MovePort: Multimodal Dataset of EMG, IMU, MoCap, and Insole Pressure for Analyzing Abnormal Movements and Postures in Rehabilitation Training

Xinyu Jiang, Jianfeng Li, Zhuozhuang Zhu, Xiangyu Liu, *Member, IEEE*, Yangyang Yuan, *Graduate Student Member, IEEE*, ChihHong Chou, Shengjie Yan, Chenyun Dai, *Member, IEEE*, and Fumin Jia

***Abstract*— In most real world rehabilitation training, patients are trained to regain motion capabilities with the aid of functional/epidural electrical stimulation (FES/EES), under the support of gravity-assist systems to prevent falls. However, the lack of motion analysis dataset designed specifically for rehabilitation-related applications largely limits the conduct of pilot research. We provide an open access dataset, consisting of multimodal data collected via 16 electromyography (EMG) sensors, 6 inertial measurement unit (IMU) sensors, and 230 insole pressure sensors (IPS) per foot, together with a 26-sensor motion capture system, under different *MOVE*ments and *PO*stures for *R*ehabilitation *T*raining (MovePort). Data were collected under diverse experimental paradigms. Twenty four participants first imitated multiple normal and abnormal body postures including (1) normal standing still, (2) leaning forward, (3) leaning back, and (4) half-squat, which in practical applications, can be detected as feedback to tune the parameters of FES/EES and gravity-assist systems to keep patients in a target body posture. Data under imitated abnormal gaits, e.g., (1) with legs raised higher under excessive electrical stimulation, and (2) with dragging legs under insufficient stimulation, were also collected. Data under normal gaits with low, medium and high speeds are also included. Pathological gait data from a subject with spastic paraplegia further increases the clinical value of our dataset. We also provide source codes to perform both intra- and inter-participant motion analyses of our dataset. We expect our dataset can provide a unique platform to promote collaboration among neurorehabilitation engineers.**

***Index Terms*— Motion analysis, neurorehabilitation, gait segmentation, dataset, database.**

Manuscript received 18 October 2023; revised 16 February 2024 and 27 May 2024; accepted 14 July 2024. Date of publication 18 July 2024; date of current version 25 July 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62173094, in part by Shanghai Sailing Program under Grant 22YF1430800, in part by the Shanghai Municipal Science and Technology Major Project under Grant 2018SHZDZX01, ZJ Lab, and in part by the Shanghai Center for Brain Science and Brain-Inspired Technology. *(Xinyu Jiang and Jianfeng Li contributed equally to this work.) (Corresponding authors: Shengjie Yan; Fumin Jia.)*

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 Fudan University under Application No. FE23166I, and performed in line with the Declaration of Helsinki.

Xinyu Jiang, Jianfeng Li, Zhuozhuang Zhu, Yangyang Yuan, and Shengjie Yan are with the School of Information Science and Technology, Fudan University, Shanghai 200433, China (e-mail: sjyan@fudan.edu.cn).

Xiangyu Liu is with the College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai 200093, China.

ChihHong Chou and Chenyun Dai are with the School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

Fumin Jia is with the Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China, and also with the State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Fudan University, Shanghai 200433, China (e-mail: jfmin@fudan.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2024.3429637

I. Introduction

ALMOST 15% of the world's population live with disabilities. As for motor function disabilities, the most common causes include stroke, paraplegia, and spinal cord injury (SCI), etc. For most people with motion disabilities, advanced smart rehabilitation techniques [1], [2] can greatly improve the effect of rehabilitation training. In most real world rehabilitation training, patients are trained to regain motion capabilities with the aid of functional electrical stimulation (FES) [3]. Recent breakthrough also demonstrates epidural electrical stimulation (EES)-based spinal cord neuromodulation can restore trunk and leg motor functions after complete paralysis [4]. FES/EES-aided rehabilitation training is usually carried out first in a specific venue with available gravity-assist systems [5] to prevent falls, and then in their daily life with robotic wearable devices [6].

