LSTM Network-Based Estimation of Ground Reaction Forces During Walking in Stroke Patients Using Markerless Motion Capture System

Ryo Sugai, Shintaro Maeda, Ryosuke Shibuya, Yusuke Sekiguchi, Shin-Ichi Izumi, Mitsuhiro Hayashibe, *Senior Member, IEEE*, and Dai Owaki, *Member, IEEE*

***Abstract*—We propose a novel approach for estimating ground reaction forces (GRFs) during walking in stroke patients using a markerless motion capture (MMC) system, specifically the Azure Kinect, and a long short-term memory (LSTM) network. GRFs are crucial indicators of walking ability, but their measurement typically requires force plates, which are not readily available in clinical settings. Our study aimed to assess the feasibility of applying artificial neural networks (ANNs) to estimate GRFs in stroke patients using MMC. Our findings demonstrate that the estimated GRFs can serve as reliable clinical indicators of gait ability, with comparable estimation error in the vertical direction for both healthy individuals and stroke patients (L/R: 10.39/9.88% and P/H: 10.70/10.06%). The proposed neural-network-based approach to GRF estimation is more accessible and cost-effective than traditional force plate measurements and has the potential to enhance the development of personalized rehabilitation programs for stroke patients. This research fills a critical gap in the field of medical robotics, providing a practical and innovative method for assessing gait quality, planning, and monitoring rehabilitation in stroke patients.**

***Index Terms*—Waking, stroke, rehabilitation, ground reaction force (GRF), neural network (NN).**

I. Introduction

WALKING is a crucial motor function that plays an essential role in maintaining independence and improving quality of life. Ground reaction forces (GRFs) are measurable indicators of walking ability, making it easier to identify gait abnormalities and create individualized rehabilitation programs [1], [2], [3], [4], [5], [6], [7], [8]. Three-dimensional (3D) GRF information includes the load that supports the body and the magnitude of the braking and propulsive forces, which are delivered by the interaction between the body and the ground. Typically, body kinematics and kinetic data obtained from 3D motion capture and force plates are used together in gait analysis [9], [10], [11], [12], [13]. However, force plates are expensive and not readily available, which limits their use in clinical settings. Therefore, alternative approaches are necessary to obtain GRF data for developing individualized rehabilitation programs [13], [14].

In previous studies, researchers have attempted to estimate GRFs without using expensive force plates, with one common approach being the use of kinematics data obtained from motion capture systems [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]. Optical motion capture (OMC) is frequently used for GRF estimation because of its ability to accurately record kinematic data via multiple cameras and reflective markers directly attached to the human body [15], [16], [17]. However, OMC is limited to laboratory settings with specialized equipment, making it difficult to apply in real-world scenarios, e.g., hospitals. Inertial motion capture (IMC) is a more accessible solution for estimating GRFs because it can acquire kinematic data by IMU sensors attached to the human body [18], [19], [20], [21]. IMC provides significant advantages in measuring human movement and activity in diverse tasks [22], [23] and environments [24], [25]. Furthermore, markerless motion capture (MMC) systems, such as Azure Kinect (Microsoft, USA) and Realsense (Intel, USA), offer the potential to estimate GRFs without markers or specialized equipment. MMC is more accessible and cost-effective and does not interfere with the subject's natural movements [26], [27]. The use of MMC for estimating GRFs is expected to further improve accessibility because of its portable and stand-alone configuration [28], [29], [30], [31], [32]. However, it is essential to investigate the accuracy and reliability of these approaches compared to the "gold standard" of force plate measurement, particularly in clinical settings [14], [33], [34].

Advanced artificial neural networks (ANNs), such as feed-forward neural networks (FFNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks [35], enable GRF estimation from kinematic data. These ANNs can

Manuscript received 31 March 2023; revised 21 June 2023; accepted 13 July 2023. Date of publication 30 August 2023; date of current version 27 October 2023. This article was recommended for publication by Associate Editor A. Padilha Lanari Bo and Editor P. Dario upon evaluation of the reviewers' comments. This work was supported in part by the Tateishi Science and Technology Foundation (2023, Research Grant A) under Grant 2231006, and in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP19K22855 and Grant JP20H04260. *(Corresponding author: Dai Owaki.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Tohoku University, Graduate School of Medicine under Application No. 2022-1-194.

Ryo Sugai, Shintaro Maeda, Mitsuhiro Hayashibe, and Dai Owaki are with the Department of Robotics, Graduate School of Engineering, Tohoku University, Sendai 980-8579, Japan (e-mail: owaki@tohoku.ac.jp).

Ryosuke Shibuya and Yusuke Sekiguchi are with the Department of Physical Medicine and Rehabilitation, Tohoku University Graduate School of Medicine, Sendai 980-8575, Japan.

Shin-Ichi Izumi is with the Department of Physical Medicine and Rehabilitation, Tohoku University Graduate School of Medicine, Sendai 980-8575, Japan, and also with the Graduate School of Biomedical Engineering, Tohoku University, Sendai 980-8575, Japan.

Digital Object Identifier 10.1109/TMRB.2023.3310196

learn walking biomechanics from supervised data, including the relationship between kinematics (input) and kinetics (output) data, which enables GRF estimation for gait analysis. Whereas an FFNN processes the entire gait cycle at once, RNNs and LSTMs can learn the time dependencies of time-series data, making them useful for the estimation of not only GRFs but also biomechanical features. Previous studies have demonstrated the feasibility of using ANNs to estimate GRFs, but the majority of these studies have focused on healthy subjects [15], [16], [19], [25].

In contrast, the aim of our study was to apply ANNs to estimate GRFs in stroke patients who often exhibit abnormal gait patterns [1], [4], [5], [7], [8]. Additionally, whereas researchers of previous studies used various types of motion capture systems, such as optical and inertial systems (OMC/IMC), we aimed to use an MMC system, specifically the Azure Kinect, to enable the estimation of GRFs in a more accessible and cost-effective manner [28]. To the best of our knowledge, no previous studies have investigated the estimation of GRFs in stroke patients using an MMC system and ANNs. Therefore, our method provides a novel approach to estimating GRFs in patients with gait disorders using cutting-edge technology and has the potential to contribute to the development of more effective clinical assessments and interventions for these patients.

II. Materials and Methods

A. Subjects

Sixteen healthy subjects (F/M: 7/9, age: 62.5 ± 9.5 years, height: 160.1 ± 8.8 cm, weight: 58.5 ± 12.3 kg) and seven hemiplegic patients (F/M: 2/5, age: 66.3 ± 9.6 years, height: 165.1 ± 11.0 cm, weight: 60.5 ± 10.7 kg) were recruited from Tohoku University Hospital. Subjects were excluded from the study if they had a history of osteoarticular disease of the lower extremity. The subjects recruited were of 20 years of age and older and were required to be able to walk independently for at least 7 m. We assessed hemiparetic patients according to hemiparesis severity and the ability to perform movements outside the extensor and flexor synergy patterns using the Brunnstrom recovery stages [36]. The Brunnstrom stage test is a standardized index employed by physical therapists to track and measure a patient's motor recovery progression following a stroke, ranging from initial paralysis to regained fine motor control. Experienced physical therapists (Y.S. and R.S.) performed these tests while applying standardized protocols. All subjects provided signed informed consent, and the protocol was approved by the Ethics Committee of Tohoku University, Graduate School of Medicine (2022-1-194).

