A Serious Game System for Upper Limb Motor Function Assessment of Hemiparetic Stroke Patients

Yuanbo Jiang, Zhen Liu, Tingting Liu, Minhua Ma, Min Tang, and Yanjie Chai

***Abstract*—Stroke often results in hemiparesis, impairing the patient's motor abilities and leading to upper extremity motor deficits that require long-term training and assessment. However, existing methods for assessing patients' motor function rely on clinical scales that require experienced physicians to guide patients through target tasks during the assessment process. This process is not only time-consuming and labor-intensive, but the complex assessment process is also uncomfortable for patients and has significant limitations. For this reason, we propose a serious game that automatically assesses the degree of upper limb motor impairment in stroke patients. Specifically, we divide this serious game into a preparation stage and a competition stage. In each stage, we construct motor features based on clinical a priori knowledge to reflect the ability indicators of the patient's upper limbs. These features all correlated significantly with the Fugl-Meyer Assessment for Upper Extremity (FMA-UE), which assesses motor impairment in stroke patients. In addition, we design membership functions and fuzzy rules for motor features in combination with the opinions of rehabilitation therapists to construct a hierarchical fuzzy inference system to assess the motor function of upper limbs in stroke patients. In this study, we recruited a total of 24 patients with varying degrees of stroke and 8 healthy controls to participate in the Serious Game System test. The results show that our Serious Game System was able to effectively differentiate between controls, severe, moderate, and mild hemiparesis with an average accuracy of 93.5%.**

***Index Terms*—Stroke, serious game, fuzzy Inference, motor function assessment, upper extremity assessment.**

Manuscript received 14 February 2023; revised 1 May 2023; accepted 26 May 2023. Date of publication 30 May 2023; date of current version 14 June 2023. This work was supported in part by the Ningbo Science Technology Plan Projects under Grant 2021S091 and Grant 2020Z082 and in part by the Natural Science Foundation of Zhejiang Province under Grant LZ23F020005. *(Corresponding author: Zhen Liu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Medical Ethics Committee of Medical School of Ningbo University under Application No. NBU-W-2022-002.

Yuanbo Jiang, Zhen Liu, and Yanjie Chai are with the Faculty of Information Science and Technology, Ningbo University, Ningbo 315211, China (e-mail: 810268678@qq.com; liuzhen@nbu.edu.cn; chaiyanjie@nbu.edu.cn).

Tingting Liu is with the College of Science and Technology, Ningbo University, Cixi 315300, China (e-mail: liutingting@nbu.edu.cn).

Minhua Ma is with the Games Academy, Falmouth University, TR11 4RH Cornwall, U.K. (e-mail: m.ma@falmouth.ac.uk).

Min Tang is with the Ningbo Rehabilitation Hospital, Ningbo 315040, China (e-mail: 2398493251@qq.com).

Digital Object Identifier 10.1109/TNSRE.2023.3281408

I. Introduction

STROKE is a disease with a high morbidity, disability, and mortality rate, which can seriously endanger the life of the patient if it strikes [1]. Stroke kills 10% of people worldwide and is the leading cause of death and disability in adults [2]. Patients disabled by stroke often require motor rehabilitation to help them regain their motor abilities by stimulating neural reconstruction through continuous motor training. The development of a rehabilitation training task requires the patient to complete a scale assessment that allows the therapist to understand the patient's disease status [3].

The Fugl-Meyer Assessment for Upper Extremity (FMA-UE) is a valid, feasible, and well-designed clinical test scale often used as a clinical research tool to assess changes in upper extremity motor deficits after stroke [4]. However, it is time-consuming for patients to complete the FMA-UE. In addition, the scale is highly subjective and requires regular assessment by an experienced therapist to guide patients through the assigned tasks. If a rehabilitation therapist is unable to assess a patient at home, the patient usually needs to return to the hospital for assessment, which not only consumes a lot of healthcare resources but also wastes a lot of the patient's time [5]. Not only that, the scales lack the sensitivity to assess subtle but important changes in exercise performance [6], as they rely on scoring criteria rather than a continuous measurement structure. Finally, the long and tedious scale assessment may cause discomfort to some patients. Therefore, there is a need to design a more interesting and faster-automated assessment method. In recent years, serious games have received more attention in the medical field and are widely used in gait balance training [7], assisted motor rehabilitation [8], [9], and assisted diagnosis [10]. Compared to previous approaches, game-based rehabilitation programs are more interesting and can be combined with interactive tasks to actively engage patients in rehabilitation [11]. In addition, serious games are a low-cost medical aid that can save medical resources [12]. Studies have been conducted to classify hemiplegia levels by assessing patients' kinematic information [13]. Kinematic assessment is a more accurate, objective, and sensitive assessment method [14], which describes the motion of the human body in space and time, including velocity, acceleration, and angular displacement, and it is increasingly used in upper limb assessment [15]. Considering the unique advantages of serious games and kinematics in medical rehabilitation, we focused on designing an upper limb movement assessment method based on serious games and incorporating kinematic information into it.

Existing serious game-based upper limb assessment methods for stroke patients typically collect information about the patient's hand in real-time during game interaction and combine machine learning or neural network models to assess the patient [16], [17], [18]. Although several of these works have achieved good results in assessing patients' upper limb motor abilities based on serious games, there are still three problems. Problem 1: Serious game design problem: Existing methods collect information from patients only in serious games. However, stroke patients often suffer from physical and neurological effects during intensive games and often show tension [19]. The assessment results only reflect how patients are doing during the game, not how they are doing in their daily life. Problem 2: Joint deficit problem: The existing methods rely only on the motion information of the hand joints to assess the upper limb motor ability, ignoring the influence of other joints of the upper limb on motor ability. In rehabilitation assessment, it is necessary to combine the kinematic characteristics of the hand, elbow, shoulder, and other joints to assess the patient's upper limb motor ability. Problem 3: Data labeling problem: Existing methods use data with labels to train machine learning models, but labels are difficult to obtain in real rehabilitation settings and require experienced physicians to manipulate, so it is difficult to label all data. The above three problems lead to the fallacy of haphazard generalization in the existing upper limb movement assessment methods, which do not fully reflect the upper limb movement ability of stroke patients.

To solve the above problems, we proposed a set of serious games to assess the upper limb motor ability of stroke patients by combining clinical a priori knowledge. Robinson et al. [20] suggested that stroke patients often have physical and neurological impairments, which are manifested by decreased physical function, emotional loss, and eased tension and agitation. Based on the above, to solve the first problem, we divided the serious game into a preparation stage and a competition stage. In the preparation stage, there is no continuous interactive task when the patient is physically active and mentally relaxed. At this point, the system reminds the patient to lift the ball into a specific area to start the game. The purpose of this stage is to guide the patient to perform guided movements, and the upper limb information collected during this process can reflect the patient's ability to perform daily life. In the competition stage, we design intensive, interactive tasks in which the system will play against the patient in a ping-pong match format. The purpose of this stage is to guide the patient to perform spontaneous movements and collect kinematic information reflecting the patient's upper limbs. The data collected by this method is more representative of the patient's actual condition. To address the second question, first, referring to the assessment content in the FMA-UE [21], we assessed the patient's upper limb motor ability in daily life based on tremor information, stretching information, timing information, and control information together in the game preparation stage. Secondly, according to the upper limb kinematic assessment content, we regarded hand-elbow-shoulder as a motion whole in the game competition stage, where each joint reflects the kinematic information of the upper limb. Then, following this idea, we designed several kinematic features to reflect the above information of the patient. For problem 3, we designed membership functions for the upper limb motion features by combining the experience of rehabilitation therapists with fuzzy inference methods to avoid the problem of relying on large-scale data annotation and to be highly robust to unfamiliar gamers. In summary, the main contributions of this paper include the following.

We propose a novel set of serious game design ideas for the assessment of motor impairment in stroke, divided into two stages: preparation and competition. The information collected in the preparation stage better reflects the user's upper extremity daily mobility, while the information collected in the competition stage better reflects the patient's upper extremity kinematic information. The assessment results based on these two stages are more representative.

We analyze the upper extremity movements of stroke patients from a more comprehensive and complete perspective, considering kinematic information of the hand, elbow, and shoulder joints. We also combine a priori knowledge of rehabilitation medicine to design motion features that describe the patient's upper extremity movements in detail. We also use a fuzzy inference method without sample markers to assess upper extremity motor deficits in stroke patients.

We recruited 24 hemiplegic patients and 8 healthy participants to experience the Serious Game System and labeled the patients' upper limb abilities by manual FMA-UE, which was used to examine the results of subsequent Serious Game System assessments. The results show that the Serious Game System achieved 93.5% accuracy in assessing upper limb hemiparesis in stroke patients.

In this paper, we discuss related work in Section II, describe data collection methods in Section III, present the system design in Section IV, and provide results and discussion in Sections V and VI, respectively. Finally, we conclude our work in Section VII.

II. Related Work

In this section, we present an automatic assessment of motor ability and a serious game-based assessment of motor ability, respectively. We discuss the strengths and weaknesses of these studies separately and used to refine the design of our approach.

A. Automatic Assessment Methods for Motor Ability

Automated assessment methods for motor ability can be classified into two types: wearable and non-wearable, based on the type of device used. Wearable devices can accurately

TABLE I
PARTICIPANT INFORMATION

Level of Hemiparesis	Number	Female	Left paralysis	Age	FMA Score
Severe	8	3	4	57.8 ±5.6	14.5 ±9.9
Obvious	8	2	3	55.2 ±10.6	42.5 ±7.8
Moderate	8	4	2	56.7 ±10.8	62.3 ±2.6
Control	8	4	N.A.	52.7 ±9.3	N.A.

