

# B-Spline Modeling of Inertial Measurements for Evaluating Stroke Rehabilitation Effectiveness

Yi-Ting Hwang<sup>1</sup>, Yu-Qian Tung, Chun-Shu Chen<sup>2</sup>, and Bor-Shing Lin<sup>3</sup>, *Senior Member, IEEE*

**Abstract**—Patients who experience upper-limb paralysis after stroke require continual rehabilitation. Rehabilitation must be evaluated for appropriate treatment adjustment; such evaluation can be performed using inertial measurement units (IMUs) instead of standard scales or subjective evaluations. However, IMUs produce large quantities of discretized data, and using these data directly is challenging. In this study, B-splines were used to estimate IMU trajectory data for objective evaluations of hand function and stability by using machine learning classifiers and mathematical indices. IMU trajectory data from a 2018 study on upper-limb rehabilitation were used to validate the proposed method. Features extracted from B-spline trajectories could be used to classify individuals in the 2018 study with high accuracy, and the proposed indices revealed differences between these groups. Compared with conventional rehabilitation evaluation methods, the proposed method is more objective and effective.

**Index Terms**—B-spline, linear mixed model, inertial measurement unit (IMU), rehabilitation, virtual reality (VR).

## I. INTRODUCTION

STROKE can occur when a blood vessel in the brain clogs or ruptures and can cause cerebral ischemia, cerebral infarction, and even death. In 2020, one in six deaths attributed to cardiovascular disease was associated with stroke [1].

Manuscript received 22 February 2023; revised 28 May 2023 and 29 August 2023; accepted 4 October 2023. Date of publication 10 October 2023; date of current version 18 October 2023. This work was supported in part by the National Science and Technology Council in Taiwan under Grant NSTC 112-2221-E-305-001-MY3; in part by the University System of Taipei Joint Research Program under Grant USTP-NTPU-TMU-109-03 and Grant USTP-NTPU-NTOU-112-01; in part by the Faculty Group Research Funding Sponsorship by the National Taipei University under Grant 2023-NTPU-ORD-01; and in part by the Academic Top-Notch and Features Field Project Funding Sponsorship of National Taipei University, Taiwan, under Grant 111-NTPU\_ORDA-F-004. (*Corresponding author: Bor-Shing Lin.*)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Chang Gung Medical Foundation under IRB No. 102-3680B, and performed in line with the Declaration of Helsinki.

Yi-Ting Hwang and Yu-Qian Tung are with the Department of Statistics, National Taipei University, New Taipei City 237303, Taiwan (e-mail: hwangyt@gm.ntpu.edu.tw; karta252525@gmail.com).

Chun-Shu Chen is with the Graduate Institute of Statistics, National Central University, Taoyuan 320317, Taiwan (e-mail: cschen1207@ncu.edu.tw).

Bor-Shing Lin is with the Department of Computer Science and Information Engineering, National Taipei University, New Taipei City 237303, Taiwan (e-mail: bslin@mail.ntpu.edu.tw).

Digital Object Identifier 10.1109/TNSRE.2023.3323375

Stroke is also a leading cause of long-term disability; more than half of stroke survivors aged 65 years or older have reduced mobility [2]. Only two-thirds of individuals who have experienced stroke can walk independently, and less than half of those who lose basic upper-limb function regain it after 12 months [3], [4]. Such patients require upper-limb rehabilitation to restore their ability to complete daily tasks, such as showering, eating, and dressing, without assistance. In rehabilitation, a therapist selects individualized treatments and tasks, such as brick stacking, card flipping, and biking, in accordance with the needs of the patient [6]. Patients must perform these tasks repeatedly to regain the motor function of an impaired body part. However, patients might develop negative attitudes toward this repetitive treatment, limiting the effectiveness of rehabilitation [5]. To enhance patient enthusiasm for rehabilitation, [7], [8] have proposed computer-based systems for poststroke rehabilitation at home. Moreover, numerous virtual reality (VR)-based upper-limb systems have been designed to increase patient motivation for rehabilitation [9], [10], [11], [12], [13], [14]. Comprehensive reviews of these systems are provided in [15] and [16].

Therapists evaluate the effectiveness of rehabilitation by using scale-based tests, such as the Fugl–Meyer test, Action Research Arm Test, Box and Block Test, Jebsen–Taylor hand function test, and Brunnstrom stage test [17]. In computer- or VR-based rehabilitation systems, range of motion, movement speed and duration, hit rate, and subjective scale parameters may be used to assess improvement [9], [10], [11], [12], [13], [14]. The limitation of such indices is that they can only be used to perform subjective evaluations with a final outcome; individualized treatment plans cannot be designed if no detailed movement information is provided [18].

In rehabilitation-oriented games, digital systems with sensors are often used to record patient movements. In [19], a wearable system with several embedded inertial measurement units (IMUs) was designed to detect complex trunk and shoulder movements and to visualize these movements on a smart device. In [13], a system with a motion tracking device (MTD) was designed to track movement trajectories and used by patients playing a VR game. In [20], a full-body sensor system with 14 IMUs was used to capture position and orientation data for the body segments of four patients with stroke; these data were used to derive joint angles, which were in turn used to evaluate rehabilitation

progress. Descriptive statistics were used to summarize the data collected using the aforementioned sensor system. In [14], two Nintendo Wiimotes (wireless movement sensors) were used to track the movements of patients with stroke who were undergoing rehabilitation; these movements were mapped onto a three-dimensional body model that could be viewed on a computer screen. A modified wearable inertial sensing system that comprised eight IMUs with triaxial accelerometers and gyroscopes was developed to compute spatiotemporal kinematic metrics for assessing patients' upper-limb movements after stroke [21]. In [22], the associations between kinematic variables obtained from four sensors and clinical assessments were investigated. In [23], a seven-camera Vicon system was used to capture position data for body segments at a sampling rate of 120 Hz; these data were then used to calculate upper-body angles. The participants in [23] were asked to repeat a tasks multiple times, and the trial-to-trial variability of each kinematic parameter was computed. The performance of the the system used was evaluated using only subjective scale tests, such as the Fugl–Meyer test. The data collected by the IMUs were used only for visualization or computing the maximum patient movement angles during the evaluation process.

The trajectories of accelerometer and gyroscope signals obtained by IMUs during tasks provide crucial information regarding the movement ability of patients. When a simple task, such as forward movement, is executed, the IMU trajectory should be smooth. The smoothness of a trajectory is a potential stability index for assessing rehabilitation effectiveness. However, accelerometer and gyroscope data are obtained from signals transmitted by multiple devices. When a server cannot receive and process data instantly, the data spikes, packet loss, and data gaps may occur [24]. In [25], a functional data technique was used to remove noise, and IMU trajectory coefficients determined using this technique were useful for characterizing the motor impairment levels of patients with stroke. In the present study, a functional data technique was used to fit IMU data and obtain movement trajectories. Evaluation indices for movement stability were developed on the basis of these trajectories for objectively assessing the improvement in the motor abilities of patients with stroke after rehabilitation. The data collected in [13] were used to demonstrate the feasibility of these indices. The proposed indices are simple and can be calculated from IMU data. Assessing the improvement in motor ability after rehabilitation is crucial, and such assessment is currently performed using only subjective scale-based instruments. Future systems can use the indices proposed in this paper for objective and improved evaluations of the effectiveness of rehabilitation.

