An Intelligent Motor Assessment Method Utilizing a Bi-Lateral Virtual-Reality Task for Stroke Rehabilitation on Upper Extremity

**CHIA-RU CHUNG[1], MU-CHUN SU[1], SI-HUEI LEE[2],
ERIC HSIAO-KUANG WU[1], LI-HSIEN TANG[1],
AND SHIH-CHING YEH[1]**

[1]Department of Computer Science and Information Engineering, National Central University, Taoyuan 320, Taiwan
[2]Department of Physical Medicine and Rehabilitation, Taipei Veterans General Hospital, National Yang-Ming University, Taipei 11221, Taiwan
CORRESPONDING AUTHOR: ERIC HSIAO-KUANG WU (hsiao@csie.ncu.edu.tw)

ABSTRACT   Virtual reality (VR) has been widely adopted by therapists to provide rich motor training tasks. Time series data of motion trajectory accompanied with the interaction of VR system may contain important clues in regard to the assessment of motor function, however, clinical evaluation scales such as Fugl-Meyer Assessment (FMA), Wolf Motor Function Test (WMFT), and Test D'évaluation Des Membres Supérieurs Des Personnes Âgées (TEMPA) are highly depended in clinic. Further, there is not yet an assessment method that simultaneously consider motion trajectory and clinical evaluation scales. The objective of this study is to establish an evidence-based assessment model by machine-learning method that integrated motion trajectory of a VR task with clinical evaluation scales. In this study, a VR system for upper-limb motor training was proposed for stroke rehabilitation. Clinical trials with 20 stroke patients were performed. A variety of motor indicators that derived via motion trajectory were proposed. The correlations between motor indicators and clinical evaluation scales were examined. Further, motor indicators were integrated with evaluation scales to develop a machine-learning based model that represents an evidence-based motor assessment approach. Clinical evaluation scales, FMA, TEMPA and WMFT, were significantly progressed. A few motor indicators were found significantly correlated with clinical evaluation scales. The accuracy of machine-learning based assessment model was up to 86%. The proposed VR system is validated to be effective in motor rehabilitation. Motor indicators derived from motor trajectory were with potential for clinical motor assessment. Machine learning could be a promising tool to perform automatic assessment.

INDEX TERMS   Stroke rehabilitation, motor training, virtual reality, machine learning.

Clinical and Translational Impact Statement—A VR task for motor rehabilitation was exanimated via clinical trials. Integrating motor indices with clinical assessment, a machine-learning model with accuracy of 86% was developed to evaluate motor function.

I. INTRODUCTION

Stroke is a leading cause of death in low- and middle- income countries. Globally, 70% of strokes and 87% of both stroke-related deaths and disability-adjusted life years occur in these developing countries [1]. Stroke causes brain death due to poor blood flow into cells inside the brain [2].

In traditional stroke rehabilitation, relevant nerve tissue is stimulated via continuous motor training to restore as much motor function as possible [3], [4], [5]. However, long-term traditional physical rehabilitation tasks become routine and monotonous, leading to a lack of motivation in patients. To assess motor function, doctors and therapists often use traditional evaluation scales like Fugl-Meyer Assessment (FMA) [6], Wolf Motor Function Test (WMFT) [7], and Test D'évaluation Des Membres Supérieurs Des Personnes Âgées (TEMPA) [8]. However, traditional evaluation scales are subjective and a therapist is required for each assessment. This would become an issue if therapists were limited [9].

The goal of interactive training tasks is to increase motivation and concentration in patients. We could achieve this goal through VR. Virtual reality is a computer-based technology that allows users to interact with simulated environments [10]. Virtual reality is more accessible and affordable in motor rehabilitation, increasing patient motivation to perform repetitive rehabilitation tasks [11]. Therefore, much research has focused on VR-based motor rehabilitation [12], [13], [14]. There is also much research demonstrating that VR could lead to significant improvement in stroke rehabilitation [15], [16], [17], [18]. Furthermore, VR could help patients who are un- willing to perform stroke rehabilitation tasks [19]. Moreover, the patient's training goals could be adjusted according to their physical conditions [20].

Virtual-reality stroke-rehabilitation systems are also able to collect a huge amount of sensing data (motion trajectory, electromyography) via sensors [21], [22], [23], [24]. In most research, VR stroke-rehabilitation systems are only used for motor training, and patient skeleton and electromyography data are collected by the sensors. However, because therapists mainly score evaluation scales, this research did not utilize the physiological data to evaluate the patient's condition. Although, some research [25], [26], [27] has analyzed the patient's physiological data collected from sensors in order to implement motor assessment functions and help patients to understand their rehabilitation progress. However, research has not discussed the correlation between proposed motor indicators and traditional evaluation scales.

Artificial intelligence and machine learning have developed rapidly in recent years and are widely used in data mining, computer vision, natural language processing, speech recognition, etc. In the medical domain, some researchers use the clustering method (K-means) to cluster impairment levels [28]. Some researchers use supervised machine learning algorithms to classify a patient's status [29], [30]. However, due to a lack of correlation with traditional evaluation scales and clinical references, therapists are unconvinced of its effectiveness and capabilities.

To address the issues mentioned above, we wanted to build an evidence-based motor assessment. Our research designed motor indicators with upper-limb data collected by sensors and analyzed the correlation between motor indicators and evaluation scales. Furthermore, we applied machine learning methods to establish an assessment model based on evidence-based motor indicators.

II. RELATED WORK

In recent years, many researchers have adopted VR training tasks in stroke rehabilitation. Some research has proved the efficiency of VR rehabilitation by comparing pretest and posttest data. Sheehy et al. [31] verified the feasibility of using home-based VR post stroke via a parallel randomized feasibility trial. Aramaki et al. [32] verified that VR is a viable tool for rehabilitating stroke patients with the Canadian Occupational Performance Measure (COPM) and the Participation Scale. Lee et al. [33] verified game-based VR canoe paddling training as an effective rehabilitation therapy using a modified functional reach test (mFRT) and a manual function test (MFT). Choi and Paik [34] proposed a mobile VR upper-extremity rehabilitation program and verified its effects using FMA-UE, B-stage, and manual muscle testing. Although this research has shown the efficiency of VR rehabilitation tasks, physiological data collected from the sensors was not sed to assess the stroke patient's status. Therefore, therapists were unable to assess a stroke patient after using VR rehabilitation tasks.

Recently, some research has extracted features from sensor data during rehabilitation tasks to assess stroke patients. Phienphanich et al. [25], Lu et al. [26], Lee et al. [27] and Kashi et al. [35] extracted movement features based on biomechanical metrics, including velocity, jerk, index of curvature, and angles of the joints. These movement features considered range of motion, smoothness, and compensation [27]. However, this research did not consider the correlation with traditional evaluation scales (such as FMA, WMFT, or TEMPA). Lee et al. [36] extracted features based on the linguistic guideline of the FMA. However, many other traditional evaluation scales are used to assess a stroke patient's impairment level. Considering features from the FMA only might not be enough to assess a stroke patient.