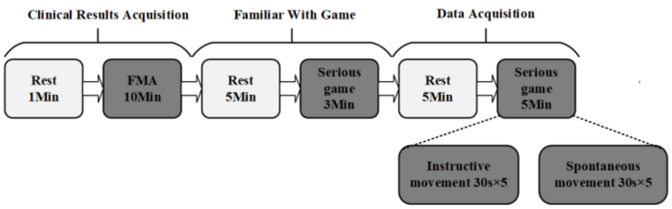


Fig. 1. Data collection flowchart.

Fig. 2. A patient using an experimental device.

collect upper limb movement data, while non-wearable devices offer a less-invasive experience that is easier for patients to accept.

Lee et al. [22] developed a smart wristband device that continuously captures acceleration signals from the wrist joint of acute stroke patients through spontaneous movements or specific movement commands. However, this method is limited in scope as it focuses solely on a specific joint of the upper extremity and fails to analyze overall motion characteristics of other joints. Wang et al. [23] proposed a multimodal-based method for assessing upper limb motility in patients with brain injury. They combined electromyography (EMG) signals and patient upper limb motion information, analyzing the similarities and differences between these two modalities of information. Additionally, they integrated three learning models, including machine learning, neural networks, and ensemble learning, for motor impairment classification. Although they were the first to use multimodal information for comprehensive upper limb assessment in the field, collecting EMG signals requires electrodes to be attached to the patient's skin, which is not only resource-intensive but also may cause discomfort and limit deployment in the patient's living environment. Furthermore, their method emphasizes muscle electrical signals rather than analyzing overall motion characteristics, potentially resulting in information redundancy. Additionally, using deep models as target classifiers requires a large number of manually annotated data samples. Bosecker et al. [24] developed a set of robots for assessing upper limb dyskinesia, which centrally collected kinematic and kinetic data. Similarly, Zhang et al. [25] proposed an upper limb rehabilitation robot that collected patient data information and used machine learning methods for classification. However, utilizing machine learning as a classifier still has significant limitations due to the problem of data labeling. Previous studies requiring physical contact with the patient to collect data can lead to discomfort and can affect assessment results. In contrast, Lee and Olesh et al. [22], [26] utilized a contactless approach with a depth camera to collect patient motion information, which is less invasive and more easily deployable. They also developed a rule-based binary logic classification method that evaluates based on prior logical rules, eliminating the need for a large number of labeled samples.

Overall, automated methods for assessing motor ability circumvent the subjectivity associated with traditional FMA evaluations, providing inspiration for future work. However, from the patient's perspective, these methods can be monotonous and may result in reluctance or non-cooperation. Therefore, it is necessary to design engaging motor impairment assessment methods. Our approach is based on serious games, ensuring accuracy while providing comfort and entertainment.

B. Automatic Assessment Method for Motor Ability Based on Serious Games

In recent years a large number of researchers have worked on the development of serious games for the assessment of motor abilities with the aim of identifying patients' lesions and assessing their motor abilities during recreation. Methods based on serious games have the unique advantages of being convenient and entertaining. In this section, we discuss the approach to motor assessment using serious games. Serradilla et al. [27] pioneered the development of a serious game for the assessment of motor impairment in stroke patients. Based on this, they tested the motor function of patients' upper extremities and administered the actual test to 33 stroke patients. Their results show that it is just as possible to assess stroke patients accurately in a serious game. Bai and Song [16], developed a serious game for upper limb training and assessment based on home scenarios, using only a depth camera and a posture sensor, to obtain better assessment results. In a similar vein, Cho and Pei et al. [28], [29] focused on a home setting and designed a method for assessing motor ability in stroke patients by using a depth camera to collect data from the patient's hand to assess upper limb motor ability. However, they ignored the holistic nature of hand-elbow-shoulder motion, and a single analysis of hand data cannot be applied to all patients and has major limitations. the Lee et al. [18] task relied only on perceived data from the game to assess patients' motor ability, and their results lacked convincingness. They used perceived data from a serious game and a motor rehabilitation scale as input to their assessment model, using a machine learning model as a classifier. This approach requires a large amount of data annotation and the results validated on a small data set are not convincing. Oña et al. [17] proposed a virtual version of the FMA scale test using the serious game to assess patients' upper limb abilities and balance abilities simultaneously. Although the above work has made some attempts to assess patients' motor impairment abilities using serious games, there are still major problems with the game design. First, the behavior of patients in the game does not reflect their daily life state, so the assessment results are limited to the results in the game, but often medical treatments, which are developed in the daily environment, lead to a failure to combine assessment results and treatments. Moreover, feature selection is limited, with most serious games for stroke relying on data from the hand joints and ignoring information from the elbow and shoulder. The above problems limit the application of serious games in the field of motor assessment. For this reason, we propose a novel design idea in developing serious games. The game is divided into a preparation stage and a competition stage, in which tasks are designed to be as easy and simple as possible. These data can reflect the state of patients in their daily life. In the competition stage, we designed the interactive game of playing ping pong with a high frequency of interaction, which is difficult to occur in the daily environment of stroke. In addition, we collect data on hand, elbow and shoulder joints simultaneously in the game and design several motor features for assessing patients' motor abilities.

III. Data Collection

A. The Fugl-Meyer Assessment for Upper Extremity (FMA-UE)

The Fugl-Meyer Assessment for Upper Extremity (FMA-UE) Scale is an impairment scale based on stroke presentation and is used to assess upper extremity sensorimotor function. A maxi mum score of 66 on the FMA-UE indicates normal upper extremity function. The scale has good validity and reliability and is commonly used to classify the severity of upper extremity stroke impairment [4]. A score of less than 32 on the FMA-UE is defined as severe hemiparesis, between 32 and 57 as moderate hemiparesis, and between 58 and 66 as mild hemiparesis.

B. Participants

Data were collected from the Neurorehabilitation Department of Ningbo Rehabilitation Hospital. Participants recruited for this study were all paralyzed in the upper extremities due to a stroke within the past year. Participants were informed of the data collection prior to being tested. A total of 25 hemiplegic patients were recruited for this test, of whom 1 declined to participate in the program, giving a total of 24 hemiplegic patients. As shown in Table I, 8 were severely hemiplegic (57.8±5.6 years) with a mean FMA score of 14.5±9.9; 8 were moderately hemiplegic (55.2±10.6 years) with a mean FMA score of 42.5±7.8; and 8 were mildly hemiplegic (56.7±10.8 years) with a mean FMA score of $\alpha\beta$ 62.3±2.6. In addition to this, 8 healthy participants (52.7±9.3 years) were recruited as controls. All participants were free of severe cognitive impairment, understood the content of the game, followed the physician's instructions and interactions, understood the purpose of the experiment, and agreed to the experiment. Participants are watched by a physician during the experiment to ensure they do not have accidents.

C. Protocol

The data collection process is shown in Fig. 1. First, after a 1-minute rest period, participants were assessed for FMA upper extremity motor function by two professional therapists, and throughout the process, participants were required to follow the therapists' instructions. This procedure lasted for 10 minutes and was designed to obtain the results of the participant s' clinical assessment for subsequent evaluation of the classification accuracy of the Serious Game System. Subsequently, after a 5-minute break, the serious game was performed for 3 minutes, and the purpose of the procedure was to familiarize the participants with the operation of the game and to avoid distractions due to the game experience. Finally, after a 5-minute break, a formal serious game test was performed in which each patient performed five instructive and spontaneous movements for approximately 1 minute each, during which the data generated during the game were collected.

Fig. 2 shows a participant undergoing a serious game test. Our experimental equipment consisted of a computer, monitor, and Kinect depth camera. The monitor was placed directly in front of the participant at a distance of 1.5 meters to ensure that the experiment would not be affected by the participant's vision. While the participants were playing the game, the Kinect camera would acquire their motion information in real-time and save it in the computer database. During the data collection phase, the average duration of the game played by the 32 patients was 5.12 ± 0.23 minutes.

IV. Serious Game System

The framework of our serious game system for upper limb motor assessment is shown in Fig. 3. First, participants were assessed by two experienced therapists through a manual FMA-UE to obtain qualitative evaluation results of the clinical assessment. Then, participants underwent a serious game assessment. During this time, we used a depth camera to collect the patient's upper extremity data during the preparation and game stages. During the preparation stage, we collected information on tremor, extension, timing and control, focusing on the patient's ability to perform daily activities of the upper extremity; during the competition stage, we collected kinematic information on the hand, elbow, and shoulder, focusing on overall upper extremity ability and responsiveness. To assess the aforementioned abilities, we designed several motor features to describe the patient's performance in the serious game. Then, we screened the features for significance by non-parametric tests. Then correlation analysis was performed, and the features with significance and the top three correlation