## II. METHODS

### A. Modeling IMU Trajectories

A wearable inertial sensing system often contains an accelerometer, a gyroscope, or a magnetometer. During a task, signals from the accelerometer, gyroscope, or magnetometer are collected by a microprocessor. Under normal conditions,

these signals have a smooth trajectory over time, and their rate of change is small. Thus, the rate of change of an IMU signal over the entire time interval for a task is a potential index for assessing the functional ability of a patient. Because wearable inertial measurement systems provides only discrete data, extreme raw data values can strongly affect the obtained rate of change of a signal. A functional model can be used to fit the raw signal and determine its rate of change.

Observed signals are a function of time, and the number of signals for a task depends on the sampling rate. A linear model for determining a parametric function for IMU signals is described in the following text. Let  $y_i$ ,  $i = 1, \dots, N$ , denote a sample trajectory collected from an IMU; let  $t_i$ ,  $i = 1, \dots, N$ , denote the corresponding recording time; and let  $N$  denote the number of recorded observations. A functional model for fitting the signal trajectory for a variable can be expressed as follows:

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, \dots, N, \quad (1)$$

where  $f(\cdot)$  is a prespecified smoothing function and  $\epsilon_i$  denotes independent random variables with a mean of 0. Given a class of basis functions  $\{\phi_j(\cdot), j = 1, \dots, K\}$ , the smoothing function can be expressed as follows:

$$f(t) = \sum_{j=1}^{K+d+1} \phi_j(t)c_j, \quad (2)$$

where  $K$  is a prespecified number of bases and  $d$  represents the degree of the polynomial function. To improve fit, the time interval can be divided into  $K + 1$  disjoint intervals of  $\tau_k$ , with  $k = 1, \dots, K$ , where  $a < \tau_1 < \tau_2 < \dots < \tau_K < b$ . The terms  $a$  and  $b$  are the lower and upper time bounds, respectively; these parameters are known as inner knots. Moreover, boundary knots can be selected such that  $f(t)$  is differentiable near the boundary. Let the knots be expressed as follows:

$$\begin{aligned} \xi_1 \leq \xi_2 \leq \dots \leq \xi_d \leq \xi_{d+1} < \xi_{d+2} \leq \dots < \xi_{d+K+1} \\ < \xi_{d+K+2} \leq \xi_{d+K+3} \leq \dots \leq \xi_{2d+K+2}, \end{aligned} \quad (3)$$

where  $\xi_{d+2} = \tau_1, \dots, \xi_{d+K+1} \equiv \tau_K, \xi_{d+1} = a$ , and  $\xi_{d+K+2} = b$ . The other knots ( $\xi_1, \dots, \xi_d$  and  $\xi_{d+K+3}, \dots, \xi_{2d+K+2}$ ) are selected arbitrarily but are often assumed to be equal to the boundary knots. A  $B$ -spline basis function was selected in this study. For  $d > 0$ , a  $B$ -spline basis function can be defined recursively as follows:

$$B_k^d(x) = \frac{x - \xi_k}{\xi_{k+d} - \xi_k} B_k^{d-1}(x) - \frac{\xi_{k+d+1} - x}{\xi_{k+d+1} - \xi_{k+1}} B_{k+1}^{d-1}(x), \quad (4)$$

where  $k = 1, \dots, K + d + 1$ ,

$$B_k^0(x) = \begin{cases} 1, & \xi_k \leq x < \xi_{k+1} \\ 0, & \text{else} \end{cases} \quad (5)$$

and  $B_k^0(x) \equiv 0$  if  $\xi_k = \xi_{k+1}$ . The spline fits are robust with respect to  $d$ . Cubic polynomials ( $d = 3$ ) are the common standard because the resulting curve appears perfectly smooth [26].

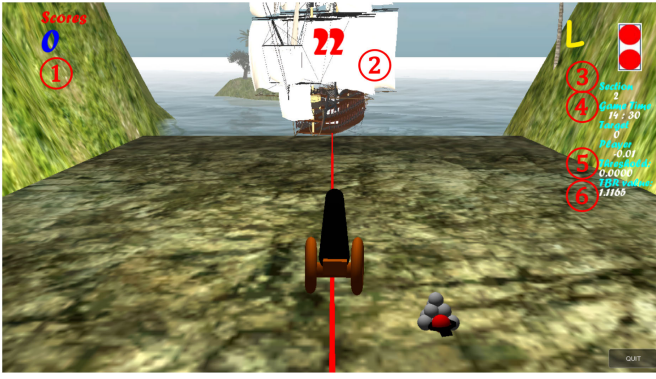


Fig. 1. Screenshot of VR rehabilitation game. The red numbers indicate the 1) patient score, 2) round timer, 3) current objective, 4) game timer, 5) theta–beta ratio (TBR) threshold for the patient, 6) and real-time TBR for the patient.

According to (2), the rate of change of a signal over time is  $df(t)/dt$ . Thus, we propose two metrics based on the rate of signal change over  $(a, b)$  for evaluating the stability of the trajectory; these metrics are expressed as

$$S_T^I = \int_a^b \left| \frac{df(s)}{ds} \right| ds \quad (6)$$

and

$$S_T^{II} = \int_a^b \left( \frac{df(s)}{ds} \right)^2 ds. \quad (7)$$

### B. Example of a VR Upper-Limb Rehabilitation System

Data regarding a VR rehabilitation game from the study of Lin et al. [13] were used to test whether the proposed method for IMU trajectory modeling could be used to evaluate rehabilitation. An experiment was conducted at Chang Gung Memorial Hospital in Taoyuan, Taiwan. Eighteen patients with stroke were enrolled in the study. Prior to the start of the experiment, written consent was obtained in accordance with the Declaration of Helsinki. The Institutional Review Board of Chang Gung Medical Foundation approved the experiment (IRB No. 102-3680B).

The considered system comprised an MTD, an electroencephalography (EEG) device, and a VR game that was designed by psychiatrists and therapists to increase patient interest and improve patient motion. A screenshot during gameplay is presented in Fig. 1.

The MTD required approximately 1 min for calibration, and each patient was asked to hold the MTD still during this time until the calibration process is complete. They then had 5 s to perform their best pronation and supination movements with the affected forearm, enabling the VR game to place game elements in accordance with the patients' maximum movement range. In each game, the patient first had 25 s to move the MTD backward and forward to load a shell into a cannon; the game proceeded to aiming when the cannon had been loaded. After loading, the cannon would fire automatically 10 s later. During this time, the patient must rotate the MTD to aim the cannon at a ship and maintain this position until the cannon fired. Finally, the patient was given 5 s to rest before the next

round. If a patient's mean hit rate exceeded 80%, the difficulty of the game is automatically increased; that is, the angle at which the ship might appear and thus the maximum range of motion required for aiming increased. The difficulty would not decrease, regardless of patient accuracy. EEG data were used to enhance the patient's attention; if the patient's concentration waned, the system displayed a stimulus to increase patient focus. Additional details regarding the system are presented in [13].