In the whole process of rehabilitation training, modeling on the motions or body postures is required to provide feedback to tune the parameters of FES/EES and gravity-assist systems. For example, at the onset of each rehabilitation training, electrical stimulation of key muscles in the lower limbs could help patients stand and keep balance. Real time recognition and monitoring on body postures can contribute to more precise stimulation and triggering timely fall prevention. During the walking training, the parameters of electrical stimulation is essential because excessive or insufficient stimulation can both

TABLE I
COMPARISON WITH PREVIOUS REPRESENTATIVE DATASETS

	No. Subjects	Modality	Mode
Iwama et al. [7]	4007	Camera	Normal gait from healthy subjects
Schreiber et al. [8]	50	(1) MoCap (full body) (2) EMG (lower limb) (3) Forceplate	Normal gait from healthy subjects
Moreira et al. [9]	16	(1) MoCap (lower limb) (2) EMG (lower limb) (3) Forceplate	Normal gait from healthy subjects
Luo et al. [10]	30	IMU (full body)	Normal gait from healthy subjects on outdoor irregular/ uneven surfaces
Hausdorff et al. [11]	64	Force-sensitive resistors	Normal gait from healthy subjects, pathological gait from ALS, HD and PD patients
Chatzaki et al. [12]	29	(1) IMU (on feet) (2) IPS	Normal gait from healthy subjects, pathological gait from PD patents, gait when turning for all subjects
Ours	25	(1) MoCap (full body) (2) EMG (lower limb) (3) IMU (full body) (4) IPS	Body postures, normal gait from healthy subjects, imitated abnormal gait with legs raised higher and dragging legs from healthy subjects, pathological gait from spastic paraplegia patient

lead to abnormal gait patterns. Gait analyses can also help to regulate the stimulation parameters, so that the correct muscles can be activated in each phase of a gait cycle. For each patient, an initial motion analysis model is usually needed at the beginning of the rehabilitation training, which is challenging without any training data. Additionally, a better understanding on abnormal motions in rehabilitation-related scenarios is also helpful in pilot studies of diverse research directions.

However, the lack of motion analysis dataset designed specifically for FES/EES-aided rehabilitation applications largely limits the conduct of pilot research. Most previous motion analysis and gait segmentation models have been validated on normal gaits. In rehabilitation training, a predefined motion analyses model is expected to show high generalizability to abnormal gaits. Moreover, considering FES and EES have been widely applied in rehabilitation training, where the two most typical gaits, (1) with legs raised higher under excessive stimulation, and (2) with dragging legs under insufficient stimulation, have been neglected in previous datasets. For example, the OU-ISIR dataset [7] is one of the most widely used vision-based motion analysis datasets. Schreiber et al. established a multimodal gait dataset using motion capture (MoCap) system, electromyography (EMG) sensors and ground reaction forces (GRF) [8]. Moreira et al. [9] recently open-sourced a lower-limb kinematic, kinetic, and EMG dataset collected from young healthy humans during walking at controlled speeds. However, all these datasets focused on normal gaits. As for motion analysis datasets on abnormal or pathological gaits, Hausdorff et al. [11] provided a dataset collected from patients with Parkinson's disease (PD), Huntington's disease (HD), and amyotrophic lateral sclerosis (ALS). However, in their raw data, only force-sensitive resistors (with the output roughly proportional to the force under the foot) were used for signal measurement, which cannot provide sufficient information for motion analyses. Likewise, Chatzaki et al. [12] developed a gait dataset on both healthy subjects and PD patents, but the motion-related data were collected only from feet with limited overall information on the full body. Details on representative datasets were shown in Table I.