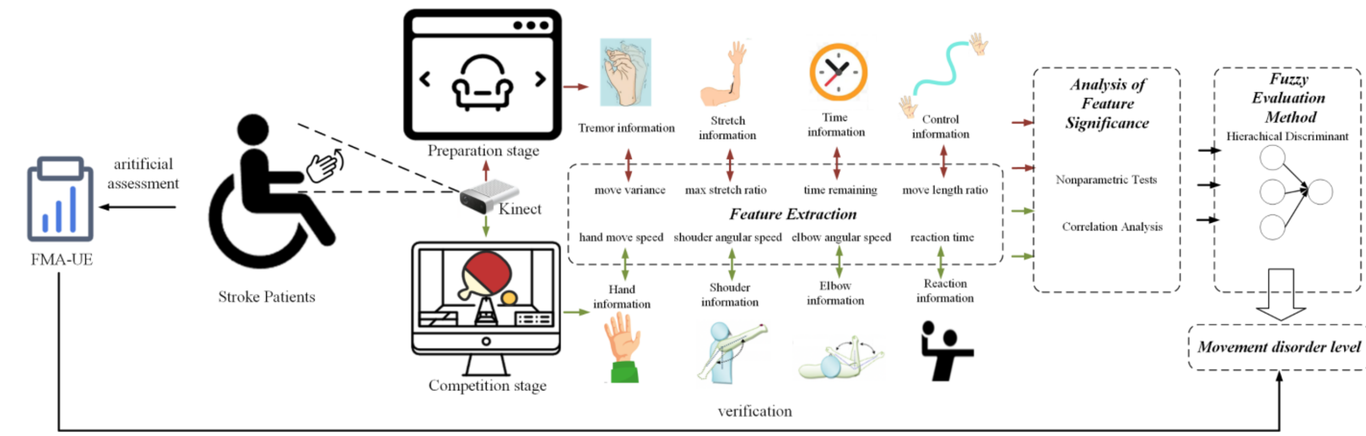
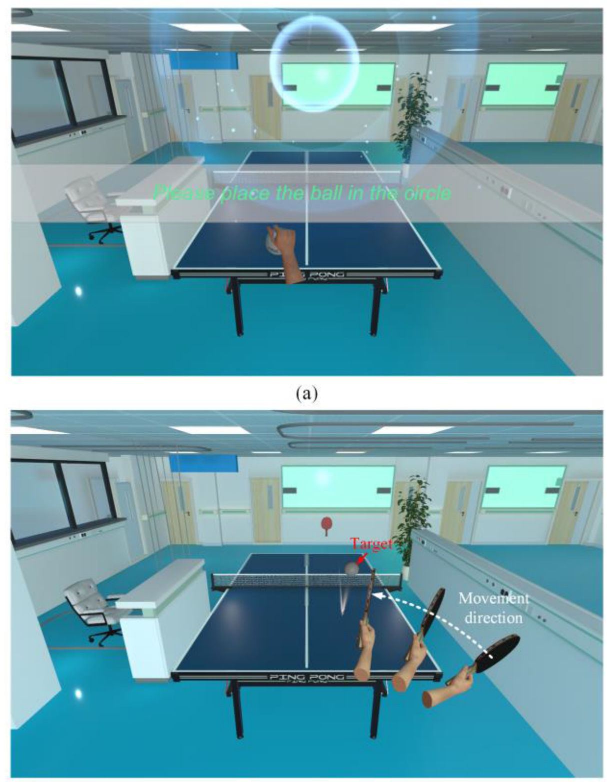


Fig. 3. The framework diagram illustrates the steps to identify hemiparesis in a serious game system.

coefficients were used as the input to the hierarchical fuzzy system. After that, we designed a membership function for each input feature in each stratified fuzzy subsystem to assess patients' motor ability by combining prior knowledge of rehabilitation and fuzzy mapping. Finally, we validated the accuracy of the serious game system in assessing the level of upper limb paralysis of patients.

A. Serious Game Design

Our serious game was developed using the Unity3D engine and compiled in C#. The game involves patients participating in a ping pong tournament. The game is divided into two stages, the preparation stage and the competition stage. The aim of the preparation stage is to make the patient do instructive movements, and the aim of the competition stage is to make the patient do spontaneous movements. In the preparation stage, the patient needs to control the movement of the affected hand. The Kinect Azure depth camera acquires information about the patient's joint points and maps the position of the patient's hand to the virtual hand in the virtual scene. As shown in Fig. 4(a), the patient's goal is to grab the ping pong ball on the table and move it to the range indicated by the aperture to complete the task and enter the competition stage. In this stage, the system does not give the patient a time limit cue in order to make the patient as relaxed as possible. Instead, considering that some patients with more severe motor impairment cannot complete the target task after catching the ball, the system automatically times the task. When the patient does not complete the task for more than 30 seconds, the system will end the preparation stage and enter the competition stage. In the competition stage, the patient needs to control the ping-pong paddle in the virtual scene and compete with the computer; as shown in Fig. 4(b), the patient needs to try his best to imitate the standard striking posture and swing his arm rapidly to hit the ping-pong ball. The faster the patient swings, the faster the ball returns, and the more likely he will win the game. In addition, both the size of the ping-pong ball and the speed of the ping-pong ball return was appropriate for elderly patients to avoid affecting the assessment results.



(a)

(b)

Fig. 4. Two game modes: (a) is the preparation stage screen and (b) is the competition stage screen.

B. Feature Design

1) Preparation Stage: In the preparation stage, the patient needs to use the paralyzed hand to control the virtual hand in the virtual scene to grasp the ping pong ball on the table and move it to a specific area. We considered four motion features, *i.e.*, movement variance, movement length ratio, remaining time, and maximum stretch ratio. As shown in Fig. 5(a), when the patient controls the virtual hand to move along the standard trajectory. In this case, the movement variance is smaller, the maximum stretch ratio is larger, the remaining time is longer, and the movement length ratio is larger. Fig. 5(b) shows the patient's motion along an irregular motion trajectory. In this case, the movement variance is larger, the maximum stretch ratio is larger, the remaining time is shorter, and the movement length ratio is smaller. The formulae for each motion feature are described in detail below.

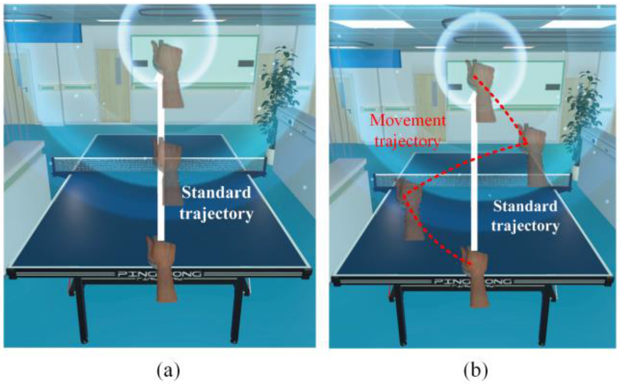


Fig. 5. Different movement trajectories in the preparation stage: (a) is moving along the standard trajectory, (b) is moving along the irregular trajectory.

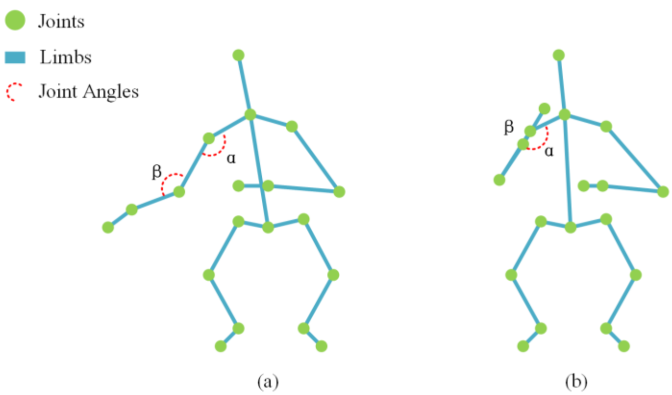


Fig. 6. Skeleton diagram of the user's swinging action during the competition stage: (a) is the start of the action, (b) is the end of the action. where α is the shoulder angle and β is the elbow angle.

-Movement variance: The deviation between the actual trajectory of the patient's hand movement and the standard trajectory was calculated using the straight line between the starting point and the endpoint as the standard trajectory, as shown in (1).

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} \Delta d_i^2}{n}} \tag{1}$$

where σ is the movement variance, Δd is the movement deviation, and n is the sampling frequency. The movement variance can be used to express the tremor information of the patient' s.

-Movement length ratio: the actual distance of the patient's hand movement during the preparation stage time divided by the straight-line distance between the starting point and the endpoint, as shown in (2).

$$m = \frac{\Delta D}{d} \tag{2}$$

where m is the movement length ratio, d is the actual movement distance, and ΔD is the linear distance between the two points. The larger the movement length ratio, the more control the patient has over the affected arm.

-Remaining time: the time left after the patient completes the target task in the preparation stage. When the patient starts to move the ball, the system will automatically record the time. When more than 30 seconds are spent in the stage, the system will automatically end the stage and enter the competition stage, as shown in (3).

$$\Delta T = \begin{cases} 30 - T, & T < 30 \\ 0, & T \geq 30 \end{cases} \tag{3}$$

where ΔT is the time remaining on serve and T is the time on serve, $\Delta T = 30 - T$ when $T < 30$ seconds and $\Delta T = 0$ when $T \geq 30$ seconds.

-Maximum stretch ratio: The system asks the patient to move the ball within a specific area, which requires the patient to lift the arm, at which point the system records the relative height of the patient's elbow to shoulder and hand to the elbow, as shown in (4).

$$p_{max} = Max\left(p_i = \frac{(h_i - e_i) + (e_i - s_i)}{l}\right) \tag{4}$$

where p_{max} is the maximum stretch ratio, h_i is the height of the hand, e_i is the height of the elbow, s_i is the height of the shoulder, and l is the length of the arm. The maximum stretch ratio reflects the patient's ability to extend the upper extremity.

2) Competition Stage: In the competition stage, the patient will play a ping pong match with the computer, and the patient needs to judge where the ball will land and hit the ball back. Fig. 6 shows the skeleton diagram of the patient, which records the information of the skeleton point position of the patient's whole body from before to after the ball is hit. From the skeleton point information, we can get the joint angle information, α is the shoulder joint angle, and β is the elbow joint angle. When the patient completes a round of strokes, the above joint angles change significantly, from which we record the velocity information of the patient's hand, elbow, and shoulder. Referring to hand information alone can lead to biased results. Some patients' hands may move through the trunk rather than through the upper limbs, so it is necessary to refer to the elbow and shoulder information, and the formulae for calculating each motion feature are described in detail below.