B. Data Collection

In this study, we used an MMC device (Azure Kinect DK, Microsoft Corp., Redmond, WA, USA) and four 90 cm × 60 cm force plates (Anima Corporation, Chofu, Tokyo, Japan). We used the Azure Kinect Body Tracking SDK v1.1.0 (distributed by Microsoft, [37]) to acquire 32 joint positions. To verify the data accuracy from the Kinect, we used measurement data from an OMC (MAC 3D System, Motion Analysis Corporation, Santa Rosa, CA, USA) consisting of eight cameras as a reference. Reflective markers were attached at the body part locations of subjects shown in Table I [38]. The Azure Kinect, force plate, and OMC sampling frequencies were set to 30 Hz, 1200 Hz, and 120 Hz, respectively. Fig. 1 shows the experimental environment, in which the OMC camera is highlighted with blue dotted squares, the force plates with a red dotted square, and the Azure Kinect with a yellow dotted square. The three-axis (mediolateral, anterior-posterior, and vertical direction as shown in Fig. 1) ground reaction forces (GRFs) were obtained from the force plates. The Azure Kinect was placed approximately 8 m away from the starting point of the walk. To synchronize the kinematic data of the subjects, we started recording Azure Kinect and OMC data simultaneously. Subjects were asked to walk at a comfortable walking speed on a 7 m outward path with embedded force plates (6–12 trials for each subject). We measured the walking task so that each trial included data from at least one step on both sides (for patients, the paretic and non-paretic sides).

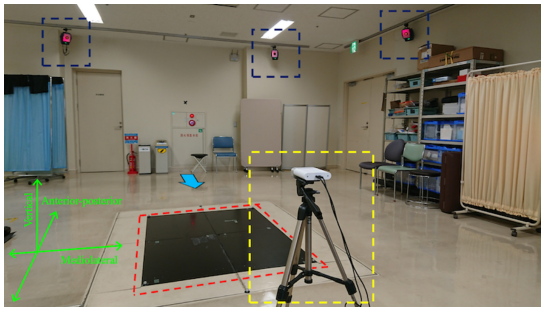


Fig. 1. Experimental environment. Eight OMC cameras (blue dotted squares are three of them), the four 90 cm × 60 cm force plates (red dotted square), and the Azure Kinect (yellow dotted square). The Azure Kinect was placed approximately 8 m away from the starting point of the walk. The three-axis (mediolateral, anterior-posterior, and vertical direction) GRFs were obtained from the force plates.

TABLE I
Reflective Marker Location Attached at Subject Body Parts for the OMC System [38]

Segment	Placement of Markers
Trunk	spinous process of the 7th cervical vertebrae, spinous process of the 10th thoracic vertebrae, deepest point of incisura jugularis, most caudal point of the sternum, xiphoid process of the sternum, the position in the middle of the right scapula,
Upper arm	both acromia,
Pelvis	both anterior superior iliac spines, both posterior superior iliac spines,
Thigh	both prominence of the greater trochanters, both the lower lateral 1/3 surface of the thigh, both the lateral femoral epicondyles, both the medial femoral epicondyles,
Shank	both prominence of the lateral malleolus, both prominence of the medial malleolus, both the lower 1/3 of the shank,
Foot	both dorsal aspect of the second metatarsal heads, both aspect of the achilles tendon insertion on the calcaneus,

C. Signal Processing and Data Preparation

We applied fourth-order Butterworth low-pass filtering to the measured data to remove noise effects. The cutoff frequency was set to 6 Hz for the Azure Kinect and OMC, and 10 Hz for the force plate. After resampling all data to 120 Hz, we synchronized the signals. A coordinate transformation was performed from a coordinate system fixed at the position of the Azure Kinect (depth camera) to a coordinate system defined at the subject's pelvis.

In this study, the GRFs were normalized by the subject's body weight (BW) and gravitational acceleration:

$$\overline{\mathbf{GRF}} = \frac{\mathbf{GRF}}{\mathrm{BW} \cdot g}, \tag{1}$$

where the GRF units were converted from [N] to [BW]. The acceleration of gravity g was set as 9.80665. The BW value 1.0 was the GRF value corresponding to gravity relative to body weight. The GRFs during the swing phase were calculated as zero.

D. RNN: Long Short-Term Memory Network

Because the GRFs are time-series data, Recurent Neural Networks (RNNs) were used as the neural network model for estimation. RNN is a type of neural network that can transfer information in time [39]. RNNs have a self-referential (self-loop) structure, which allows them to retain past information and carry it over to the next computational step. In particular, RNNs exhibit excellent performance in predicting time-series data, for example, in speech recognition and natural-language processing [40]. In this study, we used LSTM [35], which is a type of RNN. In RNNs, a longer series results in more weights that are multiplied and more gradients that disappear, resulting in difficulties associated with learning long-term dependence. LSTM overcomes these difficulties by backpropagating the error without applying weights. A schematic of an LSTM cell is shown in Fig. 2.

LSTM follows an RNN architecture with added features, namely the memory cell (cell state) and gate, which are not found in the standard RNN model. The memory cell stores and transfers past information, which enables the prediction of long-term dependency data. Gates are categorized into three types, namely the forget gate (f_t), input gate (i_t), and output gate (o_t). Each gate modifies the information stored in the memory cell or determines the output from the LSTM layer. The features of each gate are described as follows.

- Forget gate: deletes needless information from the memory cell, according to (2)
- Input gate: controls how much new information is added to the memory cell, according to (3) and (4)
- Output gate: assigns a weight to the modified information and calculates the LSTM output, according to (5)

The computations carried out by these gates are expressed by

$$f_t = \sigma\left(x_t W_x^{(f)} + h_{t-1} W_h^{(f)} + b^{(f)}\right) \tag{2}$$

$$\tilde{c}_t = \tanh\left(x_t W_x^{(g)} + h_{t-1} W_h^{(g)} + b^{(g)}\right) \tag{3}$$

$$i_t = \sigma\left(x_t W_x^{(i)} + h_{t-1} W_h^{(i)} + b^{(i)}\right) \tag{4}$$

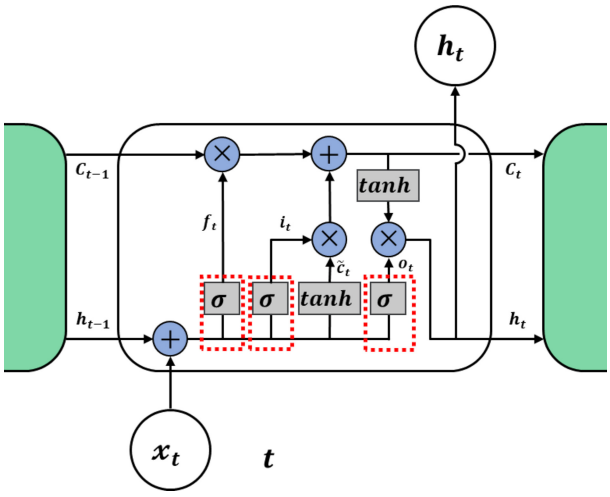


Fig. 2. Schematic of a single LSTM cell. Memory cell (cell state, c_t), which stores and transfers past information, enabling the prediction of long-term dependency data. Gates are categorized into three types: the forget gate (f_t), input gate (i_t), and output gate (o_t). Each gate modifies the information stored in the memory cell or determines the output from the LSTM layer.