The experiment comprised three equally sized groups: a control group (A) that participated in conventional rehabilitation, a group (B) that played the VR game, and a group (C) that played the VR game and used the EEG device. The EEG data were used only for monitoring and ensuring patient attention. Besides conventional rehabilitation, every week for 4 weeks, the patients in groups B and C performed three game-based rehabilitation experiments that lasted 35 min; each experiment comprised two 15-min play sessions that interrupted by a 5-min rest period. The patients participated in a total of 24 rehabilitation sessions, each of which comprised up to 22 games lasting a maximum of 40 s. The total score, hit rate, number of games, and maximum wrist turning angle were recorded for each patient during each session. The MTD could track forearm pronation and supination trajectories and angles on the basis of acceleration, angular velocity, and magnetic field strength measurements collected by embedded triaxial accelerometers, gyroscopes, and magnetometers, respectively. Data were recorded every 1/20 s in the  $x$ -direction,  $y$ -direction, and  $z$ -direction. Nine trajectories were obtained for each game. In this paper, ACX, ACY, and ACZ denote the accelerometer measurements; GRX, GRY, and GRZ denote the gyroscope measurements; and MAGX, MAGY, and MAGZ denote the magnetic field measurements. A sample acceleration trajectory in the  $z$ -direction for a patient playing the game is displayed in Fig. 2. Because the objective of the present study was to identify new metrics for IMU trajectory data, the patients in group A were not included in the current analysis.

### C. Statistical Analysis

The duration for each game was not fixed; thus, the time interval of each game was standardized to  $[0, 1]$  as follow. A new time variable was defined as follows:

$$t_i^* = \frac{t_i - \min(t_i)}{\max(t_i) - \min(t_i)}, \quad i = 1, 2, \dots, N. \quad (8)$$

To combine the data for each session, the MTD measurements were normalized as follows: Let  $y_i$ ,  $i = 1, \dots, N$ , denote the set of measurements for one game. The normalization equation is given in (9):

$$y_i^* = \frac{y_i - \bar{y}}{s_y}, \quad i = 1, 2, \dots, N, \quad (9)$$

where  $\bar{y}$  is the mean and  $s_y$  is the standard deviation of the set. Fig. 2 displays the normalized ACZ trajectory for one game; in this figure, "o" denotes a raw data data point.

The functional model defined in (1) was used to fit the normalized trajectory  $(t_i^*, y_i^*)$ ,  $i = 1, 2, \dots, N$ . The  $B$ -spline

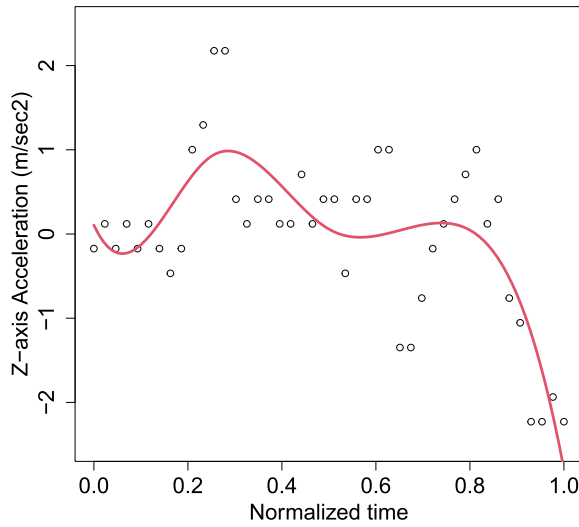


Fig. 2. Acceleration trajectory in the z-direction for a game, where “o” denotes the raw trajectory and the solid line denotes the fitted trajectory.

basis function was selected in this study under the assumption that  $d = 3$  if only three inner knots existed ( $\tau_1 = 0.25$ ,  $\tau_2 = 0.50$ , and  $\tau_3 = 0.75$ ; that is,  $K = 3$ ). When  $d = 3$ , the resulting fitting curve had seven estimated coefficients:  $\hat{c}_j$ ,  $j = 1, 2, \dots, 7$ .

Patients undergoing rehabilitation typically perform the same motion numerous times during each session and attend multiple rehabilitation sessions. Thus, the number of rehabilitation sessions was set to  $R$ , and the number of patients was set to  $n$ .  $N_{ir}$  denoted the number of games for the  $r$ th session and  $i$ th patient, where  $r = 1, 2, \dots, R$ , and  $i = 1, 2, \dots, n$ . Because (4) is a recursive formula, the estimated curve had no closed form. To combine data for each subround, the time intervals  $(a, b)$  were partitioned into  $m$  equal subintervals of  $h = (b - a)/m$ , which were denoted as  $s_0 = a$ ,  $s_1 = a + h$ ,  $\dots$ ,  $s_m = b$ . Here,  $\hat{f}_{irk}(s_t)$  denotes the fitted curve for the  $k$ th game in the  $r$ th session for the  $i$ th patient, where  $t = 0, 1, 2, \dots, m$ ,  $k = 1, 2, \dots, N_{ir}$ ,  $r = 1, 2, \dots, R$ , and  $i = 1, 2, \dots, n$ . This curve was derived through replacement of  $B_j^d(x)$ ,  $j = 1, 2, \dots, K + d + 1$ , and the estimated coefficients in (2). Because the number of games varied between sessions and patients, the fitted value for the  $r$ th rehabilitation session was defined as the average of all repeated motions during the session as follows:

$$\hat{f}_{ir}(s_t) = \frac{1}{N_{ir}} \sum_{k=1}^{N_{ir}} \hat{f}_{irk}(s_t), \quad (10)$$

where  $t = 0, 1, 2, \dots, m$ ,  $r = 1, 2, \dots, R$ , and  $i = 1, 2, \dots, n$ .