-Maximum hand acceleration: The maximum value of the instantaneous acceleration of the patient's hand movement, as shown in (5).

$$\delta_{max} = Max\left(\delta_i = \frac{\Delta v_i}{\Delta t_i}\right) \tag{5}$$

where δ_{max} is the maximum acceleration of the hand, Δv_i is the instantaneous velocity difference of the current hand, Δt_i is the instantaneous time difference, and the maximum acceleration reflects the explosive force of the patient's upper limb.

-Average hand velocity. In this round, all instantaneous hand velocities of the patient are added up and then averaged, as shown in (6).

$$\overline{v} = \frac{1}{n}\sum_{i=1}^{n}\left(v_i = \frac{\Delta d_i}{\Delta t_i}\right) \tag{6}$$

where v_i is the hand movement speed, Δd_i is the instantaneous movement distance, and Δt_i is the instantaneous time difference. The average hand speed reflects the patient's upper body strength and endurance throughout the competition.

-Maximum shoulder angular velocity: The maximum value of the instantaneous angular velocity of the patient's shoulder, as shown in (7).

$$\omega_{max} = Max\left(\omega_i = \frac{\Delta\alpha_i}{\Delta t_i}\right) \tag{7}$$

where ω_{max} is the maximum angular velocity of the shoulder, $\Delta\alpha_i$ is the instantaneous displacement angle of the shoulder, and Δt_i is the instantaneous time difference. The maximum angular velocity of the shoulder reflects the explosive force of the patient's shoulder.

-Average shoulder angular velocity: All instantaneous angular velocities of the patient's shoulder during the round were summed and then averaged, as shown in (8).

$$\bar{\omega} = \frac{1}{n}\sum\nolimits_{i=1}^{n}\frac{\Delta\alpha_i}{\Delta t_i} \tag{8}$$

where $\bar{\omega}$ is the average shoulder angular velocity, $\Delta\alpha_i$ is the instantaneous shoulder displacement angle, and Δt_i is the instantaneous time difference. The average shoulder angular velocity reflects the durability of the patient's shoulder joint strength during the competition.

-Maximum elbow angular velocity: the maximum value of the instantaneous angular velocity of the patient's elbow, as shown in (9).

$$\rho_{max} = Max\left(\sigma_i = \frac{\Delta\beta_i}{\Delta t_i}\right) \tag{9}$$

where ρ_{max} is the maximum elbow angular velocity, $\Delta\beta_i$ is the instantaneous elbow displacement angle, and Δt_i is the instantaneous time difference. The maximum angular velocity of the elbow reflects the explosive force of the patient's elbow.

-Average elbow angular velocity: All instantaneous angular velocities of the patient's elbow during the round were summed and then averaged, as shown in (10).

$$\bar{\rho} = \frac{1}{n}\sum\nolimits_{i=1}^{n}\frac{\Delta\beta_i}{\Delta t_i} \tag{10}$$

where $\bar{\rho}$ is the average elbow angular velocity, $\Delta\beta_i$ is the instantaneous elbow displacement angle, and Δt_i is the instantaneous time difference. The average elbow angular velocity responds to the persistence of the patient's elbow force during the competition.

-Reaction time: the time for the patient's hand to move in the direction of the ball after the computer hits the ball back, as shown in (11).

$$\Delta t_{reaction} = t_{action} - t_{return} \tag{11}$$

where $\Delta t_{reaction}$ is the reaction time of the patient's movement after the computer returns the ball. t_{action} is the time when the hand movement occurs and t_{return} is the time when the computer returns the ball.

C. Feature Significance Analysis

Statistical analyses were performed using IBM Statistical Package for Social Sciences™(SPSS) version 26 for Windows. Due to the small sample size of the data and the fact that most of the variables were non-normally distributed, a non-parametric test was used. We select some features for the construction of the classification model. To determine the statistical significance of each feature, the nonparametric Kruskal-Wallis test was used, and features with $p < 0.05$ were identified as statistically significant. And to test the ability of each variable to distinguish different groups in each layer, the statistical significance of each pair of category species was tested by the multiple comparison method.

To further analyze the significance of the features and select them to build a classification model, we performed Spearman correlation analysis on the features and ranked the features by correlation coefficients to filter the most relevant features to the classification results for each layer of the fuzzy system to maximize the detection probability and thus build a good classification model [30]. In order to build a classification model with the limited size of our dataset, we selected the top three features for each layer of the subsystem in terms of relevance coefficient to prevent an excessive number of rules in the fuzzy inference system.

D. Fuzzy Inference System (FIS)

Fuzzy Inference System (FIS) is one of the most practical tools in the context of fuzzy theory for dealing with nonlinear but ill-defined mappings of input variables to some output variables. Its construction consists of the following steps.

1) Determine the Membership Function: In order to extract the fuzzy features of the input information, it is necessary to fuzzify the input information with the membership function [31]. Since the trapezoidal membership function and triangular membership function are more general [32], we use these two membership functions. For all input variables, we use the trapezoidal membership function with the formula shown in (12), where a, b, c, and d are the four vertices of the trapezoidal membership function. For all output variables, the triangular membership function is applied, and its formula is shown in (13), where e, f, and g are the three vertices of the triangle membership function.

$$\mu_A(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & x \ge d \end{cases} \tag{12}$$

$$\mu_B(x) = \begin{cases} \frac{x-e}{f-e}, & e < x < f \\ 1, & x = f \qquad i \\ \frac{g-x}{g-f}, & f < x < g \end{cases} \tag{13}$$

2) Define Fuzzy Rules: Fuzzy rule evaluation is one of the keys to a fuzzy system. The number of fuzzy rules depends on the number of input features and their membership functions [33]. If there are too many input variables will lead to an excessive number of rules in the system, making it difficult

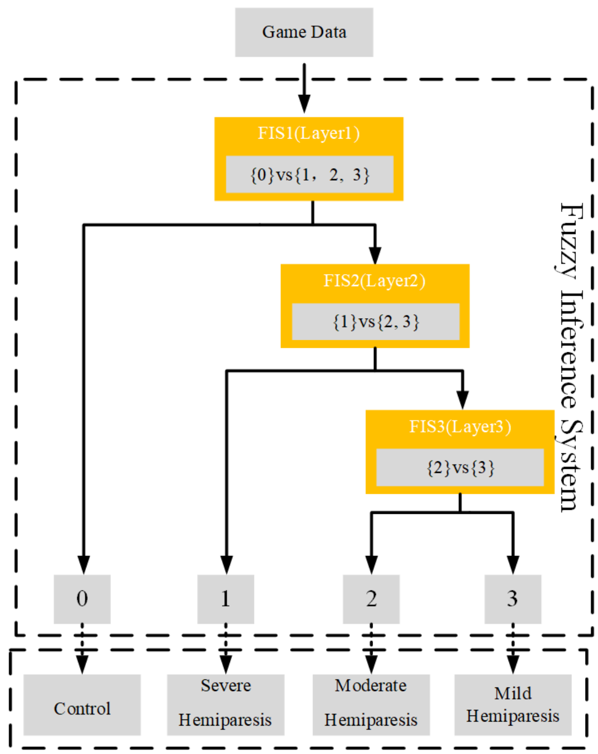


Fig. 7. Structure diagram of hierarchical fuzzy inference system.

to keep track of them [34]. Therefore, using hierarchical fuzzy inference can effectively avoid excessive fuzzy rules [35].

As shown in Fig. 7, we construct the system as a tree-like structure with three FIS subsystems. The first FIS subsystem can initially determine whether the patient belongs to the healthy control group or the hemiparesis group; on the basis of this FIS subsystem, a second layer is constructed with another FIS subsystem to determine whether it belongs to the {moderate, mild} hemiparesis group or the severe hemiparesis group. Finally, another FIS subsystem is used to determine whether it belongs to the mild hemiparesis group or the moderate hemiparesis group, forming a three-layer fuzzy inference system. Each subsystem uses only a portion of the features of the patient game data, which greatly reduces the number of fuzzy rules. Second, the hierarchical structure correlates with the clinical presentation of the patient. We found a significant difference in game performance between participants in the control group compared to those in the hemiparesis group, with participants in the hemiparesis group performing lower than the control group in both the preparation and competition stages, so they were differentiated in the first layer. In the second layer, patients with severe hemiparesis had severe arm paralysis compared to those with moderate and mild hemiparesis, had great difficulty performing the exercise, and had the worst performance within the game. A subset of patients was unable to lift their arms to complete the task during the preparation stage of the game, whereas this was not the case for patients with moderate and mild hemiparesis, so the {moderate, mild}, and severe groups were distinguished in this layer. After excluding the control and severe groups, the mild and moderate groups were distinguished at the end, as they both had some control over the paralyzed hand and performed more similarly.

3) Inference Engine and Defuzzification: The fuzzy interface engine is applied to integrate the identified fuzzy sets and to consider the fuzzy rules and the associated fuzzy regions separately. The fuzzy inference process uses ”min-max inference” to compute the conclusions of the rules based on the input values of the system [36]. The result of this process is named a ”fuzzy conclusion”. The applicability or ”truth value” of a rule comes from the combination of rule antecedents. Since the conjunction is defined as ”minimal,” rule evaluation consists of detecting the smallest (minimal) rule antecedent, which is considered to be the rule's truth value. This true value is then applied to all consequences of the rule. If any fuzzy output is a consequence of more than one rule, this output is set to the highest (maximum) truth value of all rules containing their consequences. The result of the rule evaluation is a set of fuzzy conclusions reflecting the effect of all rules with a truth value greater than 0 [37].