TABLE II
INPUT JOINT POSITIONS AND OUTPUT GRF DATA FOR THE LSTM MODEL

Input	NECK, SPINE_CHEST, SHOULDER(R/L), ELBOW(R/L), WRIST(R/L), KNEE(R/L), ANKLE(R/L), FOOT(R/L)
Output	3-axis ground reaction forces on both legs

$$o_t = \sigma\left(x_t W_x^{(o)} + h_{t-1} W_h^{(o)} + b^{(o)}\right), \tag{5}$$

where $\sigma(\cdot)$ is the sigmoid function, x is the input vector, h is the output vector, W is the weight matrix for x and h, and b is the bias at each gate computation. Each superscript indicates the gate type, and each subscript indicates the time step. The computation of the cell state (c_t) and the output of the LSTM model (hidden state h_t) at times t are determined by (6) and (7), respectively.

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \tag{6}$$

$$h_t = o_t \odot \tanh(c_t), \tag{7}$$

where $\odot$ denotes Hadamard product.

E. Model Building and Optimization

Table II includes the joint positions acquired by the Azure Kinect, which were used as input for the neural network. The measured GRFs from force plates were used as the supervised data to train the network. We optimized the LSTM model architecture, such as the number of layers, number of units for each layer, and dropout rate for each layer, and the hyperparameters, such as learning rate, batch size, window width, and its optimizer, as shown in Table III, by using Optuna (Preferred Networks, [41]), a python library capable of performing Bayesian optimization. We calculated an evaluation index of optimization using a 10-fold cross-validation with the Mean Absolute Error (MAE) described by the following equation:

$$\mathrm{MAE}_j = \frac{1}{N_j} \sum |y_p(t) - y_m(t)|, \tag{8}$$

TABLE III
HYPERPARAMETER RANGE OF LSTM MODEL ARCHITECTURE FOR OPTIMIZATION WITH OPTUNA [41]

Hyperparameter	Min	Max	interval
Number of hidden layers	1	3	1
Number of hidden layer units	8	256	4
Dropout rate	0.2	0.5	0.05
Window width	1	50	1
Hyperparameter	**options**		
Batch size	64, 128, 256, 512, 1024, 2048		
Learning rate	1e-4, 5e-4, 1e-3, 5e-3, 1e-2, 5e-2, 0.1		
Optimizer	SGD, Adagrad, RMSprop, Adam		

TABLE IV
OPTIMIZED PARAMETERS OF LSTM MODEL ARCHITECTURE

Hyperparameter	optimized value
Number of hidden layers	2
Number of hidden layer units 1	216
Number of hidden layer units 2	76
Dropout rate 1	0.20
Dropout rate 2	0,25
Window width	27
Batch size	64
Learning rate	5e-3
Optimizer	Adam

where N_j, $y_p(t)$, $y_m(t)$ represent the number of evaluation data for a fold, the estimated GRFs, and measured GRFs, respectively. We mixed a data set of healthy subjects and stroke patients in the LSTM network training. In the 10-fold cross-validation, we divided the dataset into ten subsets by subject, and each subset contains data from one to three subjects to avoid bias in the subject data. We used nine of the ten subsets as training data to train the LSTM network and calculated the evaluation value MAE_j by using the remaining one subset as validation data. This calculation for each fold was performed ten times, changing the subset used as the validation data. We finally calculated the average MAE of the ten times as the evaluation value for one hyperparameter set. In this study, we performed this optimization through 70 iterations.

F. Performance Evaluation

We built the LSTM model with the hyperparameters with the highest scores during the optimization. The hyperparameters optimized using Optuna are shown in Table IV. The dropout rates 1 and 2 represent the rates of the dropout layers located after the hidden layers 1 and 2, respectively. Fig. 3 shows the LSTM neural network model constructed based on these hyperparameters. Using the structure of this model, we trained the neural network weights to estimate the GRFs of healthy and stroke hemiplegic patients.

For evaluating model performance, we also used a 10-fold cross-validation for a mixed data set of healthy subjects and stroke patients. The correlation coefficient r and the normalized root mean square error (nRMSE [%]) were employed as evaluation metrics, expressed by

$$r = \frac{\frac{1}{N}\sum_{t=1}^{N}\left(y_p(t)-\hat{y_p}(t)\right)\left(y_m(t)-\hat{y_m}(t)\right)}{\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y_p(t)-\hat{y_p}(t)\right)^2}\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y_m(t)-\hat{y_m}(t)\right)^2}}, \tag{9}$$

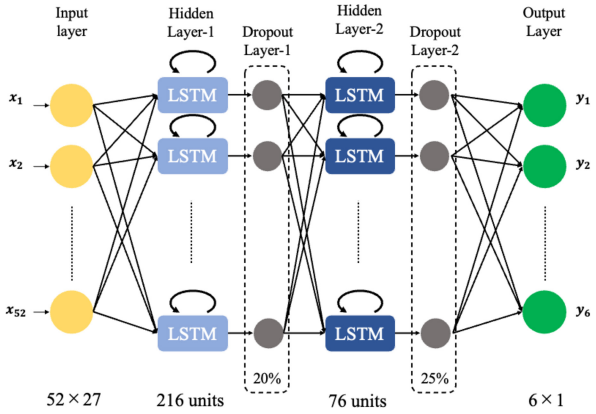


Fig. 3. Optimized LSTM neural network structure for GRF estimation.

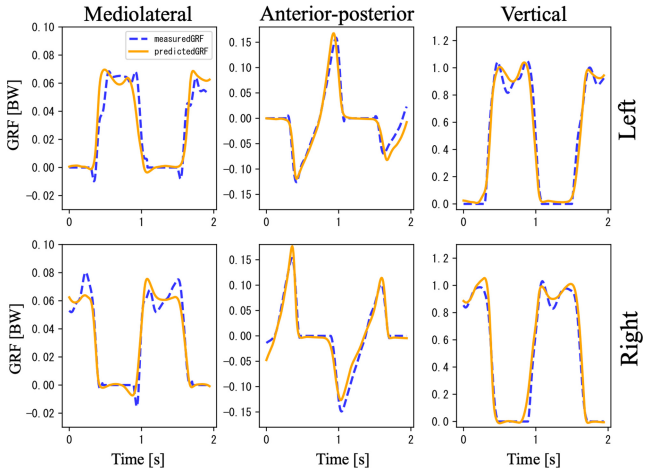


Fig. 4. Comparison of estimated GRFs (orange solid lines) and measured GRFs (blue dotted lines).

$$\mathrm{nRMSE} = \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y_p(t)-y_m(t)\right)^2}}{\max(\mathrm{y_m(t)})\text{-}\min(\mathrm{y_m(t)})} \times 100, \tag{10}$$

where $y_p(t)$ and $y_m(t)$ represent the estimated and measured GRF, respectively. Hat symbols represent average values.