To facilitate more studies on motion analysis during rehabilitation training, here we provide an open access dataset, consisting of multimodal data collected via 16 EMG sensors, 6 inertial measurement unit (IMU) sensors, and 230 insole pressure sensors (IPS) per foot, together with a 26-sensor MoCap system, under different *MOVE*ments and *PO*stures for *R*ehabilitation *T*raining (MovePort dataset). Data were collected under diverse experimental paradigms. Twenty four participants first imitated multiple normal and abnormal body postures including (1) normal standing still, (2) leaning forward, (3) leaning back, and (4) half-squat, which in practical applications, can be detected as feedback to tune the parameters of FES/EES and gravity-assist systems. Data under imitated abnormal gaits (1) with legs raised higher (to simulate excessive electrical stimulation), and (2) with dragging legs (to simulate insufficient stimulation), were also collected. Data under normal gaits with low, medium and high speeds are also included in our dataset. One more subject with spastic paraplegia also participated in our experiment. We also provide source codes to perform both intra- and inter-participant motion analyses of our dataset. The gait deviation index (GDI) [13], a multivariate measure of overall gait pathology, was also implemented to measure distance between each gait pattern and the normal gait template in our dataset. We expect our dataset can provide a unique platform to promote a wide range of research and collaboration among neural rehabilitation engineers.

II. MATERIALS

A. Subjects

24 healthy subjects (aged 21–35 years, 14 males, 10 females) and 1 subject with spastic paraplegia (17 years old, male) participated in our data collection experiment. The subject with spastic paraplegia has shown pathological gait for more than three years, with increased muscle tension in both lower limbs. All subjects were informed the research purpose and experimental details, and provided the informed consent. The experiment was approved by the ethics committee of Fudan University (approval number: FE23166I). Among 24 healthy subjects, 17 subjects completed the full experiment, with 7 subjects completing $\sim$ 90% experiments. We open source data from all subjects to make the best use of as many data as possible. The details of missing experiment paradigms for the 7 subjects can be found in the "Missing Data" section.

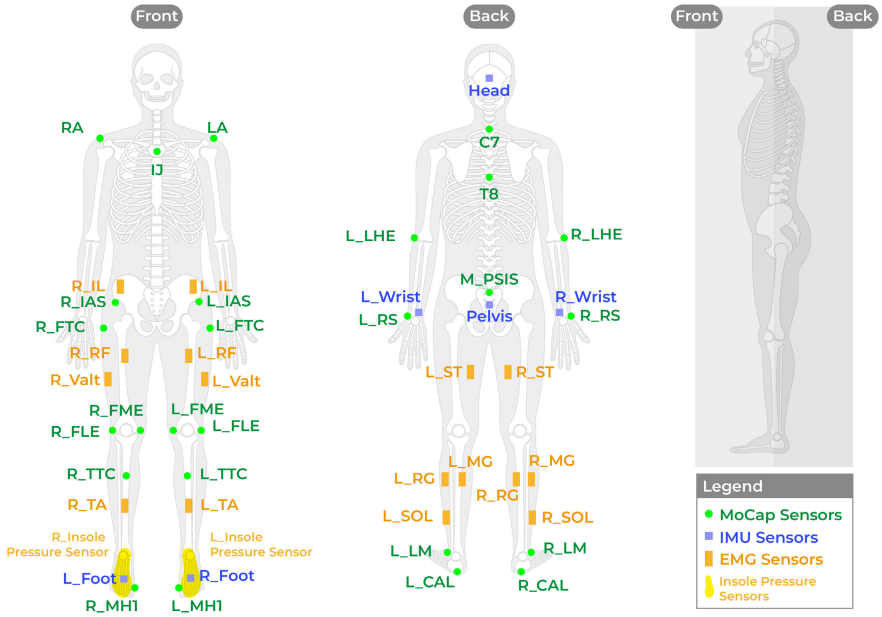

Fig. 1. Sensor Placement. For the notation of each sensor location, "L_" and "R_" denotes the same corresponding location of the left and right part of human body, respectively.

The storage size of our dataset is 15.5 GB. The directory structure of our dataset is presented in Fig. 7 in Appendix.