The fuzzy output is converted to clear output using the fuzzification process, and the main defuzzification methods are the mean of maximum (MOM) method, smallest of maximum (SOM) method, largest of maximum (LOM) method, bisector of area (BOA) method, and center of area (COA) method [38]. In our fuzzy inference model, the COA method is applied to all FISs.

4) Performance Validation: We tested the classification performance of each category, *i.e.*, accuracy, recall, precision, and F1-score [39], using a fuzzy inference system. We also used common machine learning methods used to make comparisons, namely Support Vector Machine (SVM), Classification and Regression Tree (CART), and k-Nearest Neighbours (kNN). Significant features with $p < 0.05$ were used for model training. Due to the small sample size of the dataset, a leave-one-out cross-validation method was used to calculate the average classification performance of the model [40]. In order to verify the necessity of the features used, the performance of the classifier was calculated for using all features, using only preparation stage features, and using only competition stage features, respectively.

V. Results

A. Game visualization results

The system recorded participants' motion information in real-time during the preparation stage of the first round and then displayed their motion trajectory lines in a visualized form through the Unity engine (as shown in Fig. 8). The motion trajectories of the control, mild, moderate and severe group are shown in the figure, and each color line represents one participant, and a total of 32 participants were recorded. In terms of the graph as a whole, there is a large difference between people without hemiparesis and those with hemiparesis. The control group's movements are very stable and do not shift excessively when performing guided movements, and all of them are able to complete the task efficiently in a shorter period of time. Mild hemiparesis, on the other hand, produces minimal motor deviation compared to other degrees

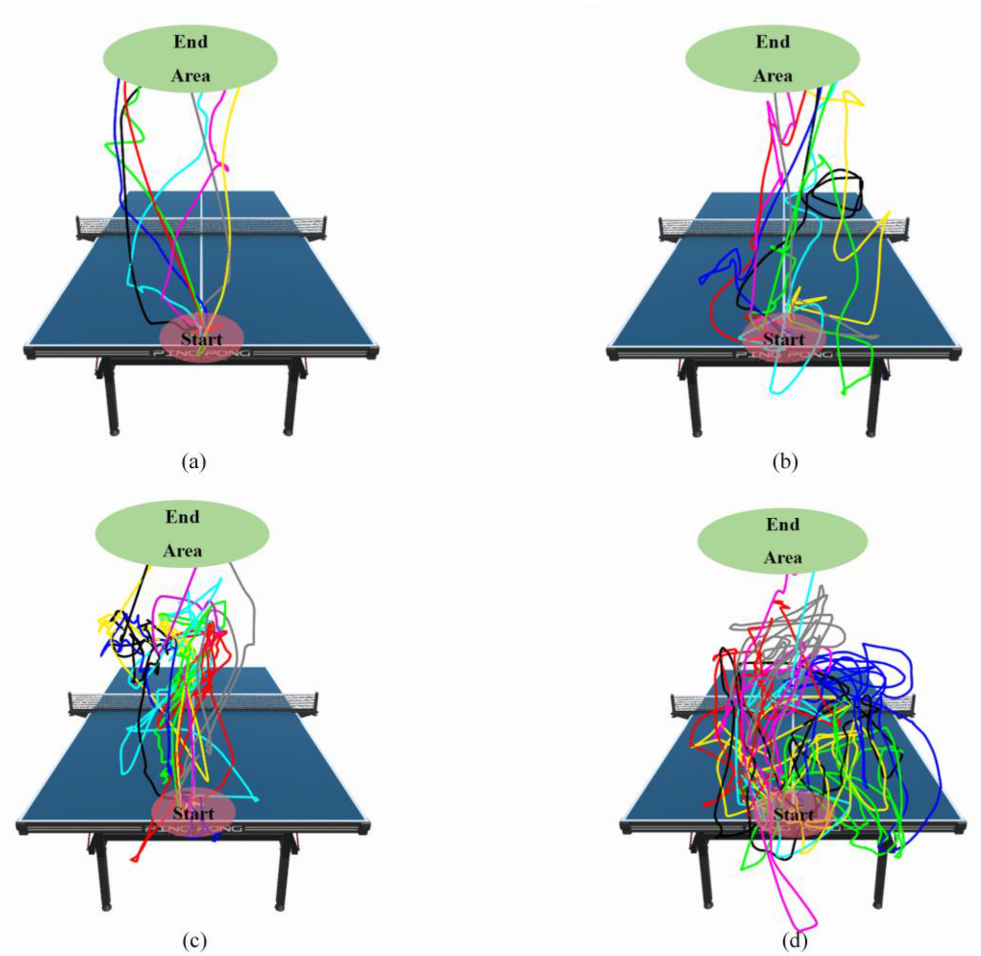


Fig. 8. Trajectory chart for patients in the preparation stage: (a) is control, (b) is mild hemiparesis, (c) is moderate hemiparesis, (d) is severe hemiparesis.

and all are able to complete the task. Moderate hemiparesis not only produced greater deviations, but also more jitter, and a small number of participants had limited ability to stretch to complete the task, thus taking significantly more time. Severe hemiparesis was the worst performer, with all participants producing great deflections and jitter, and only a few participants were able to complete the task within the time, and most would fail to complete the task before the time ran out.

We sliced all the velocity profiles of spontaneous movements during the competition stage to obtain the velocity slices of swinging movements for each participant. The velocity slices of all participants were then averaged by class to obtain the velocity profiles of the oscillatory movements of the control, mild, moderate, and severe groups (as shown in Fig. 9). The velocity of the control group increased smoothly and steadily, and there was a large difference between the control group and the hemiparesis group. The velocity curve of the hemiparesis group was not smooth, and the change trend had an abrupt rise or fall. The severe group had the lowest overall velocity and an insignificant upward velocity trend. The moderate and mild groups were in between the control and severe groups, with irregular velocity changes.

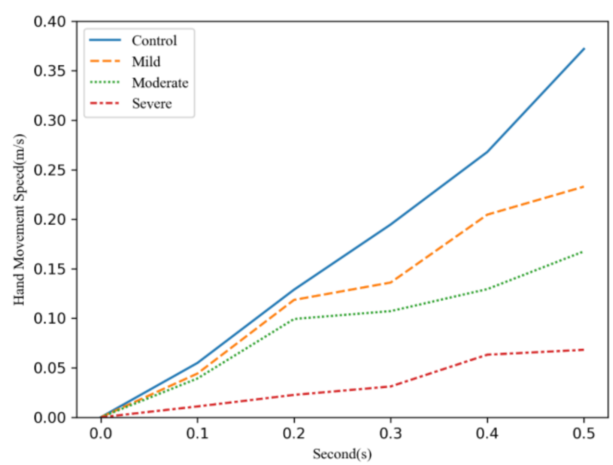


Fig. 9. Speed profile of a swinging motion. Blue curve is control, orange curve is mild hemiparesis, green curve is moderate hemiparesis, red curve is severe hemiparesis.

B. Feature significance analysis

Fig. 10 (a) shows the mean and standard deviation of the features for the four categories of participants. All features are positively correlated except for movement variance and

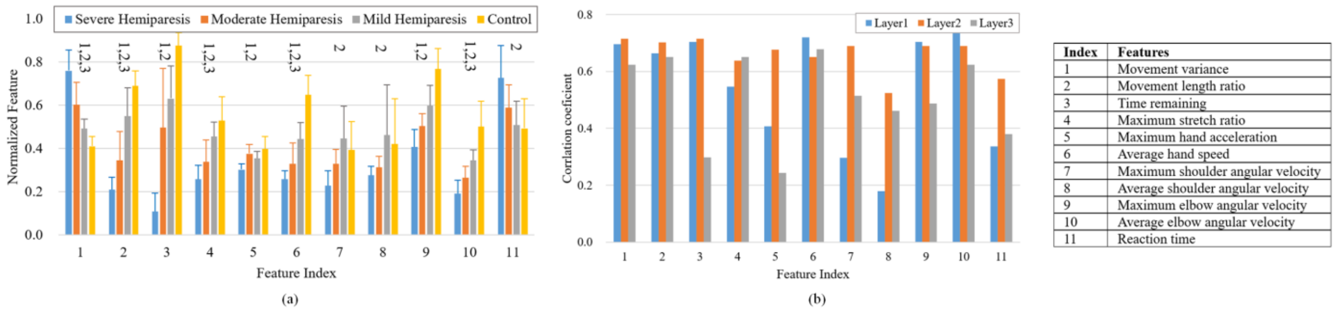


Fig. 10. Significance analysis of features. (a) means and standard deviations of characteristics of participants in different categories and (b) AUC values of characteristics with respect to distinguishing categories.