G. Statistical Analysis

We conducted statistical analysis using two-way ANOVA with side (2 levels: right and left sides for healthy subjects; paretic and non-paretic sides for hemiplegic patients) as a within-subject factor and group (2 levels: healthy subjects and hemiplegic patients for Fig. 5/ Tab. VI or 3 levels: Brunnstrom stage IV, V, VI for Fig. 7) as an independent factor. When we found a significant difference, we conducted Tukey's Honestly Significant Difference (HSD) test. The statistical significance level was set at a p value of 0.05 for all analyses. We calculated the partial eta squared (η_p^2) and Cohen's d as an estimate of the effect size. Statistical analyzes were performed using a statistical software package (R ver 4.3.0, The R Foundation).

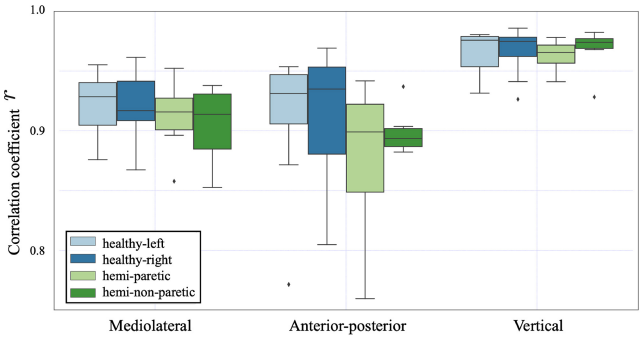
III. RESULTS

Fig. 4 shows the estimated GRFs with the LSTM model and the measured GRFs. We found that the vertical GRF was

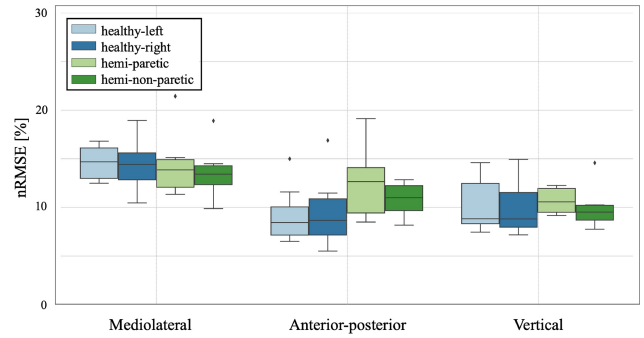
TABLE V
3-AXIS GRF ESTIMATION ACCURACY r AND ESTIMATION ERROR NRSME FOR HEALTHY AND HEMIPLEGIC SUBJECTS

		healthy subjects		hemiplegic patients	
		left	right	paretic	non-paretic
Medio-lateral	r	0.920 (0.027)	0.919 (0.031)	0.911 (0.032)	0.905 (0.034)
	nRMSE [%]	14.59 (1.66)	14.62 (2.69)	14.54 (3.68)	13.70 (3.01)
	difference of nRMSE [%]	0.03		0.84	
Anterior–posterior	r	0.911 (0.055)	0.916 (0.054)	0.877 (0.068)	0.899 (0.020)
	nRMSE [%]	**9.15** (2.63)	**9.30** (3.34)	12.64 (3.99)	**10.81** (1.82)
	difference of nRMSE [%]	0.15		1.83	
Vertical	r	**0.965** (0.018)	**0.967** (0.020)	**0.963** (0.013)	**0.967** (0.020)
	nRMSE [%]	**10.39** (2.74)	**9.88** (2.71)	**10.70** (1.37)	**10.06** (2.40)
	difference of nRMSE [%]	0.51		0.64	

The values in the table represent mean values. Values in parentheses represent standard deviations.



(a) Correlation coefficients r



(b) nRMSE [%]

Fig. 5. Distribution of correlation coefficients r and nRMSE for the left and right legs for healthy subjects and the paretic and non-paretic sides for hemiplegic patients. Estimation errors (nRMSE) were comparable in the vertical direction for both healthy individuals and stroke patients (L/R: 10.39/9.88% and P/H: 10.70/10.06%). Black points indicate outliers.

able to reproduce the two peaks. The estimated GRF during the swing phase was close to zero. Table V indicates the accuracy and error for estimating the GRFs calculated by the correlation coefficient r and nRMSE. The values in the table represent mean values. Values in parentheses represent standard deviations. Fig. 5 shows the distribution of correlation coefficients r and nRMSE for the left and right legs for healthy subjects and the paretic and non-paretic sides for hemiplegic patients.

The results in Tab. V indicates that the accuracy of the paretic side was worse in all directions. Compared to healthy subjects, hemiplegic patients show the largest change in estimation error in the anterior–posterior direction, worsening by approximately 3.5%. Hemiplegic patients have correlation coefficients below 0.90 in the anterior–posterior direction, 0.877 (paretic side) and 0.899 (healthy side), indicating lower estimation accuracy than in the other directions. In contrast, the estimation errors in the mediolateral and vertical directions show no significant difference between healthy subjects and hemiplegic patients.

Next, we performed a verification on the anterior-posterior and vertical GRF peaks as shown in Fig. 6. The two peaks of the anterior-posterior GRF are called the peak in breaking and peak in propulsion, respectively. The two peaks of the vertical GRF are called the 1^{st} peak and 2^{nd} peak, respectively. We evaluated the estimation error of the peak value as the ratio of the estimated value to the measured (true) value:

$$\text{Amplitude Ratio}[\%] = \frac{y_{peak}}{\hat{y}_{peak}} \times 100, \tag{11}$$

Table VI lists the estimation error of the peak values. We found that the estimated vertical GRF peak was smaller than the measured value. The ratio of the vertical peak values was greater than 90% (92.22–99.85%), and the standard deviation was less than 10% (4.57–7.95%). Peak values on the anterior–posterior axis averaged 89–126%. The standard deviations were approximately 20%, indicating a large variability in the estimated values. Here, the mediolateral GRF was excluded from this analysis because the nRMSE accuracy exceeded 13% as shown in Tab. V.