Some research has proposed a classification method according to movement features to classify a patient's impairment level. Biswas et al. [37] used the regularized Mahalonobis distance-based K-means clustering to classify elementary arm movements to different levels of impairment. Miao et al. [28] presented a DTW-KNN joint algorithm to classify multiple training completion levels. Kashi et al. [35] used the RAndom k-labELsets (RAkEL) algorithm via a random forest as the base classifier to establish a multi-label classification model based on stroke patient data. Each of these studies use movement features to implement clustering. However, their clustering methods did not consider traditional evaluation scales. Evaluation scales could provide a reliable, convincible analysis aspect to doctors, therapists, and patients. Our re- search proposes a classification method based on traditional evaluation scales.

Based on the movement data of VR rehabilitation tasks, some research proposed motor assessment methods to provide therapists and patients with a quantitative result to verify improvements of the stroke patient. Wang et al. [38] proposed a system where the patient can complete treatment during a rehabilitation game. Nevertheless, the quantitative results of this paper's rehabilitation system were not validated using traditional evaluation scales. Liao et al. [39] proposed scoring functions for the automated assessment of the quality of physical rehabilitation tasks. However, these scoring functions also lacked correlation with traditional evaluation scales.

Our study presents VR stroke-rehabilitation tasks containing a Kinect sensor to collect upper-limb movement data. A bi-lateral VR task, including the ball-throwing and ball-catching with two arms back and forth, is designed to train weight-shifting and eye-hand coordination. We propose

an evidence-based analysis approach that contains upper-limb movement features, a multi-classification method, and a machine learning approach. We extracted features from upper-limb movement data collected from Kinect and used statistics analysis to verify the correlation between upper-limb movement features and traditional evaluation scales. For the multi-classification method, we utilized the K-means clustering algorithm [37] to divide patient impairment levels into three categories based on the traditional evaluation scales. In comparison to previous studies, our VR stroke-rehabilitation tasks have higher reliability and interpretable ability.

III. METHOD

A. VIRTUAL REALITY SYSTEM DESIGN

1) SYSTEM INTRODUCTION

Our VR stroke-rehabilitation system combines Kinect [40], 3D VISION stereo glasses, a 3D display projector, a 3D display card, and other hardware devices. It aims to provide patients with rehabilitation tasks for their upper limbs. It includes a bi-lateral VR task for upper-limb extension, balance, and hand-eye coordination with the Unity 3D engine. The physical setting of our VR stroke-rehabilitation system is shown in Fig. 1.
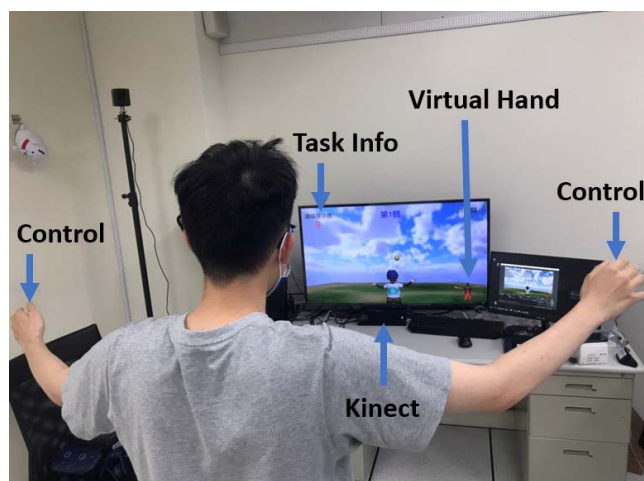

FIGURE 1. Physical setting of our VR stroke-rehabilitation system which combines Kinect [40], 3D VISION stereo glasses, a 3D display projector, a 3D display card, and other hardware devices. A bi-lateral VR task for upper-limb extension, balance, and hand-eye coordination with the Unity 3D engine were included in this system.

2) TASK CONTENT

To train the extension and muscle function of the patient's arms, the VR task in our system include both throwing and catching a ball with both arms. The participate is asked to stretch both arms to catch the ball continuously, which is moving in a parabola. Kinect was applied to perform motion capture. Before starting the task, therapists set the difficulty of the task that includes the speed of the ball, the range of the flying, and the number of balls to catch. Further description

for the difficulty of the task would be illustrated in the coming section. At the beginning of the task, a ball appeared at starting location which was at the left side or right side upon the setting, as shown in Fig. 2. The user had to touch the ball with the hand closer to it, triggering the ball to move along the parabola to the other side, and catch the ball with another hand back and forth. Meanwhile, the user can see his/her current performance including the number of successful ball-catching movements, the number of failed ball-catching movements, and the number of consecutive successful ball-catching movements.
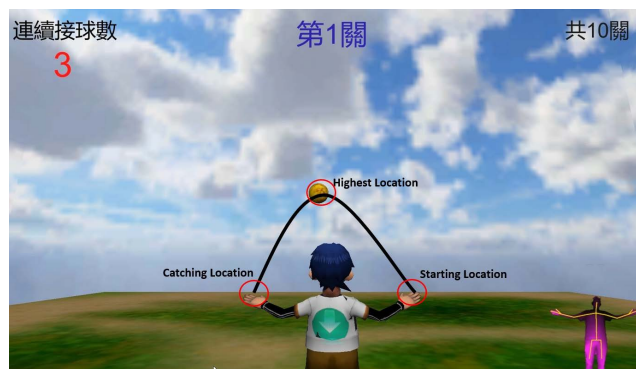

FIGURE 2. Illustration of bi-lateral VR task. At the beginning of the task, a ball appeared at Starting Location which was at the left side or right side upon the setting. The user had to touch the ball with the hand closer to it, triggering the ball to move along the parabola to the other side, and catch the ball with another hand back and forth.

3) DIFFICULTY DESIGN MECHANISM

The difficulty of the bi-lateral VR task can be set using the following parameters.

- The horizontal distance of the sphere falling: the distance can be set according to the patient's range of upper-limb extension. The bi-lateral VR task provides various range options to suit the patient's upper-limb abilities: 30% to 50%, 30% to 75% and 30% to 100% of the patient's arm when his/her arm was totally strengthened.
- The number of successful catches: therapists can adjust the number to test how long the patient could continue exercising.
- The speed of the sphere falling: speed is in units of gravitational acceleration (G). Therapists could set this to 1/9G, 2/9G, and 3/9G so that patients could train their reaction time and hand-eye coordination according to their impairment level.

B. FINAL STAGE

After many discussions with therapists, we determined what therapists want most in regards to patients' movement quality: reaction time, hand-eye coordination quality, range of upper limb movement, a comparison between the affected arm and the normal arm, and the ball's speed difference between front and rear at the highest point. Note that the range of upper limb movement for the affected and the normal

arm might be different. Evaluating their difference would be critical. The importance of ball speed is the patient's preparation time and reaction time. When a patient's arm is injured, his/her muscle would have difficulty controlling which could be evaluated by different ball speeds. In other words, different ball speeds would affect how long patients take to raise their hands and prepare. Assume that after the right hand touches the ball, the left hand must be raised to the target position before landing. It is not an easy task for the patients whose arm is injured. We designed motor indicators in accordance with therapists' needs, and the schematic diagram is shown in Fig. 2. Features 1 through 6 are related to horizontal and vertical motor ability based on different axes in the upper-limb's skeleton data. Features 7 through 19 are related to the motor ability of the whole body. The major reason for differentiating the velocity before of the arm when the ball is at the highest location or after was that when the ball is at the highest point, the user should know that the landing point of the ball, and then moves his/her arm. Hence, evaluating the spending time could estimate the patient's reaction ability.