In this study, we assumed that  $m = 100$ . The numbers of rehabilitation sessions and patients were  $R = 24$  and  $n = 12$ , respectively. A fitted curve for the normalized z-direction acceleration trajectory for a game is displayed in Fig. 2 (solid line); the fitted z-direction acceleration for the games in a session is red, and the corresponding average curve for the session is black. Although the curves vary between games, the average fitted curve is smooth and follows the line  $y = 0$ . Fig. 4 presents the average fitted curves for all 24 sessions for

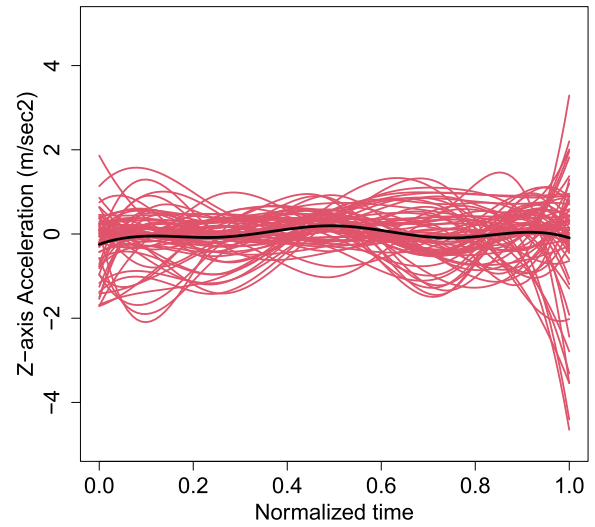


Fig. 3. Trajectories for the z-direction acceleration for a session. Fitted curves for the games are in red, and the average session trajectory is in black.

groups B and C. The curves for group C have less variation than do those for group B.

The integrals of (6) and (7) were estimated by replacing  $f(s)$  by  $\hat{f}_{ir}(s_t)$  as follows:

$$S_{ir}^I = \sum_{t=1}^{100} \frac{|\hat{f}_{ir}(s_t) - \hat{f}_{ir}(s_{t-1})|}{s_t - s_{t-1}} \times (s_t - s_{t-1}) \quad (11)$$

and

$$S_{ir}^{II} = \sum_{t=1}^{100} \left( \frac{\hat{f}_{ir}(s_t) - \hat{f}_{ir}(s_{t-1})}{s_t - s_{t-1}} \right)^2 \times (s_t - s_{t-1}), \quad (12)$$

where  $r = 1, 2, \dots, 24$  and  $i = 1, 2, \dots, 12$ .

The fitted curve in Fig. 2 trends toward 1; to avoid an excessively strong boundary effect, alternative estimates of the rates of signal change in (6) and (7) were determined as follows:

$$S_{ir}^{III} = \sum_{t=1}^{90} \frac{|\hat{f}_{ir}(s_t) - \hat{f}_{ir}(s_{t-1})|}{s_t - s_{t-1}} \times (s_t - s_{t-1}) \quad (13)$$

and

$$S_{ir}^{IV} = \sum_{t=1}^{90} \left( \frac{\hat{f}_{ir}(s_t) - \hat{f}_{ir}(s_{t-1})}{s_t - s_{t-1}} \right)^2 \times (s_t - s_{t-1}), \quad (14)$$

where  $r = 1, 2, \dots, 24$  and  $i = 1, 2, \dots, 12$ . For simplicity, the four indices expressed in (11) – (14) are denoted as I, II, III, and IV, respectively. Nine variables were recorded, resulting in 36 stability indices.

#### D. Evaluation

Three aspects of the feasibility of the aforementioned indices for rehabilitation evaluation were considered.

First, we evaluated whether the fitted curves of groups B and C differed by inputting the average of the estimated seven coefficients for each session into classification algorithms, which were used to classify the patients into groups B and

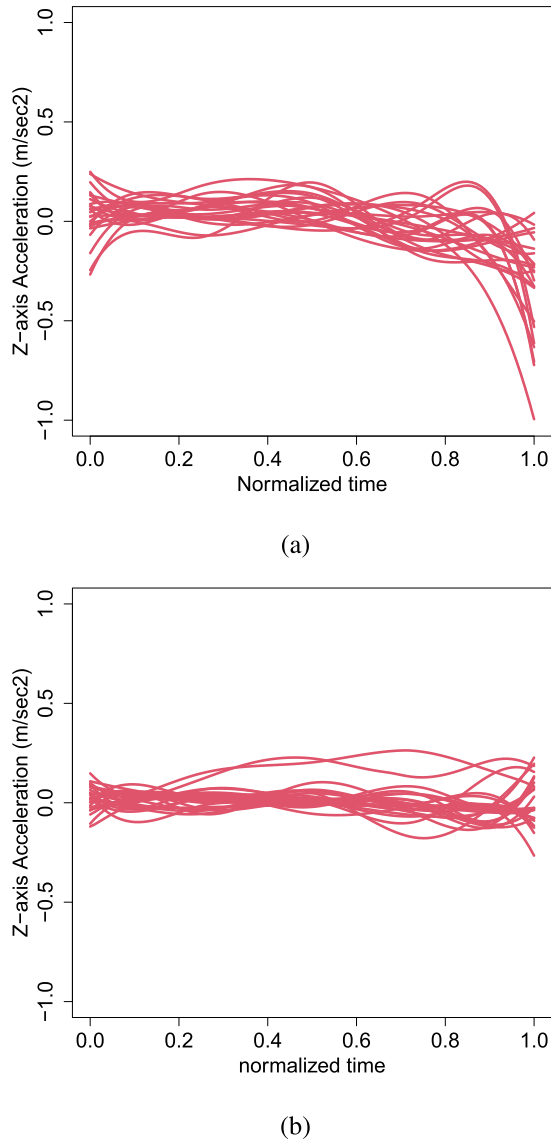


Fig. 4. Average z-direction acceleration trajectories for 24 sessions for (a) group B and (b) group C.

C. For a given motion replication  $k$  in session  $r$  for the  $i$ th patient, the vector of estimated coefficients was denoted as  $\hat{\mathbf{c}}_{ikr}$ . The vector of the average of the estimated coefficients for all motions in a session was expressed as follows:

$$\hat{\mathbf{c}}_{ir} = \frac{1}{N_{ir}} \sum_{k=1}^{N_{ir}} \hat{\mathbf{c}}_{ikr}, r = 1, 2, \dots, 24, i = 1, \dots, 12. \quad (15)$$

Each session was associated with nine variables and thus 63 features. Moreover, each of the 12 patients participated in 24 sessions, which resulted in 288 observations. Five classifiers were used, namely, the  $k$ -nearest neighbor (KNN), support vector machine (SVM), decision tree (DT), random forest (RF), and naive Bayes (NB) classifiers. Regression splines were applied using the spline package in R with the default hyperparameters. The Caret package was also used, and this package includes various classification methods and training functions. A five-fold cross-validation scheme was used for training. For the KNN classifier, the tuneGrid

function in the Caret package was selected for automatically determining the key parameters for  $k$  values of 1–30. For the SVM classifier, a radial basis function kernel was used, and the total number of unique combinations was set to ten for the grid search [27]. For the DT classifier, the cp value in tuneGrid was determined by searching from 0 to 0.05 in steps of 0.005. For the NB classifier, the default turning parameters for rf, usekernel, and adjust were used. Classification performance was assessed in terms of accuracy, specificity, and sensitivity. In this paper, sensitivity indicates the proportion of patients in group C that were correctly classified as being in group C, and specificity refers to the proportion of patients in group B that were correctly classified as being in group B. The final performance in terms of the aforementioned indices was the average value over 30 repetitions.