We investigated the relationship between the degree of paralysis and estimation accuracy. To evaluate the degree of paralysis, we used the data from the Brunnstrom test that was performed before the gait measurement. The Brunnstrom test classifies the degree of paralysis recovery into six stages, called Brunnstrom stages [42]. A higher number indicates less-severe paralysis. In this study, two patients were in Stage IV, three in Stage V, and two in Stage VI. Fig. 7 shows the comparison of estimation errors for each Brunnstrom stage. The

TABLE VI
ANTERIOR-POSTERIOR AND VERTICAL PEAK GRF ESTIMATION ERROR FOR HEALTHY AND HEMIPLEGIC SUBJECTS

		healthy subjects		hemiplegic patients	
		left	right	paretic	non-paretic
Anterior–posterior	Peak in propulsion [%]	84.1 (18.2)	93.0 (15.6)	126.4 (53.7)	**98.9** (16.4)
	Peak in braking [%]	90.1 (18.3)	88.9 (27.8)	88.7 (22.9)	108.0 (29.2)
Vertical	1^{st} peak [%]	**95.43** (**7.75**)	**95.81** (**6.71**)	92.22 (**7.95**)	**96.68** (**6.00**)
	2^{nd} peak [%]	**97.32** (**5.65**)	**97.97** (**5.70**)	**98.37** (**4.57**)	**99.85** (**6.18**)

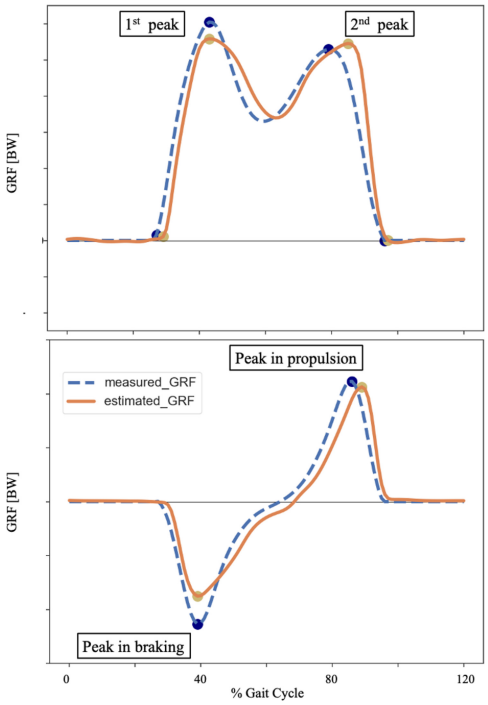


Fig. 6. Definition of GRF peak. Upper: Schematic of vertical GRF; the two peaks are called the 1^{st} peak and 2^{nd} peak, respectively. Lower: Schematic of anterior-posterior GRF; the two peaks are called the peak in breaking and peak in propulsion, respectively. The dotted blue and orange lines indicate the measured and estimated GRFs, respectively.

results indicate that a lower Brunstrom stage, i.e., the more-severe paralysis (Stage IV), tend to result in a larger estimation error.

As results of statistical analysis on the sides (right and left sides for healthy subjects; paretic and non-paretic sides for hemiplegic patients) and group (healthy subjects and hemiplegic patients) related to Fig. 5, we found no significant main effect of the sides on correlation coefficient r and nRMSE for all three-axis GRFs. A significant main effect of group on nRMSE for anterior-posterior GRF was observed ($F(1, 28) = 5.21$, $p = 0.0302$, $\eta^2 = 0.16$), showing nRMSE smaller for healthy subjects than for hemiplegic patients.

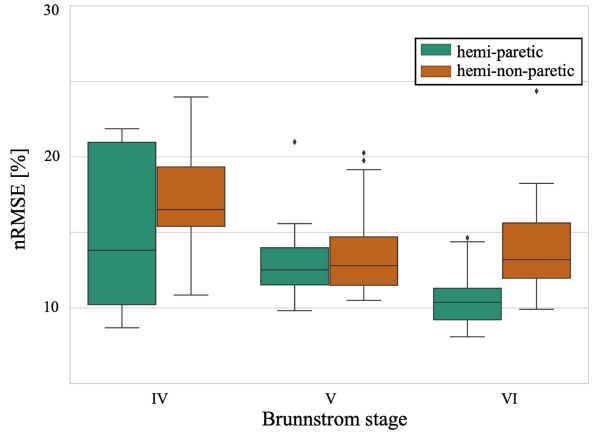
As results of statistical analysis related to Tab. VI, we found significant main effects of the group ($F(1, 226) = 42.7$, $p < 0.001$, $\eta^2 = 0.16$) as well as the interaction between the group and sides ($F(1, 226) = 23.387$, $p < 0.001$, $\eta^2 = 0.094$) on the peak amplitude of anterior-posterior GRF in propulsion. The peak amplitude in propulsion in hemiplegic patients was higher than in healthy subjects. Moreover, in hemiplegic patients, the peak amplitude in propulsion on the paretic side was higher than that on the non-paretic side. We found significant main effects of the group ($F(1, 226) = 6.87$, $p < 0.01$, $\eta^2 = 0.030$) and the interaction between the group and sides ($F(1, 226) = 8.45$, $p < 0.001$, $\eta^2 = 0.036$) on the peak amplitude of anterior-posterior GRF in breaking, showing the peak amplitude in breaking on the non-paretic side was higher than that on the paretic side. We found significant main effects of the sides ($F(1, 225) = 4.13$, $p = 0.0432$, $\eta^2 = 0.018$) and the interaction between the group and sides ($F(1, 225) = 4.01$, $p = 0.0464$ $\eta^2 = 0.018$) on the 1^{st} peak amplitude of vertical GRF, showing the 1^{st} peak amplitude of vertical GRF on the non-paretic side higher than that on paretic side in hemiplegic patients.

As results of statistical analysis on the sides (paretic and non-paretic sides) and group (Brunnstrom stage IV, V, VI) on hemiplegic patients related to Fig. 7, we found significant main effects of the group on nRMSE for mediolateral GRF ($F(2, 76) = 5.89$, $p = 0.00419$, $\eta^2 = 0.13$), anterior-posterior GRF ($F(2, 76) = 4.45$, $p = 0.0149$, $\eta^2 = 0.11$), and vertical GRF ($F(2, 76) = 4.70$, $p = 0.0119$, $\eta^2 = 0.11$), showing nRMSE higher for lower Brunstrom stage patients (Stage IV). A significant main effect of the sides on nRMSE for mediolateral GRF was observed ($F(1, 76) = 7.15$, $p = 0.00917$, $\eta^2 = 0.086$), showing nRMSE higher for mediolateral GRF on the non-paretic side.