1) FIRST TIME ENTERING THE CATCHING RANGE
Record the time when the patient first moves their arm in the catching range

2) LAST TIME ENTERING THE CATCHING RANGE
Record the time when the patient last moves their arm in the catching range.

3) CATCHING TIME
Record the time the patient successfully catches each ball.

4) AIMING TIME
Record the time when the patient's hand moves into the scope of the ball-catching range after the ball arrives at its highest location. A lower value means the patient has a faster reaction time.

5) AMOUNT OF TIME BETWEEN STABILIZATION AND CATCHING BALL
If a patient receives a higher difference in time than before, this means the patient has a higher degree of control, and the hand tremor is improving. Stabilization indicates the degree of stability of the muscle, which means that whether this movement of the hand trajectory is smooth.

6) NUMBER OF TIMES THE PATIENT GETS IN AND OUT OF CATCHING RANGE
If the number of times decreases, then the patient's upper-limb function is recovering.

7) TIME DIFFERENCE CATCHING THE BALL FROM THE NORMAL ARM TO THE AFFECTED ARM
Observing the patient's arm reactions before the ball arrives at its highest location.

8) BALL-CATCHING TIME COST BY AFFECTED ARM
The total time spent touching (catching) the ball with the affected arm during the game. A longer time cost means the patient has a better reaction ability.

9) THE TOTAL LENGTH THAT THE AFFECTED ARM JOINT POINTS MOVED DURING TASK
Observe the patient's extension ability of the affected arm.

10) MOVEMENT DIFFERENCE BETWEEN THE AFFECTED ARM AND THE NORMAL ARM
the calculation formula is as shown in

$$\frac{\textit{The movement distance by normal arm}}{\textit{The movement distance by affected arm}} \tag{1}$$

11) MAXIMUM SPEED BY THE AFFECTED ARM BEFORE THE BALL ARRIVES AT THE HIGHEST LOCATION
the calculation formula is as shown in

$$V_{MAX} = MAX\left(V_i = \frac{\Delta D_i}{\Delta t_i}\right) \tag{2}$$

where V_{MAX} is the maximum instantaneous speed; $\triangle D_i$ is the instantaneous displacement distance; $\triangle t_i$ is the instantaneous time difference.

12) AVERAGE SPEED BY AFFECTED ARM BEFORE BALL ARRIVES AT THE HIGHEST LOCATION
the calculation formula is as shown in

$$\bar{v} = \frac{1}{N}\left(\sum\nolimits_{i=1}^{N} V_i\right) \tag{3}$$

where $\bar{v}$ is the average speed; v_i is the instantaneous speed; N is the total number of seconds of training.

13) V-VARIATION BY AFFECTED ARM BEFORE BALL ARRIVES AT THE HIGHEST LOCATION
the calculation formula is as shown in

$$\sigma = \sqrt{\frac{1}{N}\left(\sum\nolimits_{i=1}^{N}(v_i - \bar{v})^2\right)} \tag{4}$$

where σ is the V-variation; V_i is the instantaneous speed; $\bar{v}$ is the average speed; The V-variation reflects the degree of speed dispersion.

14) MAXIMUM SPEED BY THE AFFECTED ARM AFTER THE BALL ARRIVED AT THE HIGHEST LOCATION
Observe the patient's maximum movement ability. The calculation formula is the same as in (2).

15) AVERAGE SPEED OF THE AFFECTED ARM AFTER THE BALL ARRIVES AT THE HIGHEST LOCATION
Observe the patient's movement variance trend. The calculation formula is the same as in (3).

16) V-VARIATION BY AFFECTED ARM AFTER THE BALL ARRIVES AT THE HIGHEST LOCATION

Observe the patient's motor ability. If the V-variation is higher after the ball arrives at the highest location, then the patient has better motor ability. The calculation formula is the same as in (4).

17) MAXIMUM HORIZONTAL EXTENSION DISTANCE OF ARM MOVEMENT

Observe the maximum horizontal range of the patient's arm movement.

18) MAXIMUM VERTICAL EXTENSION DISTANCE OF ARM MOVEMENT

Observe the maximum vertical range of the patient's arm movement.

19) MAXIMUM EXTENSION DISTANCE FROM ARM MOVEMENT IN THE FRONT-VIEW DIRECTION

Observe the maximum range of patient's arm movement.

C. CLINICAL TRIALS

1) PARTICIPANTS

Twenty subjects were enrolled in the experiment. The demographic information of participants is shown in Table 1.

TABLE 1. Demographic.

Gender	Number	Average Age (SD)
Male	12	68.25 (4.12)
Female	8	71.63 (5.34)

The enrollment criteria were:

a) *It was a unilateral stroke.*
b) *The stroke took place within the last year.*
c) *The stroke diagnosis was based on an MRI, Computed Tomography, and neurological examination.*
d) *The proximal movement of the upper stroke limb reached the Browns fourth stage (inclusive) with dyskinesia.*
e) *No obvious cognitive deficit (short intelligence test score ≤ 20 points).*
f) *Capable of following simple instructions and understanding the experimental process and rehabilitation.*
g) *Willing to join and sign the consent form. The experimental site was at the rehabilitation department of Taipei Veterans General Hospital.*

2) PROCEDURES

Participants were informed of the experimental procedure and the purpose of the research. Participants were required to sign a consent form. In addition, participants were required not to participate in other rehabilitation training activities.

The participants attended 60 min stroke-rehabilitation sessions three times per week for eight weeks, and all participants completed the 24 training sessions. To avoid extreme fatigue from continuous training, at least one day off was provided between each session. Therapists could adjust the degree of difficulty according to the participants' evaluation. The degree of difficulty of each training was set to satisfy the motor ability limit of the participant. Clinical assessments were conducted three days before training (0 week), within three days after training (eight-week endpoint).

3) MEASUREMENT

Traditional evaluation scales such as FMA, TEMPA, and WMFT have strict assessment methods, procedures and standards. If a participant scores higher on the scale, then the participant has a better recovery status. If a participant's score lower on the scale, then the participant has a worse recovery status. Because our research focuses on the upper limbs, not all items on the traditional evaluation scale were used. The FMA score range is 0 to 22 points, the TEMPA score range is 0 to 27 points, and the WMFT score range is 0 to 5 points.

4) ANALYSIS METHOD

Because the number of samples was small, data distribution was unclear and the evaluation values of the same patient's pretest and posttest data were related variables, we employed the Wilcoxon rank-4sum test, a nonparametric test used to analyze whether the pretest and posttest data were significantly different, thereby testing the effectiveness of the bi-lateral VR task. We also used Spearman correlation analysis and a double-tailed test to evaluate the correlation between the motor indicators and the traditional evaluation scales [41].