Second, the overall stability for 24 sessions was evaluated as follows:

$$\bar{S}_i^G = \frac{1}{24} \sum_{r=1}^{24} S_{ir}^G, i = 1, 2, \dots, 12; G = I, II, III, IV. \quad (16)$$

The stability of groups B and C was compared, and the Wilcoxon test was used to determine whether the overall stability of the two groups was significantly different.

Third, the progress of the rehabilitation (long-term effectiveness) was examined. Trajectories were expected to stabilize as the number of sessions increased. A linear mixed model was used to examine whether the stability declined as the number of sessions increased. This model included fixed effects and random effects. The fixed effects were those of group (GP), session (SR), and group–session (GP \* SR) interactions, and the two random effects were the intercept and slope of the trajectories. The Wald test was performed to evaluate the significance of the effects. When group–session interactions were significant, the rate of change in signal trajectory stability over sessions was significantly different between groups B and C. If the corresponding estimated coefficient was negative, the stability for group C declined faster than that for group B. The four proposed indices and original variables, including the total score, hit rate, number of games, and maximum wrist angle, were subsequently used to assess long-term rehabilitation effectiveness.

### III. RESULTS

#### A. Classification

A spline function comprises coefficients and a basis. The estimated coefficients provide information regarding trajectory characteristics. To demonstrate the effectiveness of the proposed stability indices, we examined whether the trajectory characteristics differed between groups B and C. Features were defined as the average of estimates of the coefficients over the  $N_{ir}$  games in a session. The results of a two-independent-samples  $T$ -test (Table I) revealed the variables that differed significantly between groups B and C. Among the nine variables, ACZ and MAGZ had the highest significance, whereas ACY, GRY, and MAGX had the lowest significance.

TABLE I

SIGNIFICANT DIFFERENCES BETWEEN GROUPS B AND C (IDENTIFIED ON THE BASIS OF A TWO-INDEPENDENT-SAMPLES *T*-TEST)

Variable	<i>c</i> <sub>1</sub>	<i>c</i> <sub>2</sub>	<i>c</i> <sub>3</sub>	<i>c</i> <sub>4</sub>	<i>c</i> <sub>5</sub>	<i>c</i> <sub>6</sub>	<i>c</i> <sub>7</sub>	Total
ACX	v			v	v	v	v	5
ACY	v					v	v	3
ACZ		v	v	v	v	v	v	6
GRX	v		v		v	v		4
GRY		v		v		v		3
GRZ	v	v			v	v	v	5
MAGX				v		v	v	3
MAGY	v		v	v	v			4
MAGZ	v	v	v		v	v	v	6

TABLE II

CLASSIFICATION PERFORMANCE

Methods	Tuning parameter	Accuracy	Specificity	Sensitivity
KNN	k=1	0.9290	0.8995	0.9708
SVM	$\sigma = 0.0146$ C = 1	0.9465	0.9265	0.9758
DT	cp=0.01	0.7856	0.8274	0.7595
RF	mtry = 6	0.9224	0.9348	0.9247
NB	fL = 1 adjust = 1	0.8406	0.9162	0.7847

KNN: K-nearest neighbor; SVM: support vector machine; DT: decision tree; RF: random forest; NB: naive Bayes.

TABLE III

COMPARISON OF RESULTS BETWEEN GROUPS B AND C

	Group B			Group C			<i>p</i> <sup>#</sup>
	Median	Mean	STD	Median	Mean	STD	
Total score	489.40	456.93	98.68	443.69	448.57	120.73	0.70
Hit rate	0.87	0.82	0.18	0.90	0.84	0.16	0.82
Number of games	33.38	34.14	2.57	30.71	31.12	2.30	0.06
Maximal angle	78.98	75.86	13.46	82.33	77.30	11.07	0.94

# Computed from a Wilcoxon test

Table II presents the classification results based on the average coefficients. The accuracy, specificity, and sensitivity of SVM and RF were higher than 90%; specifically, the sensitivity of SVM was 98%, and the specificity of the RF method was 93%. The KNN method had marginally lower accuracy than did the aforementioned two methods, and the DT and NB methods exhibited low performance. The accurate classification results obtained using the SVM and RF methods indicated that the characteristics of the fitted curves of groups B and C differed.

B. Overall Evaluations

The Wilcoxon test was used to compute whether the average values of the total score, hit rate, number of games, and maximum wrist angle over the 24 sessions for each patient differed between the groups (Table III). The number of games for group B was slightly higher than that for group C; this effect was marginally significant. However, although the total score and hit rate were slightly higher in group C, these effects were nonsignificant.

The *p* values obtained in the Wilcoxon text are displayed in a heat map in Fig. (5); the rows in this figure represent variables, the columns represent the four stability indices,

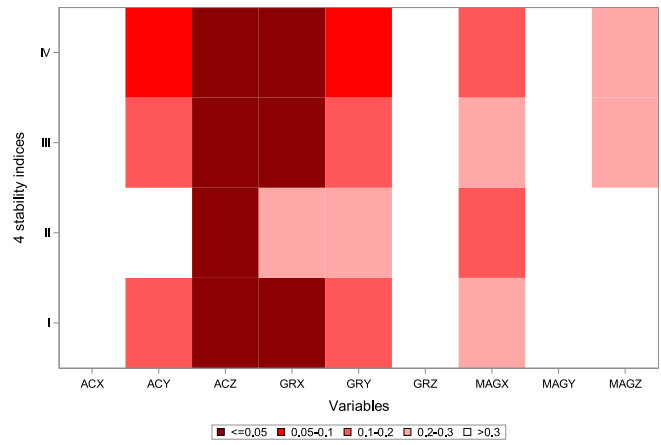


Fig. 5. Heat map for the *p* values of the differences in the proposed indices between groups B and C in the Wilcoxon test.

TABLE IV

STATISTICS OF THE FOUR INDICES FOR ACZ AND GRX

		Group B			Group C			<i>p</i> <sup>#</sup>
		Median	Mean	STD	Median	Mean	STD	
I	ACZ	0.98	1.09	0.47	0.53	0.55	0.12	0.009
	GRX	1.06	1.09	0.28	0.69	0.72	0.21	0.041
II	ACZ	2.06	2.82	2.18	0.65	0.75	0.35	0.026
	GRX	2.42	2.69	1.01	1.59	1.65	0.98	0.240
III	ACZ	0.83	0.82	0.35	0.39	0.39	0.08	0.015
	GRX	0.90	0.83	0.26	0.47	0.48	0.13	0.026
IV	ACZ	1.31	1.54	1.19	0.30	0.34	0.17	0.015
	GRX	1.54	1.42	0.75	0.46	0.51	0.30	0.009

# Computed using the Wilcoxon test

and the colors indicate the significance level (*p* < 0.05 was considered significant). ACZ had the most significant indices, followed by AC and GRX. In particular, all four indices derived from ACZ were significant. Moreover, all indices except for index II for AC and GRX were significant at the 0.05 level. Finally, the differences between groups B and C in the indices derived from ACX, GRZ, and MAGY were nonsignificant at *p* = 0.30.