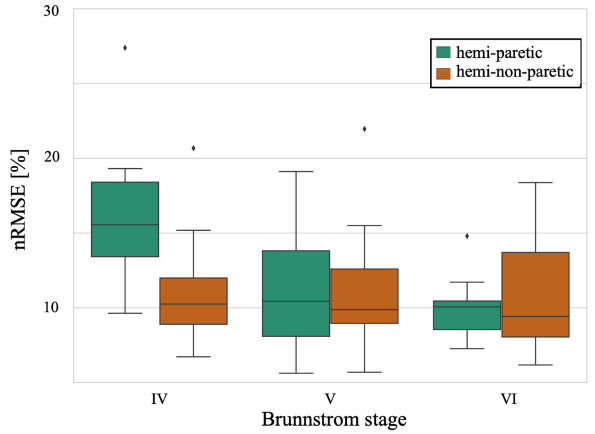
The results of Tukey's Honestly Significant Difference (HSD) test for the Brunnstrom stage group (IV, V, VI) are as follows. nRMSE of mediolateral GRF for group IV was higher than those for the V and VI groups ($p = 0.0241$, $d = 0.80$; $p = 0.00519$, $d = 0.89$). nRMSE of anterior-posterior GRF for group IV was higher than those for the V and VI groups ($p = 0.0431$, $d = 0.66$; $p = 0.0174$, $d = 0.86$). nRMSE of vertical GRF for group IV was higher than those for the VI group ($p = 0.00914$, $d = 1.2$).

IV. DISCUSSION

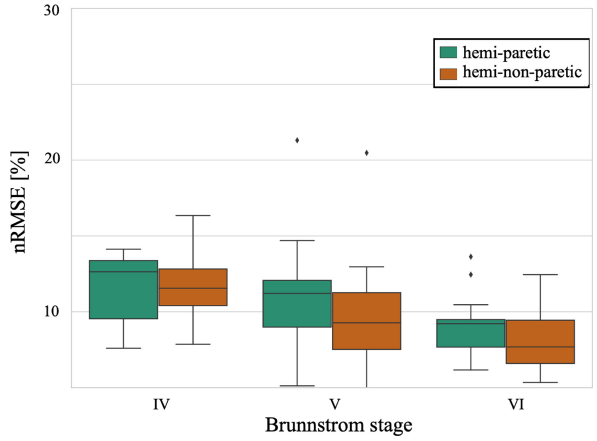
This study was aimed at estimating 3D GRFs during walking in stroke patients with high accuracy by combining MMC with a depth camera and RNNs. Because the estimation of GRFs for healthy subjects only is not conducive to actual gait analysis, we verified the estimation accuracies of the GRFs for both healthy subjects and hemiplegic patients.



(a) Mediolateral directions.



(b) Anterior-posterior directions.



(c) Vertical directions.

Fig. 7. Comparison of estimation errors in the three directions for patients in each Brunnstrom stage. A lower Brunnstrom stage, i.e., more-severe paralysis Stage IV, resulted in a larger estimation error for all three-axis GRFs due to their asymmetrical GRFs. Black points indicate outliers.

In the data for hemiparetic patients, we found that the GRF in the anterior–posterior direction compared to the vertical direction was not well estimated (Tab. V and Fig. 5). A possible cause is reduced muscle activity on the paretic side [10], [43], [44], [45]. Chen and Patten [11] reported that hemiplegic patients have little plantar muscle work on the paretic side compared to healthy subjects. Jessica et al. [12] reported that the plantar flexion moment on the non-paretic side increased to compensate for the reduced plantar muscle activity on the paretic side, contributing to anterior propulsion. Another possible cause is that the learned neural network would not be expressive enough. We assumed that the LSTM was unable to sufficiently learn such GRF patterns, resulting in a larger error. In this case, the information from the Kinect's skeletal model alone could not sufficiently account for muscle strength and reduced propulsive force. The estimation may have been influenced by the training data of healthy subjects with relatively large peak values and stroke patients with relatively lower plantar muscle work on the paretic side, thus deviated from the actual measured values. We consider another possible cause as the poor accuracy of the Kinect's foot position estimation. We could improve the estimation accuracy by increasing the accuracy of position information by adding other sensors, e.g., attaching IMU sensors [20], [22], [23], [24], [25] to the feet or EMG sensors [46], [47] that measure muscle activity, or by adding information from a musculoskeletal model [30], [31], [32] that can simulate muscle activity. A further possible approach to improve the prediction performance of the LSTM itself is to extend the model to the Hammerstein LSTM (H-LSTM) model [48], [49]. In time-series prediction, the variables at time step t depend nonlinearly on the variables in the past time series $t-n,\ldots,t-2,t-1$. The nonlinear relationship underlying time-series data limits the prediction ability of the standard LSTM. To solve this problem, the H-LSTM model was proposed [48], [49], which consists of two components, a static nonlinear and a dynamic linear module, which are responsible for handling nonlinear transformations and time-series prediction, respectively. To apply this model to the time-series prediction, we can use neural networks as the static nonlinear and dynamic linear modules. We have reported an accuracy improvement in cricket walking motion prediction using the H-LSTM model [50].

The vertical GRF achieved the best estimation accuracy among the three directions. The peak vertical GRF value was related to the subject's propulsive force [10], [51], [52], suggesting that the estimated GRFs could be used as an indicator of gait ability. Furthermore, we found that a higher Brunnstrom stage (that is, less-severe paralysis) tended to result in a smaller estimation error. The reason is that a lighter degree of paralysis results in more-symmetrical GRFs on both legs. The GRF patterns were similar to those of a healthy subject, which made the prediction easier.

The mean absolute error (MAE), commonly used as a method of training machine-learning models, carries inherent limitations when addressing issues such as data distribution, hyperparameter overfitting, time dependency, and generalization to different populations [53], [54]. Differences in data distribution can cause significant discrepancies in prediction accuracy between training data and data from a new population. Models utilizing MAE as a loss function tend to be susceptible to hyperparameter overfitting, making appropriate selection and adjustment of hyperparameters crucial. Furthermore, MAE, by treating each data point independently, may have limitations in predicting time-series data with temporal correlations. Moreover, because MAE treats all prediction

errors equally, it can be challenging to adequately evaluate errors with different impacts in a new population. To resolve these issues, other approaches should be considered, such as a combination of loss functions and evaluation metrics other than MAE, tuning of hyperparameters, or cross-validation in different populations.

In this study, the neural network was trained by a mixed data set of healthy subjects and stroke patients, with the expectation of improving the neural network's expressive ability and estimation performance for unknown subject data. Neural networks could not estimate GRFs in the mediolateral direction with high accuracy, because GRFs in the mediolateral direction vary widely from person to person. Moreover, we also consider the reason as the small data set size. The small data set size may have caused the neural network to overfit the gait patterns of the subject data used for training, resulting in a loss of generalization performance to unknown data. A larger data set with more subjects would be required to train a more generalized model.

For the measurements in this experiment, we instructed the subjects to walk at a comfortable walking speed without specifying a walking speed. Because the GRF pattern varies with walking speed, a more general estimation model should be trained by varying the speed condition step by step. Furthermore, it is necessary to consider a wide variety of variables such as body weight, body height, gait patterns, foot patterns, healthy subjects or hemiplegic patients, age groups, and gait speeds. In our present model, we normalize the GRFs of the supervised data according to the body weight of each subject using (1). Thus, by measuring the subject weight before the experiment and multiplying it by the output predicted GRF, the model can accommodate and estimate the individual differences in body weight. The length information of body segments related to body height was included in the relative positional relationships of the joint position data obtained from the Kinect. Thus, although we did not explicitly consider the relationship between height and input/output data, as a result, we assumed that information about height was included in the Kinect input data and that predictions related to this information were implemented. By using body weight and/or height as input data, we can develop an advanced prediction model that accounts for body length and weight information. LSTM is an RNN that can learn time-series dependencies. Because the Kinect joint position time-series data used as input includes information on gait and foot patterns, we assumed that the prediction model incorporated these patterns.