D. DEVELOPMENT OF EVIDENCE-BASED ASSESSMENT MODEL

Based on development of an evidence-based, convincible assessment model for stroke rehabilitation, we followed the flow diagram in Fig. 3. First, we performed clustering (K-means) based on evaluation scales to cluster participants' data. Next, we collected skeleton data which were collected when a participant was performing the VR task via Kinect and extracted features by motor indicators. Finally, we fit multi-classifiers (MLP, RBFNN, SVM) based on evaluation scales, thereby developing an evidence-based assessment model.

1) CLUSTERING OF MULTI-DIMENSIONAL EVALUATION SCALES

Clustering of multi-dimensional evaluation scales Traditional evaluation scales are reliable, convincible and highly recognized by therapists. To further evaluate the patients' recovery status, we used clustering analysis based on traditional evaluation scales' scores to place each patient into the three impairment levels. Each level could be seen as one patient's recovery status.

The K-means algorithm was used in clustering analysis. The K-means algorithm repeatedly minimized the squared error between the empirical mean of a cluster and the individual data points [37]. Traditional evaluation scale results of 20 participants were mapped into three dimensional
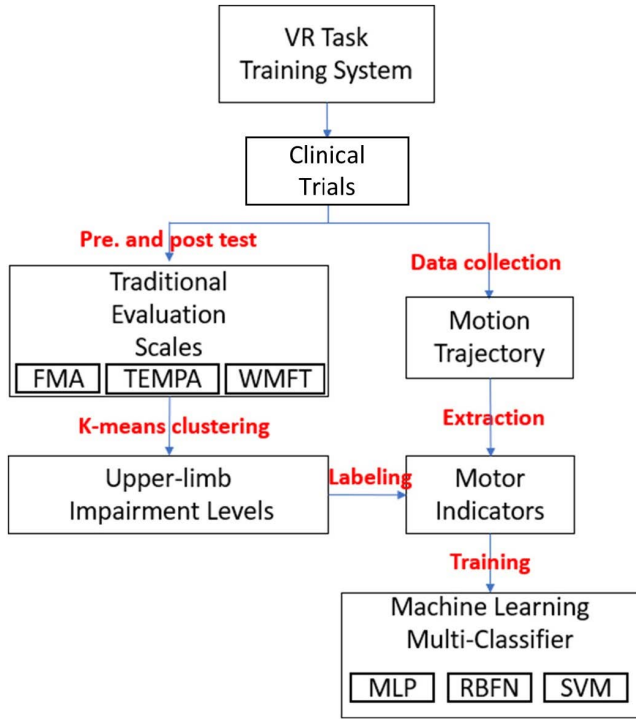
FIGURE 3. Flow diagram of the assessment model for our VR stroke-rehabilitation system. First, K-means clustering based on evaluation scales was used to cluster participants' data. Next, the skeleton data were collected when a participant was performing the VR task via Kinect, and the features were extracted as motor indicators. Finally, multi-classifiers (MLP, RBFNN, SVM) based on evaluation scales were adopted to develop an evidence-based assessment model.

coordinates (x; y; z) in the form of (FMA, TEMPA, WMFT). Different K values were taken to classify the data with the K-means algorithm. The clustering effect of each K value was validated by the silhouette coefficient [42], thereby determining the appropriate K value.

2) ASSESSMENT MODEL BY MACHINE LEARNING

Based on the K-means clustering algorithm with traditional evaluation scales' scores, we classified the participants' evaluation scales' scores into different clusters. Therefore, we were able to establish a classification model based on these labels. We used a machine learning-based classification algorithm to build an assessment model that could automatically classify stroke patients into different impairment levels.

The machine learning-based classification algorithms used in this study are composed of neural networks and support vector machines. Neural network is an artificial neural network formed by the interconnection of neurons for solving artificial intelligence problems. Neural networks have many different architectures. In this paper, a MLP (Multilayer Perceptron) and a RBFN (Radial Basis Function Network) were used. A Support Vector Machine (SVM) is a supervised machine learning method. It is widely used when there is a small number of samples and nonlinear and high-dimensional pattern recognition is needed. However, SVM might be

incapable of multivariate data. In this case, LibSVM is often used to analyze multivariate experimental data [43].

To validate the effectiveness of the assessment model, we used the impairment level classified by K-means with traditional evaluation scales as the expected value. Some of the motor indicators chosen as input features for the assessment model have been validated for correlation between motor indicators and traditional evaluation scales. To evaluate the assessment model, we tested the accuracy and the mean absolute percentage error (MAPE) for MLP, RBFN and SVM classifiers. The MAPE's calculation formula is shown in equation (5).

$$MAPE = \frac{1}{T} \sum_{k=1}^{T} \left| \frac{d_k - y_k}{d_k} \right| \times 100\% \tag{5}$$

where, d_k is the actual value of the kth data; y_k is the predicted output value of the kth data; T is the total amount of data. The relationship between MAPE and prediction effect is defined as when MAPE < 10, the prediction effect is accurate; when $50 \leq$ MAPE, the prediction effect is inaccurate.

IV. RESULTS

A. COMPARISON OF PRETEST AND POSTTEST DATA IN TRADITIONAL EVALUATION SCALES

The comparison of pretest and posttest data in three traditional evaluation scales is shown in Table 2. The scores of all evaluation scales were significantly improved (P < 0:05), indicating the bi-lateral VR task has significant efficacy for the rehabilitation of stroke patients.

TABLE 2. Traditional evaluation scale statistical results.

Scales	Pretest	Posttest	P-value
FMA	14.60 (4.39)	17.15 (3.60)	0.001**
TEMPA	-12.60 (7.93)	-10.35 (6.63)	0.005**
WMFT	3.66 (0.77)	3.84 (0.67)	0.020*

Significance Level = 0.05; *P < 0.05; ** P < 0.01.

B. CORRELATION ANALYSIS OF MOTOR INDICATORS AND TRADITIONAL EVALUATION SCALES

The correlation between horizontal motor indicators, vertical motor indicators and traditional evaluation scales in pretest and posttest data are shown in Fig. 4 (A) and Fig. 4 (B). Aiming time and the amount of time between stabilization and catching correlate significantly with FMA, TEMPA, and WMFT in pretest and posttest. However, the results were not as expected. Aiming time showed positive correlation but the amount of time between stabilization and catching showed negative correlation. We expected that if aiming time was shorter, the amount of time between stabilization and catching would be longer if participants had better motor ability. This result might be due to the degree of difficulty setting for the bi-lateral VR task.