Table IV provides the statistics of the four indices for ACZ, AC, and GRX. The median, mean, and standard deviation of these indices were significantly higher for the group B patients than for the group C patients. Moreover, the median, mean, and standard deviation of indices I and II were greater than those of indices III and IV. Although the median and mean of index II for group B were twice those for group C, the variability in index II due to GRX was larger; thus, this result was nonsignificant.

C. Progressive Evaluation

The total score over the 24 sessions for the patients in groups B and C are displayed in Fig. 6, in which the *x*-axis indicates the number of sessions. The scores increased as the number of sessions increased; however, the difference between groups B and C was small.

The total score estimates according to the linear mixed model are presented in Table V. In this table, GP is a binary variable for which 1 indicates group C and 0 indicates group

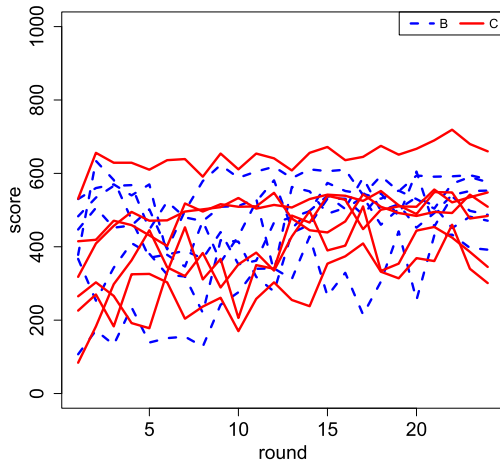


Fig. 6. Scores over 24 sessions for group B (red) and group C (blue).

TABLE V  
ESTIMATES OF TOTAL SCORE FROM LINEAR MIXED MODEL

Effect	Estimate	SE	$p$
Intercept	382.63	62.68	0.00
SR	5.94	1.99	0.00
GP	-16.92	88.64	0.85
SR*GP	0.68	2.82	0.81

B, and SR denotes the number of sessions. The interaction between SR and GP (SR \* GP) was not significant, which indicated that the total score over time was not significantly different between groups B and C. However, higher SR was significantly associated with higher scores. Similar results were observed for the hit rate and maximum wrist turning angle. The difference in the number of games for SR, GP, and SR \* GP was not significant.

A heat map of the  $p$  values obtained for SR, GP, and SR \* GP from the linear mixed model is displayed in Fig. (7); the rows in this figure represent variables, and the columns represent estimates. The labels I\_SR, I\_GP, and I\_SR \* GP indicate the estimates derived from  $\bar{S}_i^I$ ; other elements are defined similarly. The colors indicate the significance level ( $p < 0.05$  was considered significant). The most significant estimates were observed for ACZ, followed by ACY and MAGZ. A significant SR \* GP value indicated that the stability was affected by the group and the number of sessions. The obtained  $\bar{S}_i^{IV}$  values indicated that SR \* GP was significant for ACY, ACZ, and MAGZ. SR \* GP was significant for ACZ, as indicated by the  $\bar{S}_i^{III}$  values. Moreover, SR \* GP was significant for MAGZ, as indicated by the  $\bar{S}_i^I$  and  $\bar{S}_i^{II}$  values. SR was significant for all four stability indices for ACY and ACZ. After the first session, stability was not significantly different between the groups.

The most significant stability indices were derived from ACZ, and many indices derived from MAGZ were significant for SR \* GP. Thus, detailed estimates for ACZ are provided in this paper. Table VI presents the model estimates for indices I and II. The average stability of index II for ACZ was not significantly different between the groups during the first session. SR \* GP was significant, which indicated that the rate of change differed between groups B and C. The rate of

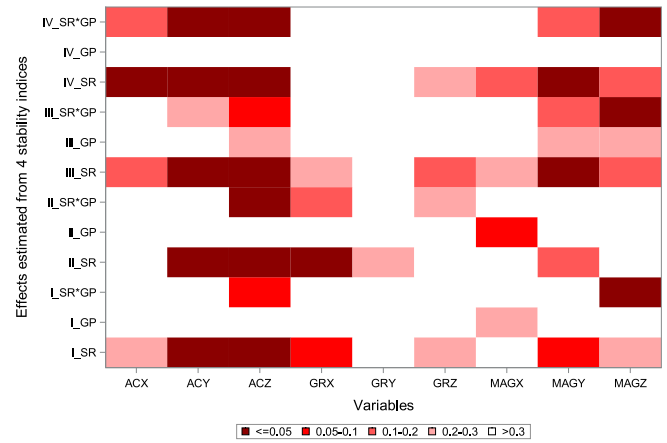


Fig. 7. Heat map of  $p$  values for indices obtained using a linear mixed model.

TABLE VI  
ESTIMATES OBTAINED FOR ACZ FROM  
INDICES I AND II AND THE LINEAR MIXED MODEL

Effect	I			II		
	Est.	SE	$p$	Est.	SE	$p$
Intercept	0.798	0.155	0.000	1.493	0.656	0.024
SR	0.023	0.009	0.010	0.106	0.035	0.002
GP	-0.235	0.219	0.307	-0.759	0.927	0.432
SR*GP	-0.025	0.013	0.053	-0.105	0.049	0.032

TABLE VII  
LINEAR MIXED MODEL ESTIMATES FOR  
ACZ BASED ON INDICES III AND IV

Effect	III			IV		
	Est.	SE	$p$	Est.	SE	$p$
Intercept	0.590	0.097	0.000	0.650	0.369	0.080
SR	0.019	0.009	0.030	0.071	0.010	0.000
GP	-0.174	0.137	0.230	-0.268	0.522	0.619
SR*GP	-0.021	0.012	0.093	-0.075	0.014	0.000

change for group B was 0.106, whereas that for group C was approximately 0.001; these values were derived by combining the effects of SR and SR \* GP. However, for index I, the rate of change in the average stability for group B significantly increased by 2.3%, and the difference in the rate of change between groups B and C was weakly significant ( $p = 0.053$ ). After the effects of SR and SR \* GP were combined, the rate of change in the average stability derived from index I for group C was approximately -0.002; that is, the average stability remained constant as the number of sessions increased.

Table VII presents the model estimates for indices III and IV. For index IV, the rate of change in SR \* GP differed significantly between groups B and C. The rate of change for group B was 0.071. After the effects of SR and SR \* GP were combined, the rate of change for group C was approximately -0.004, which indicated that index IV was constant over the time interval. The average stability of index III for ACZ did not differ significantly during the first session. However, the rate of change in the average stability for group B increased significantly by 1.9%; the difference in the rate of change between groups B and C was nonsignificant.