As discussed in the limitations, the accuracy of the prediction depends on the accuracy of the distal segment position; hence, how the input data accurately obtained the foot patterns is an issue that needs to be improved. To consider the wide variety of variables, we need more training data to construct a model that accounts for these variables, which can be a more challenging research issue from the viewpoint of data collection efforts. In this study, as a first step, we built a model to predict three-axis GRFs in healthy subjects and patients, considering the extent to which the currently available data on healthy subjects and hemiplegic patients can be effectively utilized. The authors would like to emphasize in this study that, as a standard prediction model, we have shown that we can successfully apply the prediction model to a diverse range of subjects, including hemiplegic patients. We expect that adding more data from various perspectives and constructing a GRF prediction model that considers such variables as inputs will inform a more accurate and personalized prediction model construction [55]. We plan to address these issues as future research topics. Finally, a limitation of the markerless motion capture system (MMC) that we should mention here is that its application is limited to a subgroup of post-stroke patients. In other words, not all stroke patients can walk independently, and an MMC system cannot obtain appropriate motion data for patients who require assistive devices (such as crutches or canes) or assistance from a clinician when walking.

V. Conclusion

Our study employed an Azure Kinect device, a depth camera, and a recurrent neural network (RNN) to estimate ground reaction forces (GRFs) during walking in both healthy individuals and hemiparetic stroke patients. By comparing the accuracy of the estimation between these two groups, we assessed the feasibility of applying this method to clinical gait analysis. The findings demonstrate that the estimated GRF can serve as a clinical indicator of gait ability. Although the errors for both groups were largest in the mediolateral direction, the estimation results for hemiplegic patients were comparable to those of healthy subjects in the vertical direction, which is a significant factor in clinical gait analysis. Improving the accuracy of the estimation will facilitate the application of the proposed neural-network-based approach to GRF estimation for personalized rehabilitation programs.

References

[1] S. Morita, H. Yamamoto, and K. Furuya, "Gait analysis of hemiplegic patients by measurement of ground reaction force," *Scand. J. Rehabil. Med.*, vol. 27 1, pp. 37–42, Mar. 1995.

[2] J. L. McCrory, S. C. White, and R. M. Lifeso, "Vertical ground reaction forces: Objective measures of gait following hip arthroplasty," *Gait Posture*, vol. 14, no. 2, pp. 104–109, 2001.

[3] T. H. Cruz and Y. Y. Dhaher, "Impact of ankle-foot-orthosis on frontal plane behaviors post-stroke," *Gait Posture*, vol. 30, pp. 312–316, Oct. 2009.

[4] A. M. Muniz and J. Nadal, "Application of principal component analysis in vertical ground reaction force to discriminate normal and abnormal gait," *Gait Posture*, vol. 29, no. 1, pp. 31–35, 2009.

[5] J. L. Allen, S. A. Kautz, and R. R. Neptune, "Forward propulsion asymmetry is indicative of changes in plantarflexor coordination during walking in individuals with post-stroke hemiparesis," *Clin. Biomechan.*, vol. 29, no. 7, pp. 780–786, 2014.

[6] F. Vaverka, M. Elfmark, Z. Svoboda, and M. Janura, "System of gait analysis based on ground reaction force assessment," vol. 45, no. 4, pp. 187–193, 2015.

[7] L. N. Awad, J. A. Palmer, R. T. Pohlig, S. A. Binder-Macleod, and D. S. Reisman, "Walking speed and step length asymmetry modify the energy cost of walking after stroke," *Neurorehabil. Neural Repair*, vol. 29, no. 5, pp. 416–423, 2015.

[8] L. N. Awad et al., "Reducing circumduction and hip hiking during hemiparetic walking through targeted assistance of the paretic limb using a soft robotic exosuit," *Amer. J. Phys. Med. Rehabil.*, vol. 96, pp. S157–S164, Oct. 2017.

[9] R. B. Davis, S. Õunpuu, D. Tyburski, and J. R. Gage, "A gait analysis data collection and reduction technique," *Human Movement Sci.*, vol. 10, pp. 575–587, Oct. 1991.

[10] S. J. Olney, M. P. Griffin, T. N. Monga, and I. Mcbride, "Work and power in gait of stroke patients," *Arch. Phys. Med. Rehabil.*, vol. 72, vol. 5, pp. 309–314, 1991.

[11] G. Chen and C. Patten, "Joint moment work during the stance-to-swing transition in hemiparetic subjects," *J. Biomech.*, vol. 41, no. 4, pp. 877–883, 2008.

[12] L. A. Jessica, A. K. Steven, and R. N. Richard, "Step length asymmetry is representative of compensatory mechanisms used in post-stroke hemiparetic walking," *Gait Posture*, vol. 33, no. 4, pp. 538–543, 2011.

[13] F. Ferrarello et al., "Tools for observational gait analysis in patients with stroke: A systematic review," *Phys. Ther.*, vol. 93, no. 12, pp. 1673–1685, 2013.

[14] D. M. Peters et al., "Utilization of wearable technology to assess gait and mobility post-stroke: A systematic review," *J. NeuroEng. Rehabil.*, vol. 18, p. 67, Apr. 2021.

[15] S. E. Oh, A. Choi, and J. H. Mun, "Prediction of ground reaction forces during gait based on kinematics and a neural network model," *J. Biomech.*, vol. 46, no. 14, pp. 2372–2380, 2013.

[16] W. R. Johnson, A. S. Mian, C. J. Donnelly, D. Lloyd, and J. A. Alderson, "Predicting athlete ground reaction forces and moments from motion capture," *Med. Biol. Eng. Comput.*, vol. 56, no. 10, pp. 1781–1792, 2018.

[17] M. Mundt, A. Koeppe, S. David, F. Bamer, W. Potthast, and B. Markert, "Prediction of ground reaction force and joint moments based on optical motion capture data during gait," *Med. Eng. Phys.*, vol. 86, pp. 29–34, Dec. 2020.

[18] A. Karatsidis, G. Bellusci, H. Schepers, M. de Zee, M. Andersen, and P. Veltink, "Estimation of ground reaction forces and moments during gait using only inertial motion capture," *Sensors*, vol. 17, no. 1, p. 75, 2016.

[19] F. J. Wouda et al., "Estimation of vertical ground reaction forces and sagittal knee kinematics during running using three inertial sensors," *Front. Physiol.*, vol. 9, p. 218, Mar. 2018.

[20] Y. Hutabarat, D. Owaki, and M. Hayashibe, "Quantitative gait assessment with feature-rich diversity using two IMU sensors," *IEEE Trans. Med. Robot. Bionics*, vol. 2, no. 4, pp. 639–648, Nov. 2020.