The correlation between full body motor indicators and traditional evaluation scales in pretest and posttest data
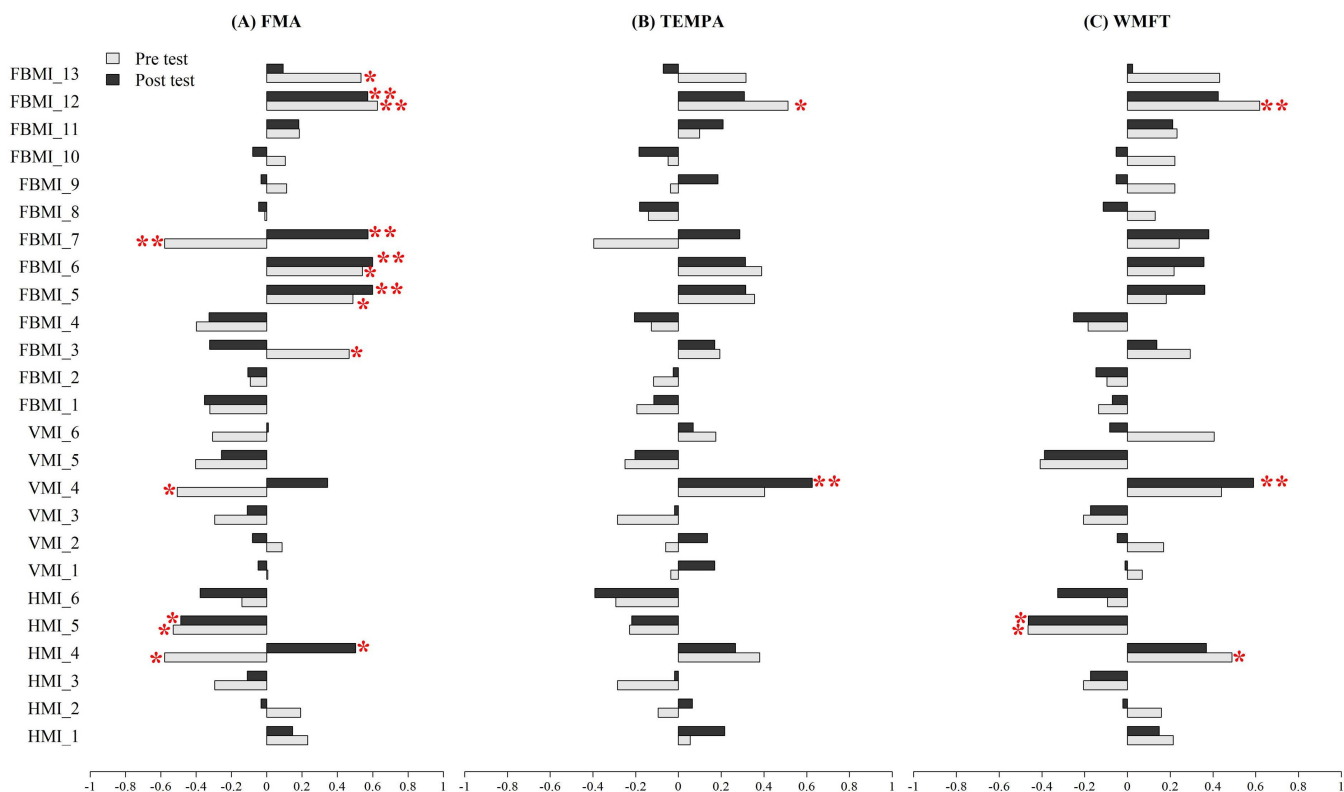
FIGURE 4. Correlation coefficients between motor indicators including horizontal motor indicator (HMI), vertical motor indicator (VMI), and full body's motor indicator (FBMI), and traditional valuations scales which were FMA, TEMPA, and WMFT in Pretest and Posttest, respectively. Note that * means $P < 0.05$ and ** means $P < 0.01$. H(V)MI_1 = first time entering the catching range evaluated by horizontal(vertical) motor indicator; H(V)MI_2 = last time entering the catching range evaluated by horizontal(vertical) motor indicator; H(V)MI_3 = catching time evaluated by horizontal(vertical) motor indicator; H(V)MI_4 = aiming time evaluated by horizontal(vertical) motor indicator; H(V)MI_5 = amount of time between stabilization and catching ball evaluated by horizontal(vertical) motor indicator; H(V)MI_6 = number of times getting in and out of catching range evaluated by horizontal(vertical) motor indicator; FBMI_1 = catching ball time difference between normal arm and affected arm; FBMI_2 = catching ball time cost by affected arm; FBMI_3 = movement distance by affected arm; FBMI_4 = movement difference between affected arm and normal arm; FBMI_5 = maximum speed by affected arm before ball arrives at the highest location; FBMI_6 = average speed by affected arm before ball arrives at the highest location; FBMI_7 = V -variation by affected arm before ball arrives at the highest location; FBMI_8 = maximum speed by affected arm after ball arrives at the highest location; FBMI_9 = average speed by affected arm after ball arrives at the highest location; FBMI_10 = V -variation by affected arm after ball arrives at highest location; FBMI_11 = maximum extension distance from arm movement in the horizontal direction; FBMI_12 = Maximum extension distance from arm movement in the vertical direction; FBMI_13 = maximum extension distance from arm movement in the front-view direction.

are shown in Fig. 4 (C). Maximum speed, average speed, *V-variation* by the affected arm before the ball arrives at the highest location and maximum extension distance from arm movement in the vertical direction correlates significantly with FMA, TEMPA, and WMFT in pretest and posttest data. Each of these motor indicators are related to muscle control ability. In addition, we expected that a faster upper-limb movement speed would mean the participant has better motor ability. However, according to Fig. 4 (C), many motor indicators have no correlation with traditional evaluation scales in pretest and posttest data. This result might be due to the small number of samples and need further investigation.

C. CLUSTER ANALYSIS BASED ON MULTI-DIMENSIONAL EVALUATION SCALES

The clustering analysis based on the K-means algorithm is used to determine the most proper K value. By fitting the clustering model by the silhouette coefficient, we got the highest silhouette value when K = 3. Therefore, the patient's impairment level was divided into three categories based on the multidimensional evaluation scales, with a total of 40 points (20 patients' evaluation scale data in pretest and posttest data). The spatial clustering result when K = 3 is shown in Fig. 5. As shown, the number of data in Cluster1, Cluster2, and Cluster3 are 15, 18, and 7 respectively.

Based on Fig. 4, traditional evaluation scales were replaced with the clustering result in reanalyzing the correlation between the motor indicators and the clustering result, as shown in Fig. 6. A total of six motor indicators correlates with the clustering results. These 6 indicators are the number of times getting in and out of catching range, aiming time, maximum speed, *V-variation* by affected arm before ball arrives at the highest location, average speed by affected arm after the ball arrives at the highest location, and maximum extension distance from arm movement in the vertical direction.

TABLE 3. Hyperparameter settings of machine learning models.