TABLE VIII  
LINEAR MIXED MODEL ESTIMATES FOR  
MAGZ BASED ON INDICES I AND II

Effect	I			II		
	Est.	SE <sup>§</sup>	<i>p</i>	Est.	SE	<i>p</i>
Intercept	1.046	0.174	0.000	2.328	0.824	0.005
SR	0.011	0.010	0.299	0.033	0.057	0.561
GP	0.175	0.246	0.494	0.142	1.166	0.905
SR*GP	-0.029	0.015	0.045	-0.058	0.080	0.472

<sup>§</sup>: standard error of the estimate

TABLE IX  
LINEAR MIXED MODEL ESTIMATES FOR  
MAGZ BASED ON INDICES III AND IV

Effect	III			IV		
	Est.	SE <sup>§</sup>	<i>p</i>	Est.	SE	<i>p</i>
Intercept	0.809	0.129	0.000	1.358	0.420	0.001
SR	0.011	0.007	0.130	0.036	0.024	0.138
GP	0.204	0.182	0.290	0.495	0.593	0.423
SR*GP	-0.031	0.010	0.002	-0.094	0.034	0.006

<sup>§</sup>: standard error of the estimate

SR \* GP was significantly associated with MAGZ for indices I, III, and IV (Tables VIII and IX). The rate of change for group B with SR was not significantly different from 0; that is, the variability remained constant as SR increased. However, the significance of SR \* GP indicated that the rate of change with SR for group C was different from 0; that is, as SR increased, the stability index and variability decreased. Finally, SR \* GP was not significant for index II because of a large standard error.

#### IV. DISCUSSION

In this study, functional analysis was performed for effectively using sensor data to evaluate a stroke rehabilitation program. Groups in a stroke rehabilitation program were differentiated according to coefficients estimated using a fitted B-spline function. Four indices based on the variation of the B-spline function were adopted to evaluate the rehabilitation program.

Recruiting sufficient suitable participants is a major challenge in clinical research. Factors such as game score, number of games, hit rate, and maximum wrist turning angle in rehabilitation games can be easily measured but do not provide information on intermediate motions. Thus, these factors might be insufficient for evaluating the overall difference between experimental groups from a small sample, as in the study of Lin et al. [13]. However, the indices proposed in this paper, which are based on intermediate data, reveal that, in the study by Lin et al., the patients in group C outperformed those in group B in terms of overall rehabilitation and progress during their rehabilitation program.

Rehabilitation programs include static actions, microactions, and intense actions. The accelerometer data collected by an IMU are most accurate for static actions. By contrast, the gyroscope data collected by an IMU enable the derivation of more accurate measures when microactions and intense actions are executed than when a static action is executed. Thus, filtering procedures are required for generating the IMU

trajectory [21], [28], [29]. The proposed spline smoothing techniques do not require data preprocessing to remove noise. The estimated spline function coefficients were found to be effective features for machine learning. In contrast to statistical parameters such as mean and standard deviation, the coefficients in functional data analysis are derived from a basis function that is fitted to the data, and these coefficients are less influenced by outliers than are such typical statistical parameters. A similar finding was reported by [25], who demonstrated that the coefficients derived from spline functions are excellent features for classifying patients with stroke by their level of motor impairment. Nevertheless, the maximum duration of each game in the study by Lin et al. was 10 s, and only three inner knots (*K*) were selected in the present study to avoid data overfitting. When IMU trajectory tracking is longer, a higher value of *K* can be selected to increase modeling accuracy.

The trajectories derived from the acceleration and magnetic field strength in the *z*-direction were key features. The tilt angle of the MTD was equal to  $\tan^{-1}\left(\frac{ACY}{ACZ}\right)$ . A higher ACZ value resulted in a higher tilt angle, which indicated greater shaking of the MTD and inferior hand function. Similarly, a lower ACZ value indicated better hand function.

The trajectory displayed in Fig. 3 can be attributed to the data collection method. At the beginning of each game, the patient was focused and static; however, motor and concentration impairment might have resulted in the patient being unable to stop immediately when required. Therefore, indices III and IV proved slightly more effective for evaluating the performance of groups B and C than did indices I and II. Furthermore, because the rate of change can be negative, the absolute norm and squared norm were used to compute the indices. Among the four considered indices, index IV was the most effective at detecting differences between groups B and C; this index is commonly used because it has excellent mathematical properties.

Except for [13], most relevant studies have collected data using a cross-sectional method [21], [22], [23]. In particular, in [23], four repetitions were performed, and standard deviation was used to examine trial-to-trial variability. Although follow-up research was conducted in [20], only descriptive statistics such as the average and standard deviation of joint range of motion were determined. The data used in the present study were collected using a sophisticated design; that is, patients were asked to play games in 24 sessions with similar conditions. The trajectories were expected to be similar after a patient became familiar with the game and achieved improved functioning. The aim of the present study was to devise a stability index for assessing improvement in motor ability during a rehabilitation program. Thus, the trajectories for each session were averaged to simplify the construction of the indices. In a future study, the difference between a trajectory and the averaged trajectory for all games in each session can be computed, and an instability index can be defined as the average of this difference. However, when the rehabilitation task is performed only once, such an instability index cannot be computed.



## V. CONCLUSION

Continual rehabilitation is required for the recovery of patients with upper-limb paralysis after stroke. An objective evaluation of rehabilitation progress must be performed to enable therapists to modify the program appropriately. Data collected by IMUs can indicate hand function but are discretized. Therefore, this paper proposes the use of B-splines for extracting the features of IMU trajectories to construct objective indices for evaluating rehabilitation effectiveness.

In the present study, the IMU trajectory data collected by Lin et al. using a VR upper-limb rehabilitation system were used with the proposed method to obtain not only feature information for classifiers but also stability indices for assessing the effectiveness of sensor-based rehabilitation systems. The best stability indices (indices III and IV) can be produced if the trajectory movement is only recorded for the target action. Compared with game-based or therapist-evaluated scores, the proposed method is more objective and effective for evaluating rehabilitation progress.