TABLE II
SUMMARY OF SIGNIFICANT FEATURES (MEAN ± SD) IN RELATION TO THE CORRELATION COEFFICIENT OF THE CORRESPONDING RANK

Index	Severe	Moderate	Mild	Control	Correlation coefficient
1	0.91±0.11	0.72±0.12	0.59±0.05	0.49±0.05	0.69(1), 0.71(2), 0.62(3)
2	0.21±0.05	0.34±0.13	0.55±0.13	0.69±0.06	0.66(1), 0.70(2), 0.65(3)
3	3.27±2.50	14.89±8.23	18.94±4.50	26.28±1.78	0.70(1), 0.71(2), 0.29
4	0.25±0.06	0.33±0.10	0.45±0.06	0.52±0.10	0.54(1), 0.63(2), 0.65(3)
5	0.30±0.02	0.37±0.04	0.35±0.03	0.39±0.05	0.40(1), 0.67(2), 0.24
6	0.02±0.004	0.03±0.009	0.04±0.007	0.06±0.008	0.71(1), 0.65(2), 0.67(3)
7	41.04±12.23	59.17±12.21	80.38±26.90	70.75±23.58	0.29, 0.68(2), 0.51
8	2.76±0.42	3.13±0.50	4.62±2.33	4.21±2.09	0.17, 0.52(2), 0.46
9	73.47±14.41	90.76±10.42	107.77±16.69	138.20±16.84	0.70(1), 0.68(2), 0.48
10	4.77±1.55	6.64±1.30	8.62±1.19	12.53±2.94	0.73(1), 0.68(2), 0.62(3)
11	2.18±0.44	1.76±0.31	1.52±0.33	1.48±0.40	0.33, 0.57(2), 0.37

reaction time, which are negatively correlated with dyskinesia. The values of each feature in the graph are normalized to facilitate comparison and viewing between features. Most features vary monotonically with motor ability. We can observe that for the features in the preparation stage, the differences between the control and hemiparesis groups are large, and all features are statistically significant when distinguishing between the two groups. And the same results were shown between the moderate and severe groups. The results were slightly different when comparing the mild and moderate groups, and all three features were statistically significant except for the remaining time. This indicates that the features of the preparation stage are important at each level of the fuzzy reasoning system. For the features of the competition stage, only the kinematic features of the hand and elbow differed more between the control and hemiparesis groups, suggesting that information about these two joints is important for differentiation in the first layer. Moreover, the difference in each feature was larger between the severe and moderate groups, suggesting that the competition stage is critical for differentiation in the second layer. In contrast, only average hand speed and average elbow angular velocity differed between moderate and mild.

Fig. 10 (b) shows the correlation coefficients for all features in each layer, with the specific values shown in Table II. If a number is marked after the correlation coefficient, it indicates that the feature is statistically significant in that layer. For differentiating the control and hemiplegic groups, it was observed that most of the features in the preparation stage and the kinematic features of the hand and elbow in the competition stage were more discriminative, while the features of extension ability, shoulder kinematic information, and reaction time were less discriminative. In this layer, the highest correlation coefficient was found for average elbow angular velocity (0.73), followed by average hand speed (0.71) and remaining time (0.70), so the discriminatory ability to combine Instructed and spontaneous movements were stronger. For the differentiation {moderate, mild} and severe groups, the discriminative power was found to be stronger for both stage features and higher for the preparation stage in terms of correlation coefficients, *i.e.*, movement variance (0.71), remaining time (0.71) and movement length ratio (0.70), thus indicating that Instructed motion was more effectively discriminated in this layer. For the discrimination of the mild and moderate groups, it was observed that the correlation coefficient values of the features were generally lower than those of the first and second layers, indicating that the two groups of participants were closer and more difficult to distinguish. The top three features in terms of correlation coefficient were average hand speed (0.67), movement length ratio (0.65), and maximum stretch ratio (0.65). Thus, the results in this layer were similar to those in the first layer, combining Instructed and spontaneous movement features for better discrimination.

C. Fuzzy Inference system

After identifying the features as input variables for the fuzzy system, we defined two fuzzy sets for the input variables, namely "Low" and "High". Table III shows the four vertex values of the trapezoidal membership function for each input variable, and the vertex values for each feature were set given the advice of the professional therapist.

After determining the fuzzy sets of the input variables, two fuzzy sets are defined for the output variables, namely "Possible" and "Impossible", and the output values of the two fuzzy sets are the probabilities of the output variables(As shown in Table IV). When the output value of "Possible" of the output variable hemiparesis in FIS1 is 0.64 and the output

TABLE III
FUZZY SET OF INPUT VARIABLES

FIS	Input Variance	Fuzzy Sets	
		Low	High
FIS1	Remaining time	[0, 0, 20, 23]	[20, 23, 30, 30]
	Average hand speed	[0, 0, 0.04, 0.05]	[0.04, 0.05, 0.1, 0.1]
	Average elbow angular speed	[0, 0, 9, 12]	[9, 12, 25, 25]
FIS2	Movement variance	[0, 0, 0.70, 0.85]	[0.70, 0.85, 1, 1]
	Movement length ratio	[0, 0, 0.2, 0.3]	[0.2, 0.3, 1, 1]
	Remaining time	[0, 0, 5, 10]	[5, 10, 30, 30]
FIS3	Movement length ratio	[0, 0, 0.3, 0.6]	[0.3, 0.6, 1, 1]
	Maximum stretch ratio	[0, 0, 0.3, 0.45]	[0.3, 0.45, 1, 1]
	Average hand speed	[0, 0, 0.03, 0.04]	[0.03, 0.04, 0.1, 0.1]

TABLE IV
FUZZY SET OF OUTPUT VARIABLES

FIS	Output Variance	Fuzzy Sets	
		Possible	Impossible
FIS1	Hemiparesis	[0, 0, 1]	[0, 1, 1]
	Control	[0, 1, 1]	[0, 0, 1]
FIS2	Severe	[0, 0, 1]	[0, 1, 1]
	{Moderate, Mild}	[0, 1, 1]	[0, 0, 1]
FIS3	Moderate	[0, 0, 1]	[0, 1, 1]
	Mild	[0, 1, 1]	[0, 0, 1]

value of “Impossible” is 0.36, then the output value of the output variable control has an output value of 0.36 for “Possible” and 0.64 for “Impossible”, and the output probability of “Possible” for the former is greater than that of the latter. The output probability of “Possible” is greater than that of the latter, i.e., it is classified as hemiparesis.

Table VI shows the fuzzy rules of the fuzzy system, all designed with the advice of the therapist. Due to the hierarchical structure of the fuzzy system, the number of fuzzy rules is greatly reduced, where the number of rules for each sub-FIS is 8, and the total number of rules for the fuzzy system is 24. For example, in rule 6, when the participant’s remaining time is “High”, the average hand speed is “Low”, and the average elbow angular speed is “High”, according to the therapist’s suggestion, the participant is at hemiparesis as impossible and at control as possible.

D. Classification Results

The classification results are shown in Table VI. We compared the use of all variables with the use of one stage variable and compared the fuzzy inference system with traditional machine learning classifiers, *i.e.*, SVM, CART, and kNN. The inputs to these classifiers were used with statistically significant features. The FIS using all features showed the best performance among all classifiers, with average accuracy, recall, precision, and F1 scores of 93.5%, 87.2%, 87.2%, and 86.2%, respectively. For the FIS using the preparation stage features, the average accuracy, recall, precision, and F1 scores were 87%, 74.7%, 75.7%, and 74.5%, respectively. For FIS using competition stage features, the average accuracy, recall, precision, and F1 scores were 81.7%, 64%, 65.5%, and 63.7%, respectively. In contrast, SVM, CART, and kNN have poorer performance, with lower accuracy and F1 scores than the fuzzy system.

VI. DISCUSSION

When a stroke causes limb paralysis, it usually leads to weakness in the upper extremities of the patient [41]. A large number of studies have been devoted to the rehabilitation and assessment of patients by means of games [42]. The serious game designed in this study aims to provide a comprehensive evaluation of daily mobility and kinematic information of the upper limbs in patients with hemiparesis. Furthermore, it is designed to be both motivating and clinical. Stroke patients do not need to participate in manual assessment but simply play the game to obtain the assessment results. The system can help hospitals reduce the investment of medical resources and relieve the pressure on physicians. Our study shows that serious games should assess patients from multiple aspects, and the design of the game should take into account the overall limb movement of patients, especially when serious games are used instead of clinical assessment, and should try to cover information of each joint. In addition, our proposed system was designed with reference to the FMA-UE and with the advice of professional therapists to enable our system can classify the degree of upper limb paralysis of participants by means of a game instead of a clinical scale.

According to the trajectory of four types of participants performing the guided movement during the preparation stage (shown in Fig. 8), the control group demonstrated a performance closest to the standard trajectory and were all able to complete the task within a short duration, without exhibiting any arm tremors. In the mild group compared to the control group, although the mild group was eventually able to lift to the end area, there was excess movement, and we observed a slight tremor in the arms of some participants, thus deviating from the standard trajectory, with the green and black curves being the most pronounced. The elapsed time also became significantly more. With increased motor impairment, most participants needed to make multiple attempts, leading to irregular trajectories appeared, and more time was needed to complete the task. When in severe hemiparesis, the range of motion was limited due to upper limb paralysis, which led to a decrease in the ability to extend the arm, which was accompanied by severe tremors and the inability to extend the arm to the prescribed area, which affected their ability to complete the task. As can be seen in Fig. 9, for performing spontaneous movements, the velocity profile of the healthy participants steadily increased, which indicates that their movements were smooth and without any jitter. In contrast, the velocity curve of the hemiparesis group was zigzagged, which indicates that the hand produced tremors during the movement toward the target, and the movement was poorly completed. In addition, participants in the severe group were unable to move their hands quickly, which is consistent with their clinical performance. This demonstrates the soundness of the game design in that not only were patients

TABLE V
FUZZY RULES

FIS	Rules	Input Variance			Output Variance	
		Remaining time	Average hand speed	Average elbow angular speed	Hemiparesis	Control
FIS1	Rule1	Low	Low	Low	Possible	Impossible
	Rule2	Low	Low	High	Possible	Impossible
	Rule3	Low	High	Low	Possible	Impossible
	Rule4	Low	High	High	Impossible	Possible
	Rule5	High	Low	Low	Possible	Impossible
	Rule6	High	Low	High	Impossible	Possible
	Rule7	High	High	Low	Impossible	Possible
	Rule8	High	High	High	Impossible	Possible
		Movement variance	Movement length ratio	Remaining time	Severe	{Moderate, Mild}
FIS2	Rule9	High	Low	Low	Possible	Impossible
	Rule10	High	Low	High	Possible	Impossible
	Rule11	High	High	Low	Possible	Impossible
	Rule12	High	High	High	Impossible	Possible
	Rule13	Low	Low	Low	Possible	Impossible
	Rule14	Low	Low	High	Impossible	Possible
	Rule15	Low	High	Low	Impossible	Possible
	Rule16	Low	High	High	Impossible	Possible
		Movement length ratio	Maximum stretch ratio	Average hand speed	Moderate	Mild
FIS3	Rule17	Low	Low	Low	Possible	Impossible
	Rule18	Low	Low	High	Possible	Impossible
	Rule19	Low	High	Low	Possible	Impossible
	Rule20	Low	High	High	Impossible	Possible
	Rule21	High	Low	Low	Possible	Impossible
	Rule22	High	Low	High	Impossible	Possible
	Rule23	High	High	Low	Impossible	Possible
	Rule24	High	High	High	Impossible	Possible