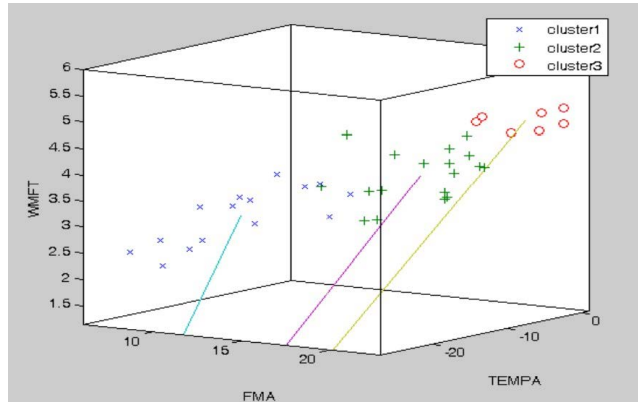| Models	Parameters	Motor indicators			
		All (25)	Pretest - posttest (7)	Posttest - pretest (7)	Clustering (6)
MLP	Learning efficiency	0.5	0.5	0.5	0.5
	MSE	0.007	0.02	0.012	0.03
	Learning times	80,000	80,000	80,000	80,000
RBFN	Learning efficiency	0.5	0.5	0.5	0.5
	MSE	0.07	0.08	0.06	0.04
	Learning times	80,000	80,000	80,000	80,000
RBFN	cost	512	32	2	32
	gamma	0.08	0.5	0.5	0.008

FIGURE 5. 3D scatter plot for the traditional evaluation scales, Fugl-Meyer Assessment (FMA) [6], Wolf Motor Function Test (WMFT) [7], and Test D'évaluation Des Membres Supérieurs Des Personnes Âgées (TEMPA) [8]. K-means clustering was based on them, and the highest silhouette value was attained when K = 3. The blue cross, green plus, and red circle are Cluster 1, Cluster 2, and Cluster 3, respectively. The corresponding number of data are 15, 18, and 7 respectively.

D. MACHINE LEARNING BASED CLASSIFICATION MODEL

Based on the above correlation analysis, we found that some motor indicators correlate with traditional evaluation scales. These motor indicators were used to train MLP, RBFN, and SVM classifiers. The initial hyperparameters, shown in Table 3, are aimed to fit the classifiers based on the overall MAPE. In neural network analysis, we first chose all (19) motor indicators to the fit neural network, as shown in Table 4. The optimal number of neurons in MLP is two, and the number of optimal neurons in RBFN is three. We also chose subclasses of motor indicators to fit MLP, RBFN classifiers, as shown in Table 5. In SVM analysis, we also chose all motor indicators and subclasses of motor indicators to fit SVM classifier. The classification result of SVM is shown in Table 6. Based on above the classification result, which uses a two-neuron architecture with all motor indicators, has the best performance of about 86% overall accuracy and 12.33% MAPE. In contrast, the highest accuracy of SVM is only 61.5%.

V. DISCUSSION

According to Table 1, bi-lateral VR task could be effectively used as rehabilitation training for stroke patients. However, based on the clustering analysis above, some motor indicators
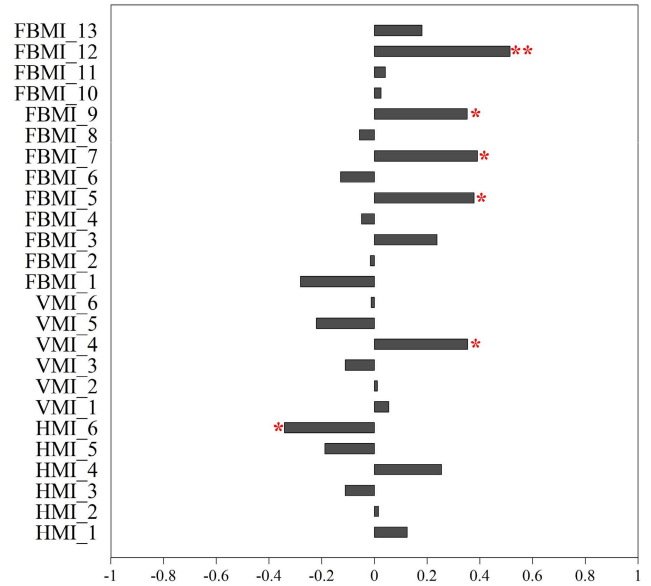

FIGURE 6. Correlation coefficients between motor indicators including horizontal motor indicator (HMI), vertical motor indicator (VMI), and full body's motor indicator (FBMI), and corresponding cluster. Note that * means P < 0.05 and ** means P < 0.01.

unexpectedly have less correlation with traditional evaluation scales.

According to Fig. 4 some of the motor indicators have significant correlation with traditional evaluation scales in pretest and posttest as expected. For example, maximum speed, average, V -variation by affected arm before the ball arrives at the highest location are all positive correlations with traditional evaluation scales. However, some of the motor indicators unexpectedly have less correlation with traditional evaluation scales in pretest and posttest data. We believe that if the number of samples increases in the future, these motor indicators would have significant correlation with all traditional evaluation scales.

For the assessment model's selection of motor indicators, we analyzed the correlation between horizontal, vertical, full body motor indicators and clustering in Fig. 6. We discovered that aiming time, maximum speed, V -variation by affected arm before the ball arrives at the highest location, maximum extension distance from arm movement in the vertical direction have both significant correlation with traditional

TABLE 4. Classification results of neural network and all motor indicators.

	Number of neurons								
	2			3			4		
	Training Accuracy	Test Accuracy	Overall Accuracy	Training Accuracy	Test Accuracy	Overall Accuracy	Training Accuracy	Test Accuracy	Overall Accuracy
Models	Training MAPE	Test MAPE	Overall MAPE	Training MAPE	Test MAPE	Overall MAPE	Training MAPE	Test MAPE	Overall MAPE
MLP	89.38%	72.5%	86%	86.25%	67.5%	82.5%	86.25%	67.5%	82.5%
	10.33%	20.31%	12.33%	11.34%	22.07%	13.49%	11.07%	22.71%	13.4%
RBFN	81.25%	60%	77%	87.5%	63%	82.6%	84.38%	63%	80.1%
	13.86%	29.24%	16.94%	11.06%	30.67%	14.98%	12.68%	30.73%	16.29%

TABLE 5. Classification results of neural networks and selected motor indicators.

	MLP (2 neurons)			RBFN (3 neurons)		
	Training Accuracy	Test Accuracy	Overall Accuracy	Training Accuracy	Test Accuracy	Overall Accuracy
Motor indicator	Training MAPE	Test MAPE	Overall MAPE	Training MAPE	Test MAPE	Overall MAPE
Pretest - posttest (7)	83.13%	60%	78.5%	81.25%	60%	77%
	16.68%	24.06%	18.1%	14.78%	35.49%	18.92%
Posttest - pretest (7)	83.13%	52.5%	77%	84.38%	60%	79.5%
	11.18%	28.84%	14.71%	8.09%	33.66%	13.2%
Clustering (6)	78.75%	60%	75%	78.12%	55%	73.5%
	19.94%	27.29%	21.41%	17.11%	35.98%	20.88%

TABLE 6. Classification results of SVM and motor indicators.

Motor indicator	Training Accuracy	Test Accuracy	Overall Accuracy
All (25)	51.88%	45%	50.5%
Pretest - posttest (7)	56.25%	42.5%	53.5%
Posttest - pretest (7)	51.88%	55%	52.5%
Clustering (6)	61.88%	60%	61.5%

evaluation scales and clustering. These motor indicators might be more important to the assessment model.