[21] Y. Hutabarat, D. Owaki, and M. Hayashibe, "Recent advances in quantitative gait analysis using wearable sensors: A review," *IEEE Sensors J.*, vol. 21, no. 23, pp. 26470–26487, Dec. 2021.

[22] Y. Hutabarat, D. Owaki, and M. Hayashibe, "Seamless temporal gait evaluation during walking and running using two IMU sensors," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2021, pp. 6835–6840.

[23] Y. Hutabarat, D. Owaki, and M. Hayashibe, "Temporal variation quantification during cognitive dual-task gait using two IMU sensors," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2022, pp. 1121–1124.

[24] E. Shahabpoor and A. Pavic, "Estimation of vertical walking ground reaction force in real-life environments using single IMU sensor," *J. Biomechan.*, vol. 79, pp. 181–190, Oct. 2018.

[25] E. Shahabpoor, A. Pavic, J. M. W. Brownjohn, S. A. Billings, L. Guo, and M. Bocian, "Real-life measurement of tri-axial walking ground reaction forces using optimal network of wearable inertial measurement units," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 6, pp. 1243–1253, Jun. 2018.

[26] Y.-J. Chang, S.-F. Chen, and J.-D. Huang, "A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities," *Res. Develop. Disabil.*, vol. 32, no. 6, pp. 2566–2570, 2011.

[27] H. M. Hondori and M. Khademi, "A review on technical and clinical impact of Microsoft Kinect on physical therapy and rehabilitation," *J. Med. Eng.*, vol. 2014, Dec. 2014, Art. no. 846514.

[28] R. Lun and W. Zhao, "A survey of applications and human motion recognition with Microsoft Kinect," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 29, no. 5, 2015, Art. no. 1555008.

[29] A. González, P. Fraisse, and M. Hayashibe, "Adaptive interface for personalized center of mass self-identification in home rehabilitation," *IEEE Sensors J.*, vol. 15, no. 5, pp. 2814–2823, May 2015.

[30] S. Skals, K. P. Rasmussen, K. M. Bendtsen, J. Yang, and M. S. Andersen, "A musculoskeletal model driven by dual Microsoft Kinect sensor data," *Multibody Syst. Dyn.*, vol. 41, pp. 297–316, Dec. 2017.

[31] Y. Hirano, D. Kushida, and H. Matsumoto, "Contactless motion analysis system using a Kinect and musculoskeletal model," in *Proc. IEEE Life Sci. Conf. (LSC)*, 2017, pp. 308–311.

[32] M. Eltoukhy, C. Kuenze, M. S. Andersen, J. Oh, and J. Signorile, "Prediction of ground reaction forces for parkinson's disease patients using a kinect-driven musculoskeletal gait analysis model," *Med. Eng. Phys.*, vol. 50, pp. 75–82, Dec. 2017.

[33] 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)*, 2014, pp. 455–460.

[34] D. Kobsar et al., "Validity and reliability of wearable inertial sensors in healthy adult walking: A systematic review and meta-analysis," *J. NeuroEng. Rehabil.*, vol. 17, p. 62, May 2020.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[36] S. Brunnstrom, *Recovery Stages and Evaluation Procedures*, (Movement Therapy in Hemiplegia: A Neurophysiological Approach). New York, NY, USA: Harper & Row, 1970.

[37] "Microsoft, azure Kinect body tracking SDK." Accessed: Mar. 27, 2023. [Online]. Available: https://learn.microsoft.com/en-us/azure/kinect-dk/body-sdk-download

[38] A. Cappozzo, F. Catani, U. D. Croce, and A. Leardini, "Position and orientation in space of bones during movement: Anatomical frame definition and determination," *Clin. Biomech.*, vol. 10, no. 4, pp. 171–178, 1995.

[39] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. 26th Int. Conf. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.

[40] K. Cho et al., "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empir. Methods Nat. Lang. Process.*, 2014, pp. 1724–1734.

[41] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2019, pp. 2623–2631.

[42] K. A. Sawner, S. Brunnstrom, and J. M. Vigne, "Brunnstrom's movement therapy in hemiplegia: A neurophysiological approach," Philadelphia, PA, USA: Lippincott, 1992.

[43] A. L. Hsu, P.-F. Tang, and M. H. Jan, "Analysis of impairments influencing gait velocity and asymmetry of hemiplegic patients after mild to moderate stroke," *Arch. Phys. Med. Rehabil.*, vol. 84, no. 8, pp. 1185–1193, 2003.

[44] P. Langhorne, F. Coupar, and A. Pollock, "Motor recovery after stroke: A systematic review," *Lancet Neurol.*, vol. 8, no. 8, pp. 741–754, 2009.

[45] G. M. Rozanski, A. H. Huntley, L. D. Crosby, A. Schinkel-Ivy, A. Mansfield, and K. K. Patterson, "Lower limb muscle activity underlying temporal gait asymmetry post-stroke," *Clin. Neurophysiol.*, vol. 131, no. 8, pp. 1848–1858, 2019.

[46] S. Sakamoto, Y. Hutabarat, D. Owaki, and M. Hayashibe, "Ground reaction force and moment estimation through EMG sensing using long short-term memory network," *Cyborg Bionic Syst.*, vol. 4, p. 16, Mar. 2023.

[47] M. T. N. Truong, A. E. A. Ali, D. Owaki, and M. Hayashibe, "EMG-based estimation of lower limb joint angles and moments using long short-term memory network," *Sensors*, vol. 23, no. 6, p. 3331, 2023.

[48] D. T. Westwick and R. E. Kearney, "Separable least squares identification of nonlinear Hammerstein models: Application to stretch reflex dynamics," *Ann. Biomed. Eng.*, vol. 29, pp. 707–718, Aug. 2001.

[49] E. J. Dempsey and D. T. Westwick, "Identification of Hammerstein models with cubic spline nonlinearities," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 2, pp. 237–245, Feb. 2004.

[50] Y. Wang, M. Hayashibe, and D. Owaki, "Prediction of whole-body velocity and direction from local leg joint movements in insect walking via LSTM neural networks," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 9389–9396, Oct. 2022.

[51] R. W. Bohannon, "Strength of lower limb related to gait velocity and cadence in stroke patients," *Physiother. Canada*, vol. 38, no. 4, pp. 204–206, 1986.

[52] C. M. Kim and J. J. Eng, "Symmetry in vertical ground reaction force is accompanied by symmetry in temporal but not distance variables of gait in persons with stroke," *Gait posture*, vol. 18, no. 1, pp. 23–28, 2003.

[53] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," 2016, *arXiv:1611.03530*.

[54] F. Chollet, *Deep Learning With Python*. Shelter Island, NY, USA: Manning Publ., 2017.

[55] M. Sartori, D. G. Lloyd, and D. Farina, "Neural data-driven musculoskeletal modeling for personalized Neurorehabilitation technologies," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 5, pp. 879–893, May 2016.