TABLE VI
CLASSIFICATION PERFORMANCE WITH LEAVE-ONE-SUBJECT-OUT CROSS-VALIDATION

Classifier	Classes	All Features				Preparation Stage Features				Competition Stage Features			
		Accuracy	Recall	Precision	F1 Score	Accuracy	Recall	Precision	F1 Score	Accuracy	Recall	Precision	F1 Score
SVM	Severe	0.80	0.42	0.68	0.51	0.79	0.37	0.65	0.47	0.80	0.50	0.62	0.55
	Moderate	0.68	0.70	0.42	0.52	0.59	0.50	0.30	0.37	0.62	0.42	0.31	0.35
	Mild	0.71	0.42	0.43	0.42	0.69	0.40	0.39	0.39	0.66	0.37	0.34	0.35
	Control	0.81	0.50	0.66	0.56	0.74	0.37	0.48	0.41	0.78	0.45	0.60	0.51
CART	Severe	0.83	0.52	0.75	0.61	0.80	0.42	0.68	0.51	0.76	0.47	0.52	0.49
	Moderate	0.71	0.65	0.45	0.53	0.65	0.50	0.36	0.41	0.59	0.30	0.24	0.26
	Mild	0.78	0.55	0.57	0.55	0.73	0.55	0.46	0.50	0.73	0.55	0.46	0.50
	Control	0.84	0.57	0.74	0.64	0.80	0.52	0.63	0.56	0.67	0.32	0.34	0.32
kNN	Severe	0.86	0.62	0.78	0.69	0.82	0.61	0.72	0.66	0.78	0.42	0.60	0.49
	Moderate	0.86	0.72	0.72	0.72	0.65	0.58	0.52	0.54	0.73	0.50	0.46	0.47
	Mild	0.80	0.72	0.59	0.64	0.75	0.62	0.57	0.59	0.66	0.47	0.36	0.40
	Control	0.91	0.82	0.84	0.82	0.85	0.73	0.76	0.74	0.81	0.60	0.64	0.61
FIS	Severe	0.93	0.85	0.87	0.85	0.88	0.70	0.80	0.74	0.79	0.50	0.60	0.54
	Moderate	0.94	0.87	0.89	0.87	0.86	0.75	0.73	0.73	0.85	0.67	0.71	0.68
	Mild	0.89	0.82	0.76	0.78	0.81	0.72	0.61	0.66	0.71	0.57	0.45	0.50
	Control	0.98	0.95	0.97	0.95	0.93	0.82	0.89	0.85	0.92	0.82	0.86	0.83

of all levels able to participate in the game but there was a strong differentiation in the performance of participants within the game.

Based on the feature analysis presented in Fig. 10 (a), the control and hemiparesis groups were effectively differentiated with significant features in both the preparation and competition stages and with higher correlation coefficients in layer 1 than in layer 2 and layer 3 (as shown in Fig. 10 (b)). Moreover, it was observed that during spontaneous movements in the competition stage, the hand and elbow kinematic features in the control group were significantly different from the hemiparesis group, which is consistent with the findings of Eva et al. [43], where normal individuals were able to produce more elbow movements during the exercise. And for distinguishing the {moderate, mild} group from the severe group, we found that the features of the preparation stage were more meaningful. This may be due to the fact that severe hemiparesis is more hampered in performing guided movements compared to spontaneous movements. We also found that the reaction time was significantly greater in the severe group compared to the other groups, which may be due to severe damage to the motor areas of the cerebral cortex, resulting in an inability to move the affected limb in a short period of time [44]. For differentiating between the mild and moderate groups, better discrimination was observed with the features of the preparation stage and the kinematic features of the hand and elbow. This suggests that the mild group not only performed better in daily activities but also had more flexible hand movements and was able to generate more movements in the elbow.

From analysis of the above features, it can be observed that kinematic information of the joints plays a crucial role in differentiating the categories due to the fact that our system quantifies elbow and shoulder movements of the patients during spontaneous movements, such as swings during the competition stage. For patients with hemiparesis, limb paralysis leads to differences in movement below the shoulder or elbow, which in turn affects the speed of hand movement. Therefore, we propose that it is reasonable to consider the hand-elbow-shoulder as a motion whole due to the analysis of each joint to avoid biases arising from non-upper limb movements.

As shown in Table VI, the fuzzy inference system designed based on the therapist's experience achieved the highest mean accuracy with better recall, precision, and F1 score. This outcome indicates that this classifier is unbiased. The highest accuracy was observed in the control and severe groups, indicating better separability between these two categories. This result is consistent with the clinical fact that normal individuals have significantly better motor abilities than hemiplegic patients, and severe hemiplegic patients have almost no motor abilities. In contrast, the accuracy of identifying mild and moderate hemiparesis was lower. We believe the reason for this is that patients with mild and moderate hemiparesis perform similarly in games, and using only somatosensory devices fails to collect fine data, such as finger grip strength, which can lead to biased classification results. Furthermore, we observed higher performance of classifiers designed using features from both the preparation stage and the competition stage. This finding justifies the design of our game, which was able to improve the classification performance by combining daily activity ability and kinematic information. Moreover, the classification performance of machine learning classifiers is not as good as that of fuzzy inference systems. This problem is caused by the small sample size of the data, which hinders the classifier from learning the correct decision boundaries [45]. Therefore, machine learning methods may not be applicable to small sample datasets.

We conducted this work to address three issues that currently exist with serious game-based assessments. Firstly, we proposed a new serious game design approach to alleviate patients' tension, loss, and irritation during gameplay. Specifically, we divided our game into two stages, a preparation stage and a competition stage. During the preparation stage, no intensive requirements are imposed, providing sufficient relaxation time for the patient's body and mind. This stage also assesses the patient's upper limb daily life ability. Secondly, we addressed the problem of missing joints in most current games by facilitating patients to perform spontaneous movements through the ping-pong game during the game stage. We used the kinematic information collected during this period on the hand, elbow, and shoulder joints to comprehensively evaluate the patients' upper limb motor functions, thereby improving the accuracy of the assessment results. Finally, we developed a classification model using a fuzzy inference method that incorporated the experience of rehabilitation therapists. The proposed model resolves labeling difficulties and overfitting issues generated by model training. The resulting classification accuracy was 93.5%, with clinical reference value. In addition, studies revealed some effective kinematic features, such as average hand speed, maximum elbow angular velocity, and average elbow angular velocity. Further investigations showed that these kinematic features were significantly associated with the extent of motor nerve damage in patients [46], particularly the elbow angular velocity. This finding provides a priori knowledge for other researchers to design automated models for the identification of upper limb hemiparesis in stroke patients. Moreover, patient's multi-round exercise data showed a significant improvement in upper limb joint motion after all 3 rounds of play, with the instantaneous motion speed of the hand, elbow, and shoulder improved by 15%, 17%, and 9%, respectively. The proposed serious game method may further improve the patients' motor ability and have both recognition and rehabilitation therapeutic effects. This finding encourages further investigation of the effect of serious games in motor rehabilitation.

A. Limitations

Although the results of our proposed assessment system are promising, some limitations should be taken into account. Firstly, the system is unable to detect the grip strength of the patient's hand during the game. This may lead to inaccurate assessment results for some patients with poor grip function but otherwise good function. To address this limitation, lightweight hardware could be added in the future to assist in the assessment. Secondly, the serious game system is currently limited to the assessment of the level of upper limb motor impairment. It remains unclear whether the system is effective for intervention training of patients' upper limbs, whether it can be applied to rehabilitation training in the future, and whether it can improve patients' upper limb motor functions.

VII. Conclusion

In this paper, a somatosensory upper extremity serious game assessment system was designed to assess the grade of upper extremity hemiparesis in patients with post-stroke hemiparesis. Tests on 32 participants showed that the game assessment system was able to distinguish well between different levels of paralysis, with an accuracy of 93.5% compared to the therapist's manual scale assessment.

Ideally, when assessing and training patients using serious games, consideration should be given to their emotional state. We design different game stages specifically for patients with dysphoric moods. In our future work, we aim to investigate the potential benefits of emotion facilitation in serious games. Leveraging the plasticity of the virtual environment to induce relaxation in patients may improve their performance or speed up the recovery process.

References

[1] C. M. Stinear, C. E. Lang, S. Zeiler, and W. D. Byblow, "Advances and challenges in stroke rehabilitation," *Lancet Neurol.*, vol. 19, no. 4, pp. 348–360, Apr. 2020.

[2] S. S. Virani, A. Alonso, H. J. Aparicio, E. J. Benjamin, and M. S. Bittencourt, "Heart disease and stroke statistics—2021 update: A report from the American Heart Association," *Circulation*, vol. 143, no. 8, pp. 254–743, 2021.

[3] R. Bonita and R. Beaglehole, "Recovery of motor function after stroke.," *Stroke*, vol. 19, no. 12, pp. 1497–1500, Dec. 1988.

