., Liu, D., Li, L.-W., & Shi, J. (2022). Neural network-based event-triggered data-driven control of disturbed nonlinear systems with quantized input. *Neural Networks*, *156*, 152–159.

Wang, Y.-H., Li, T.-H. S., & Lin, C.-J. (2013). Backward Q-learning: The combination of Sarsa algorithm and Q-learning. *Engineering Applications of Artificial Intelligence*, *26*(9), 2184–2193.

Wong, W., Juwono, F. H., & Khoo, B. T. T. (2021). Multi-features capacitive hand gesture recognition sensor: A machine learning approach. *IEEE Sensors Journal*, *21*(6), 8441–8450.

Yoon, H.-S., & Chi, S.-Y. (2006). Visual processing of rock, scissors, paper game for human robot interaction. In *2006 SICE-ICASE international joint conference* (pp. 326–329). IEEE.

Zhang, R., Lv, Q., Li, J., Bao, J., Liu, T., & Liu, S. (2022). A reinforcement learning method for human-robot collaboration in assembly tasks. *Robotics and Computer-Integrated Manufacturing*, *73*, Article 102227.