## REFERENCES

- [1] *About Underlying Cause of Death, 1999–2020*, CDC WONDER Online Database, Centers Disease Control Prevention, Atlanta, GA, USA, 2020. [Online]. Available: <https://wonder.cdc.gov/ucd-icd10.html>
- [2] C. W. Tsao et al., “Heart disease and stroke statistics—2022 update: A report from the American heart association,” *Circulation*, vol. 145, no. 8, pp. e153–e639, Jan. 2022.
- [3] C.-M. Chen, C.-C. Tsai, C.-Y. Chung, C.-L. Chen, K. P. Wu, and H.-C. Chen, “Potential predictors for health-related quality of life in stroke patients undergoing inpatient rehabilitation,” *Health Quality Life Outcomes*, vol. 13, no. 1, Aug. 2015, Art. no. 118.
- [4] R. Stockley, R. Peel, K. Jarvis, and L. Connell, “Current therapy for the upper limb after stroke: A cross-sectional survey of U.K. therapists,” *BMJ Open*, vol. 9, no. 9, Sep. 2019, Art. no. e030262.
- [5] J. Desrosiers, F. Malouin, C. Richards, D. Bourbonnais, A. Rochette, and G. Bravo, “Comparison of changes in upper and lower extremity impairments and disabilities after stroke,” *Int. J. Rehabil. Res.*, vol. 26, no. 2, pp. 109–116, Jun. 2003.
- [6] C. J. Winstein et al., “Guidelines for adult stroke rehabilitation and recovery,” *Stroke*, vol. 47, no. 6, pp. e98–e169, May 2016.
- [7] J. Parker, S. Mawson, G. Mountain, N. Nasr, R. Davies, and H. Zheng, “The provision of feedback through computer-based technology to promote self-managed post-stroke rehabilitation in the home,” *Disab. Rehabil., Assistive Technol.*, vol. 9, no. 6, pp. 529–538, Nov. 2014.
- [8] J. Parker, S. Mawson, G. Mountain, N. Nasr, and H. Zheng, “Stroke patients’ utilisation of extrinsic feedback from computer-based technology in the home: A multiple case study realistic evaluation,” *BMC Med. Informat. Decis. Making*, vol. 14, Jun. 2014, Art. no. 46.
- [9] R. J. Davies et al., “A personalized self-management rehabilitation system for stroke survivors: A quantitative gait analysis using a smart insole,” *JMIR Rehabil. Assistive Technol.*, vol. 3, no. 2, p. e11, Nov. 2016.
- [10] L. Powell, J. Parker, M. M. St-James, and S. Mawson, “The effectiveness of lower-limb wearable technology for improving activity and participation in adult stroke survivors: A systematic review,” *J. Med. Internet Res.*, vol. 18, no. 10, p. e259, Oct. 2016.
- [11] E. Tseklevs, I. T. Paraskevopoulos, A. Warland, and C. Kilbride, “Development and preliminary evaluation of a novel low cost VR-based upper limb stroke rehabilitation platform using Wii technology,” *Disab. Rehabil., Assistive Technol.*, vol. 11, no. 5, pp. 413–422, Jul. 2016.
- [12] S. Jie, Y. Haoyong, T. L. Chaw, C. C. Chiang, and S. Vijayavenkataraman, “An interactive upper limb rehab device for elderly stroke patients,” *Proc. CIRP*, vol. 60, pp. 488–493, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2212827117301373>
- [13] B.-S. Lin, J.-L. Chen, and H.-C. Hsu, “Novel upper-limb rehabilitation system based on attention technology for post-stroke patients: A preliminary study,” *IEEE Access*, vol. 6, pp. 2720–2731, Dec. 2018.
- [14] A. Warland et al., “The feasibility, acceptability and preliminary efficacy of a low-cost, virtual-reality based, upper-limb stroke rehabilitation device: A mixed methods study,” *Disability Rehabil.*, vol. 41, no. 18, pp. 2119–2134, Aug. 2019.
- [15] J. Parker, L. Powell, and S. Mawson, “Effectiveness of upper limb wearable technology for improving activity and participation in adult stroke survivors: Systematic review,” *J. Med. Internet Res.*, vol. 22, no. 1, Jan. 2020, Art. no. e15981.
- [16] D. A. Oladele, E. D. Markus, and A. M. Abu-Mahfouz, “Adaptability of assistive mobility devices and the role of the Internet of medical things: Comprehensive review,” *JMIR Rehabil. Assistive Technol.*, vol. 8, no. 4, Nov. 2021, Art. no. e29610.
- [17] T. Platz, C. Pinkowski, F. van Wijck, I.-H. Kim, P. di Bella, and G. Johnson, “Reliability and validity of arm function assessment with standardized guidelines for the fughl-meyer test, action research arm test and box and block test: A multicentre study,” *Clin. Rehabil.*, vol. 19, no. 4, pp. 404–411, Jun. 2005.
- [18] L. Yu, J.-P. Wang, Q. Fang, and Y. Wang, “Brunnstrom stage automatic evaluation for stroke patients using extreme learning machine,” in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Hsinchu, Taiwan, Nov. 2012, pp. 380–383.
- [19] Q. Wang et al., “Stroke Patients’ acceptance of a smart garment for supporting upper extremity rehabilitation,” *IEEE J. Transl. Eng. Health Med.*, vol. 6, pp. 1–9, Oct. 2018.
- [20] J. P. O. Held et al., “Inertial sensor measurements of upper-limb kinematics in stroke patients in clinic and home environment,” *Frontiers Bioeng. Biotechnol.*, vol. 6, Apr. 2018, Art. no. 27.
- [21] A. Schwarz, M. M. C. Bhagubai, G. Wolterink, J. P. O. Held, A. R. Luft, and P. H. Veltink, “Assessment of upper limb movement impairments after stroke using wearable inertial sensing,” *Sensors*, vol. 20, no. 17, p. 4770, Aug. 2020.
- [22] Z.-J. Chen, C. He, M.-H. Gu, J. Xu, and X.-L. Huang, “Kinematic evaluation via inertial measurement unit associated with upper extremity motor function in subacute stroke: A cross-sectional study,” *J. Healthcare Eng.*, vol. 2021, pp. 1–7, Aug. 2021.
- [23] Y.-W. Chen, W.-W. Liao, C.-L. Chen, and C.-Y. Wu, “Kinematic descriptions of upper limb function using simulated tasks in activities of daily living after stroke,” *Hum. Movement Sci.*, vol. 79, Oct. 2021, Art. no. 102834.
- [24] B.-S. Lin, Y.-S. Lin, I.-J. Lee, and B.-S. Lin, “A real-time missing data recovery method using recurrent neural network for multiple transmissions,” in *Recent Advances in Intelligent Information Hiding and Multimedia Signal Processing*, J.-S. Pan, A. Ito, P.-W. Tsai, and L. C. Jain, Eds. Cham, Switzerland: Springer, 2019, pp. 99–107.
- [25] Y.-T. Hwang, W.-A. Lu, and B.-S. Lin, “Use of functional data to model the trajectory of an inertial measurement unit and classify levels of motor impairment for stroke patients,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 925–935, Apr. 2022.
- [26] A. Perperoglou, W. Sauerbrei, M. Abrahamowicz, and M. Schmid, “A review of spline function procedures in R,” *BMC Med. Res. Methodol.*, vol. 19, no. 46, Mar. 2019, Art. no. 46.
- [27] G. L. Prajapati and A. Patle, “On performing classification using SVM with radial basis and polynomial kernel functions,” in *Proc. 3rd Int. Conf. Emerg. Trends Eng. Technol.*, Goa, India, Nov. 2010, pp. 512–515.
- [28] J.-S. Wang, Y.-L. Hsu, and J.-N. Liu, “An inertial-measurement-unit-based pen with a trajectory reconstruction algorithm and its applications,” *IEEE Trans. Ind. Electron.*, vol. 57, no. 10, pp. 3508–3521, Oct. 2010.
- [29] M. Ortiz-Salazar, A. Rodríguez-Liñán, L. M. Torres-Treviño, and I. López-Juárez, “IMU-based trajectory generation and modelling of 6-DOF robot manipulators,” in *Proc. Int. Conf. Mechatronics, Electron. Automot. Eng. (ICMEAE)*, Nov. 2015, pp. 181–186.