The number of neurons was considered in constructing a neural network classifier with all motor indicators. The optimal MLP and RBFN neural network models with all motor indicators were developed, as shown in Table 4. In considering the selection of motor indicators, we tried different subclasses of motor indicators to fit MLP and RBFN neural network classifiers, as shown in Table 5. However, the classification model with the highest overall accuracy and highest overall MAPE was trained by all motor indicators. We considered that although some motor indicators might not have significant correlation with traditional evaluation scales in a traditional statistics aspect, these motor indicators still impact the neural network-based classification model. The neural network models could learn implicit relations between those motor indicators. Moreover, the number of motor indicators that correlate to traditional evaluation scales might be too small for neural network models to learn well. Therefore, our neural network model still shows great potential.

According to Table 6, the SVM classification model trained with motor indicators considered traditional evaluation scales. The motor indicators with significant correlation to clustering show the highest overall accuracy in the SVM classification model. However, results of overall accuracy in Table 6 are not great.

Although several studies have devoted on developing VR system for stroke rehabilitation, it is still lack of a system which incorporated both sensor and traditional evaluation scales. More specifically, previous research adopted VR training tasks in stroke rehabilitation to show the efficiency of VR rehabilitation by comparing pretest and posttest data without sensor data and traditional evaluation scales [31], [32], [33], [34]. While several studies only considered

the sensor data to implement the evaluation [25], [26], [27], [28], [35], [37], [38], [39]. Our system was developed by considering both sensor data and traditional evaluation scales. Since the VR systems lack of similarity in previous research, comparing the results such as the performance of the proposed models would have no meaning. Yet, our results have shown that considering both sensor data and traditional evaluation scales is helpful to develop the VR system.

Because the number of samples is small, the results might not necessarily be accurate. Nonetheless, an automated motor assessment model is feasible. If we have enough samples in the future, we would consider doing the following: (1) Increase the type of evaluation scales. (2) Increase the number of clusters. (3) Implement additional motor indicators.

VI. CONCLUSION

In this paper, we developed a VR-based stroke rehabilitation system for the upper limbs. We proposed a motor assessment model with 20 stroke participants' traditional evaluation scales (FMA, TEMPA and WMFT). First, we used the K-means algorithm to cluster participants' pretest and posttest data. Furthermore, we designed many motor indicators for a bilateral VR task and analyzed their correlation with traditional evaluation scales. Maximum speed, average speed, V-variation by the affected arm before the ball arrives at the highest location and maximum extension distance from arm movement in the vertical direction correlates significantly with FMA, TEMPA, and WMFT in pretest and posttest data. Each of these motor indicators are related to muscle control ability. Moreover, we used clustering analysis and the correlation between motor indicators and traditional evaluation scales to fit the machine learning-based classification model (MLP, RBFN, SVM). Finally, our motor assessment model with MLP architecture achieved 86% overall accuracy. In the future, we would perform a larger clinical trial to enhance the accuracy of our assessment model. We would also adopt advanced machine learning methods such as DNN, RNN, and LSTM to exploit our Kinect-based skeleton data to implement a higher performance motor assessment model.

REFERENCES

[1] W. Johnson, O. Onuma, M. Owolabi, and S. Sachdev, "Stroke: A global response is needed," *Bull. World Health Org.*, vol. 94, no. 9, pp. 634A–635A, 2016.

[2] K. P. Doyle, R. P. Simon, and M. P. Stenzel-Poore, "Mechanisms of ischemic brain damage," *Neuropharmacology*, vol. 55, no. 3, pp. 310–318, Sep. 2008.

[3] S. L. Wolf et al., "Effect of constraintinduced movement therapy on upper extremity function 3 to 9 months after stroke: The EXCITE randomized clinical trial," *Jama*, vol. 296, no. 17, pp. 2095–2104, Nov. 2006.

[4] D. J. Lin, S. P. Finklestein, and S. C. Cramer, "New directions in treatments targeting stroke recovery," *Stroke*, vol. 49, no. 12, pp. 3107–3114, Dec. 2018.

[5] W. Lu, C. Wang, J. Lin, C. Sheu, and C. Hsieh, "The minimal detectable change of the simplified stroke rehabilitation assessment of movement measure," *J. Rehabil. Med.*, vol. 40, no. 8, pp. 615–619, Aug. 2008.

[6] J. Sanford, J. Moreland, L. R. Swanson, P. W. Stratford, and C. Gowland, "Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke," *Phys. Therapy*, vol. 73, no. 7, pp. 447–454, Jul. 1993.

[7] S. L. Wolf, D. E. Lecraw, L. A. Barton, and B. B. Jann, "Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head-injured patients," *Exp. Neurol*, vol. 104, no. 2, pp. 125–132, May 1989.

[8] J. Desrosiers, R. Hébert, G. Bravo, and É. Dutil, "Upper extremity performance test for the elderly (TEMPA): Normative data and correlates with sensorimotor parameters," *Arch. Phys. Med. Rehabil.*, vol. 76, no. 12, pp. 1125–1129, Dec. 1995.

[9] A. S. Pollock, L. Legg, P. Langhorne, and C. Sellars, "Barriers to achieving evidence-based stroke rehabilitation," *Clin. Rehabil.*, vol. 14, no. 6, pp. 611–617, Dec. 2000, doi: 10.1191/0269215500cr369oa.

[10] N. J. Devlin, D. Parkin, and J. Browne, "Patient-reported outcome measures in the NHS: New methods for analysing and reporting EQ-5D data," *Health Econ.*, vol. 19, no. 8, pp. 886–905, Jun. 2010. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.1608

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

[12] O. Postolache, D. J. Hemanth, R. Alexandre, D. Gupta, O. Geman, and A. Khanna, "Remote monitoring of physical rehabilitation of stroke patients using IoT and virtual reality," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 562–573, Feb. 2021.

[13] M. Rodríguez-Hernández, J.-J. Criado-Álvarez, A.-I. Corregidor-Sánchez, J. L. Martín-Conty, A. Mohedano-Moriano, and B. Polonio-López, "Effects of virtual reality-based therapy on quality of life of patients with subacute stroke: A three-month follow-up randomized controlled trial," *Int. J. Environ. Res. Public Health*, vol. 18, no. 6, p. 2810, Mar. 2021. [Online]. Available: https://www.mdpi.com/1660-4601/18/6/2810

[14] N. Norouzi-Gheidari, A. Hernandez, P. S. Archambault, J. Higgins, L. Poissant, and D. Kairy, "Feasibility, safety and efficacy of a virtual reality exergame system to supplement upper extremity rehabilitation post-stroke: A pilot randomized clinical trial and proof of principle," *Int. J. Environ. Res. Public Health*, vol. 17, no. 1, p. 113, 2020. [Online]. Available: https://www.mdpi.com/1660-4601/17/1/113

[15] K. E. Laver, B. Lange, S. George, J. E. Deutsch, G. Saposnik, and M. Crotty, "Virtual reality for stroke rehabilitation," *Cochrane Database Systematic Rev.*, vol. 11, no. 11, 2017, Art. no. CD008349.