[4] D. J. Gladstone, C. J. Danells, and S. E. Black, "The Fugl–Meyer assessment of motor recovery after stroke: A critical review of its measurement properties," *Neurorehabilitation Neural Repair*, vol. 16, no. 3, pp. 232–240, Sep. 2002.

[5] Centers for Disease Control and Prevention, "Outpatient rehabilitation among stroke survivors—21 states and the district of Columbia, 2005," *MMWR. Morbidity Mortality Weekly Rep.*, vol. 56, no. 20, pp. 504–507, 2007.

[6] E.-L. Bustrén, K. S. Sunnerhagen, and M. A. Murphy, "Movement kinematics of the ipsilesional upper extremity in persons with moderate or mild stroke," *Neurorehabilitation Neural Repair*, vol. 31, no. 4, pp. 376–386, Apr. 2017.

[7] F. Noveletto et al., "Biomedical serious game system for balance rehabilitation of hemiparetic stroke patients," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2179–2188, Nov. 2018.

[8] B. Ferreira, J. Lourenço, and P. Menezes, "A serious game for post-stroke motor rehabilitation," in *Proc. 5th Exp. Int. Conf.*, Jun. 2019, pp. 383–387.

[9] M. Ma and K. Bechkoum, "Serious games for movement therapy after stroke," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2008, pp. 1872–1877.

[10] E. D. Oña, C. Balaguer, and A. Jardón, "Towards a framework for rehabilitation and assessment of upper limb motor function based on serious games," in *Proc. IEEE 6th Int. Conf. Serious Games Appl. Health (SeGAH)*, May 2018, pp. 1–7.

[11] D. Drummond, A. Hadchouel, and A. Tesnière, "Serious games for health: Three steps forwards," *Adv. Simul.*, vol. 2, no. 1, pp. 1–8, Dec. 2017.

[12] D. Webster and O. Celik, "Systematic review of Kinect applications in elderly care and stroke rehabilitation," *J. Neuroeng. Rehabil.*, vol. 11, no. 108, pp. 1–24, Jul. 2014.

[13] N. Hussain, M. A. Murphy, and K. S. Sunnerhagen, "Upper limb kinematics in stroke and healthy controls using target-to-target task in virtual reality," *Frontiers Neurol.*, vol. 9, p. 300, May 2018.

[14] G. Kwakkel et al., "Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the stroke recovery and rehabilitation roundtable," *Int. J. Stroke*, vol. 12, no. 5, pp. 451–461, Jul. 2017.

[15] M. A. Murphy and C. Häger, "What has been gained from recent advances in kinematic analyses of the upper extremity after stroke?" *Physiotherapy*, vol. 101, p. e65, May 2015.

[16] J. Bai and A. Song, "Development of a novel home based multi-scene upper limb rehabilitation training and evaluation system for post-stroke patients," *IEEE Access*, vol. 7, pp. 9667–9677, 2019.

[17] E. D. Oña, A. Jardón, and C. Balaguer, "Automatic assessment of arm motor function and postural stability in virtual scenarios: Towards a virtual version of the Fugl–Meyer test," in *Proc. IEEE 8th Int. Conf. Serious Games Appl. Health (SeGAH)*, Aug. 2020, pp. 1–6.

[18] S. Lee, J. Cui, L. Liu, M. Su, L. Zheng, and S. Yeh, "An evidence-based intelligent method for upper-limb motor assessment via a VR training system on stroke rehabilitation," *IEEE Access*, vol. 9, pp. 65871–65881, 2021.

[19] R. W. Teasell and L. Kalra, "What's new in stroke rehabilitation," *Stroke*, vol. 35, no. 2, pp. 383–385, Feb. 2004.

[20] M. R. G. Robinson, "Neuropsychiatric consequences of stroke," *Annu. Rev. Med.*, vol. 48, no. 1, pp. 217–229, Feb. 1997.

[21] B. Singer and J. Garcia-Vega, "The Fugl–Meyer upper extremity scale," *J. Physiotherapy*, vol. 63, no. 1, p. 53, Jan. 2017.

[22] S. Lee, Y. Lee, and J. Kim, "Automated evaluation of upper-limb motor function impairment using Fugl–Meyer assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 125–134, Jan. 2018.

[23] C. Wang, L. Peng, Z. Hou, J. Li, T. Zhang, and J. Zhao, "Quantitative assessment of upper-limb motor function for post-stroke rehabilitation based on motor synergy analysis and multi-modality fusion," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 943–952, Apr. 2020.

[24] C. Bosecker, L. Dipietro, B. Volpe, and H. I. Krebs, "Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke," *Neurorehabilitation Neural Repair*, vol. 24, no. 1, pp. 62–69, Jan. 2010.

[25] M. Zhang et al., "Quantitative evaluation system of upper limb motor function of stroke patients based on desktop rehabilitation robot," *Sensors*, vol. 22, no. 3, p. 1170, Feb. 2022.

[26] E. V. Olesh, S. Yakovenko, and V. Gritsenko, "Automated assessment of upper extremity movement impairment due to stroke," *PLoS ONE*, vol. 9, no. 8, Aug. 2014, Art. no. e104487.

[27] J. Serradilla, J. Q. Shi, Y. Cheng, G. Morgan, C. Lambden, and J. A. Eyre, "Automatic assessment of upper limb function during play of the action video game, circus challenge: Validity and sensitivity to change," in *Proc. IEEE 3rd Int. Conf. Serious Games Appl. Health (SeGAH)*, May 2014, pp. 1–7.

[28] S. Cho, W. Kim, N. Paik, and H. Bang, "Upper-limb function assessment using VBBTs for stroke patients," *IEEE Comput. Graph. Appl.*, vol. 36, no. 1, pp. 70–78, Jan. 2016.

[29] W. Pei, G. Xu, M. Li, H. Ding, S. Zhang, and A. Luo, "A motion rehabilitation self-training and evaluation system using Kinect," in *Proc. 13th Int. Conf. Ubiquitous Robots Ambient Intell. (URAI)*, Aug. 2016, pp. 353–357.

[30] Y. Saeys, T. Abeel, and Y. Van De Peer, "Robust feature selection using ensemble feature selection techniques," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2008, pp. 313–325.

[31] J. Dombi, "Membership function as an evaluation," *Fuzzy Sets Syst.*, vol. 35, no. 1, pp. 1–21, Mar. 1990.

[32] A. Sadollah, "Introductory chapter: Which membership function is appropriate in fuzzy system?" in *Fuzzy Logic Based in Optimization Methods and Control Systems and Its Applications*. London, U.K.: Intechopen,2018, pp. 3–6.

[33] G. V. S. Raju, J. Zhou, and R. A. Kisner, "Hierarchical fuzzy control," *Int. J. Control*, vol. 54, no. 5, pp. 1201–1216, 1991.

[34] T. Siddique, R. Fareh, M. Abdallah, Z. Ahmed, and M. H. Rahman, "Autonomous exercise generator for upper extremity rehabilitation: A fuzzy-logic-based approach," *Micromachines*, vol. 13, no. 6, p. 842, May 2022.

[35] N. Xiong and L. Litz, "Reduction of fuzzy control rules by means of premise learning—Method and case study," *Fuzzy Sets Syst.*, vol. 132, no. 2, pp. 217–231, Dec. 2002.

[36] E. Pourjavad and A. Shahin, "The application of Mamdani fuzzy inference system in evaluating green supply chain management performance," *Int. J. Fuzzy Syst.*, vol. 20, no. 3, pp. 901–912, Mar. 2018.

[37] S. Kundu, "The min-max composition rule and its superiority over the usual max-min composition rule," *Fuzzy Sets Syst.*, vol. 93, no. 3, pp. 319–329, Feb. 1998.

[38] S. Chakraverty, D. M. Sahoo, and N. R. Mahato, *Concepts of Soft Computing*. Berlin, Germany: Springer, 2019.

[39] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *Proc. Eur. Conf. Inf. Retr.*, 2005, pp. 345–359.

[40] T.-T. Wong, "Performance evaluation of classification algorithms by K-fold and leave-one-out cross validation," *Pattern Recognit.*, vol. 48, no. 9, pp. 2839–2846, Sep. 2015.

[41] H. Nakayama, H. S. Jørgensen, H. O. Raaschou, and T. S. Olsen, "Recovery of upper extremity function in stroke patients: The Copenhagen stroke study," *Arch. Phys. Med. Rehabil.*, vol. 75, no. 4, pp. 394–398, Apr. 1994.

[42] I. Doumas, G. Everard, S. Dehem, and T. Lejeune, "Serious games for upper limb rehabilitation after stroke: A meta-analysis," *J. NeuroEng. Rehabil.*, vol. 18, no. 1, pp. 1–16, Dec. 2021.

[43] M. A. Murphy, C. Willén, and K. S. Sunnerhagen, "Kinematic variables quantifying upper-extremity performance after stroke during reaching and drinking from a glass," *Neurorehabilitation Neural Repair*, vol. 25, no. 1, pp. 71–80, Jan. 2011.

[44] S. C. Cramer and K. R. Crafton, "Somatotopy and movement representation sites following cortical stroke," *Exp. Brain Res.*, vol. 168, nos. 1–2, pp. 25–32, Jan. 2006.

[45] T. Dietterich, "Overfitting and undercomputing in machine learning," *ACM Comput. Surv.*, vol. 27, no. 3, pp. 326–327, Sep. 1995.

[46] G. Thrane, K. S. Sunnerhagen, and M. A. Murphy, "Upper limb kinematics during the first year after stroke: The stroke arm longitudinal study at the university of Gothenburg (SALGOT)," *J. NeuroEng. Rehabil.*, vol. 17, no. 1, pp. 1–12, Dec. 2020.