[16] I. J. M. de Rooij, I. G. L. van de Port, and J.-W.-G. Meijer, "Effect of virtual reality training on balance and gait ability in patients with stroke: Systematic review and meta-analysis," *Phys. Therapy*, vol. 96, no. 12, pp. 1905–1918, Dec. 2016, doi: 10.2522/ptj.20160054.

[17] R. S. Calabrò et al., "The role of virtual reality in improving motor performance as revealed by EEG: A randomized clinical trial," *J. NeuroEng. Rehabil.*, vol. 14, no. 1, p. 53, 2017.

[18] D. C. Porras, P. Siemonsma, R. Inzelberg, G. Zeilig, and M. Plotnik, "Advantages of virtual reality in the rehabilitation of balance and gait," *Neurology*, vol. 90, no. 22, pp. 1017–1025, May 2018. [Online]. Available: https://n.neurology.org/content/90/22/1017

[19] S. J. Page, P. Levine, S. Sisto, Q. Bond, and M. V. Johnston, "Stroke patients' and therapists' opinions of constraint-induced movement therapy," *Clin. Rehabil.*, vol. 16, no. 1, pp. 55–60, Feb. 2002, doi: 10.1191/0269215502cr473oa.

[20] S. K. Subramanian, C. B. Lourenço, G. Chilingaryan, H. Sveistrup, and M. F. Levin, "Arm motor recovery using a virtual reality intervention in chronic stroke: Randomized control trial," *Neurorehabilitation Neural Repair*, vol. 27, no. 1, pp. 13–23, 2013, doi: 10.1177/1545968312449695.

[21] Y. Hayashi, K. Nagai, K. Ito, S. J. Nasuto, R. C. V. Loureiro, and W. S. Harwin, "A feasible study of EEG-driven assistive robotic system for stroke rehabilitation," in *Proc. 4th IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, Jun. 2012, pp. 1733–1739.

[22] S. I. Lee et al., "Enabling stroke rehabilitation in home and community settings: A wearable sensor-based approach for upper-limb motor training," *IEEE J. Transl. Eng. Health Med.*, vol. 6, pp. 1–11, 2018.

[23] M. Panwar et al., "Rehab-Net: Deep learning framework for arm movement classification using wearable sensors for stroke rehabilitation," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 11, pp. 3026–3037, Nov. 2019.

[24] H. Zhou, H. Hu, and N. Harris, "Application of wearable inertial sensors in stroke rehabilitation," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, Jan. 2005, pp. 6825–6828.

[25] P. Phienphanich et al., "Stroke screening feature selection for arm weakness using a mobile application," *IEEE Access*, vol. 8, pp. 170898–170914, 2020.

[26] L. Lu et al., "Evaluating rehabilitation progress using motion features identified by machine learning," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 4, pp. 1417–1428, Apr. 2021.

[27] M. H. Lee, D. P. Siewiorek, A. Smailagic, A. Bernardino, and S. B. I. Badia, "Learning to assess the quality of stroke rehabilitation exercises," in *Proc. 24th Int. Conf. Intell. User Interfaces*. New York, NY, USA: Association for Computing Machinery, Mar. 2019, pp. 218–228, doi: 10.1145/3301275.3302273.

[28] S. Miao, C. Shen, X. Feng, Q. Zhu, M. Shorfuzzaman, and Z. Lv, "Upper limb rehabilitation system for stroke survivors based on multi-modal sensors and machine learning," *IEEE Access*, vol. 9, pp. 30283–30291, 2021.

[29] S. C. Yeh, M.-C. Huang, P.-C. Wang, T.-Y. Fang, M.-C. Su, P.-Y. Tsai, and A. Rizzo, "Machine learning-based assessment tool for imbalance and vestibular dysfunction with virtual reality rehabilitation system," *Comput. Methods Biomed.*, vol. 116, no. 3, pp. 311–318, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260714001667

[30] D. Rodríguez-Martín, A. Samà, C. Pérez-López, J. Cabestany, A. Català, and A. Rodríguez-Molinero, "Po sture transition identification on Pd patients through a SVM-based technique and a single waist-worn accelerometer," *Neurocomputing*, vol. 164, pp. 144–153, Sep. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231215002994

[31] L. Sheehy et al., "Home-based virtual reality training after discharge from hospital-based stroke rehabilitation: A parallel randomized feasibility trial," *Trials*, vol. 20, no. 1, p. 333, Jun. 2019.

[32] A. L. Aramaki, R. F. Sampaio, A. Cavalcanti, and F. C. M. S. E. Dutra, "Use of client-centered virtual reality in rehabilitation after stroke: A feasibility study," *Arquivos de Neuro-Psiquiatria*, vol. 77, no. 9, pp. 622–631, Sep. 2019.

[33] M. M. Lee, K. J. Lee, and C. H. Song, "Game-based virtual reality canoe paddling training to improve postural balance and upper extremity function: A preliminary randomized controlled study of 30 patients with subacute stroke," *Med. Sci. Monitor*, vol. 24, pp. 2590–2598, Apr. 2018.

[34] Y. H. Choi and N. J. Paik, "Mobile game-based virtual reality program for upper extremity stroke rehabilitation," *J. Visualized Exp.*, no. 133, 2018, Art. no. e56241.

[35] S. Kashi, R. F. Polak, B. Lerner, L. Rokach, and S. Levy-Tzedek, "A machine-learning model for automatic detection of movement compensations in stroke patients," *IEEE Trans. Emerg. Topics Comput.*, vol. 9, no. 3, pp. 1234–1247, Jul. 2021.

[36] S. Lee, Y.-S. 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.

[37] D. Biswas et al., "Recognizing upper limb movements with wrist Worn inertial sensors using k-means clustering classification," *Hum. Movement Sci.*, vol. 40, pp. 59–76, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167945714002115

[38] L. Wang, J. Liu, and J. Lan, "Feature evaluation of upper limb exercise rehabilitation interactive system based on Kinect," *IEEE Access*, vol. 7, pp. 165985–165996, 2019.

[39] Y. Liao, A. Vakanski, and M. Xian, "A deep learning framework for assessing physical rehabilitation exercises," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 468–477, Feb. 2020.

[40] D. Webster and O. Celik, "Experimental evaluation of Microsoft Kinect's accuracy and capture rate for stroke rehabilitation applications," in *Proc. IEEE Haptics Symp. (HAPTICS)*, Feb. 2014, pp. 455–460.

[41] C. Spearman, "The proof and measurement of association between two things," *Amer. J. Psychol.*, vol. 15, no. 1, pp. 88–103, 1904.

[42] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, Jan. 1987, [Online]. Available: https://www.sciencedirect.com/science/article/pii/0377042787901257

[43] N. Kato, T. Tanaka, S. Sugihara, and K. Shimizu, "Development and evaluation of a new telerehabilitation system based on VR technology using multisensory feedback for patients with stroke," *J. Phys. Therapy Sci.*, vol. 27, no. 10, pp. 3185–3190, 2015.