B. Data Collection

EMG, IMU, IPS and MoCap snesors, were used for motion measurement in our experiment. First, 16 EMG sensors (Cometa Systems) were placed on the vastus lateralis (Vlat), rectus femoris (RF), semitendinosus (ST), tibialis anterior (TA), medial gastrocnemius (MG), lateral gastrocnemius (LG), soleus muscle (SOL), and iliopsoas muscle (IL) of each subject, as presented in Fig. 1. Second, 6 IMU sensors (Xsens) were placed on the head, pelvis, right wrist, left wrist, left foot and right foot of each subject. Each IMU sensor can measure 3-axis accelerometers, 3-axis gyroscopes, and 3-axis Euler angles, with a total of 6 sensors $\times$ 9-axis per sensor $=$ 54-axis information measured. Third, 2 IPS arrays (Xsensor) with 230 sensors each array, were placed under two feet. MoCap system (Qualisys) with 26 sensors were used to capture the motion of human body. The sensor locations and notations largely overlap with the previous study (supplementary materials of [14]). The total force of each foot and the coordinates of center of pressure (COP) on both feet were saved separately. The sampling rates of EMG, IMU, IPS and MoCap sensors are 2000 Hz, 100 Hz, 60 Hz, and 100 Hz, respectively. Data of above modalities are saved in files with a ".csv" format. Using our open-sourced codes, the loaded data are $n \times t$ matrix, where n refers to number of channels (EMG and IPS) or axes (MoCap and IMU), and t refers to the number of samples.

C. Experimental Paradigm

1) Body Posture Recognition: In the first session, 24 healthy subjects were required to complete and maintain the target body postures on the ground. These body postures consists of: (1) normal standing still; (2) leaning forward, i.e. imitating a posture with the center of gravity forward; (3) leaning back, i.e. the posture with the center of gravity backward and (4) half-squat, i.e. bending their knees to imitate the posture when the support force is insufficient. On average, for each body posture of each subject, 62.1 s $\pm$ 12.9 s signals were recorded.

2) Normal Gaits From Healthy Subjects: In the second session, 24 healthy subjects were required to perform normal gaits on non-tilted treadmill at low (1 km/h), medium (2 km/h) and high (3-4 km/h, a higher but still relatively comfortable speed selected by each subject) speeds. Subjects were allowed to take a break between experiments with different speeds. On average, for each subject, 94.5 s $\pm$ 31.4 s, 80.1 $\pm$ 18.2 s, and 72.9 $\pm$ 16.0 s signals were recorded for gaits at low, medium and high speeds, respectively.

3) Abnormal Gaits From Healthy Subjects: In the third session, 24 healthy subjects were required to perform abnormal gaits on non-tilted treadmill. Two most representative abnormal gaits, namely (1) gaits with legs raised higher, and (2) gaits with dragging legs, were considered. The gaits with legs raised higher can simulate the situations when patients receive excessive electrical stimulation, while the gaits with dragging legs can simulate the situations when patients have difficulty lifting his legs or receive insufficient electrical stimulation. These two abnormal gaits are the most common ones in FES/EES-aided rehabilitation training. For gaits with legs raised higher, the speed was set to 2 km/h, and on average, 104.8 s $\pm$ 32.0 s signals were recorded from each subject. For gaits with dragging legs, the speed was set to 1 km/h, and on average, 85.7 s $\pm$ 15.4 s signals were recorded from each subject.

TABLE II
NETWORK ARCHITECTURE

Layers	Components	In/Out
Conv 1	Conv (w=[k1,k2], s=[s1,s2]) BN, ReLU MP (w=[s1,s2], s=[s1,s2]) Dropout (p=0.3)	1/32
Conv 2	Conv (w=[k1,k2], s=[s1,s2]) BN, ReLU MP (w=[s1,s2], s=[s1,s2]) Dropout (p=0.3)	32/64
Conv 3	Conv (w=[k1,k2], s=[s1,s2]) BN, ReLU MP (w=[s1,s2], s=[s1,s2]) Dropout (p=0.3)	64/128
FC 1	FC ($nn = 200$), ReLU	128/1
FC 2	FC ($nn = 100$), ReLU	1/1
FC 3	FC ($nn = 20$), ReLU	1/1
FC 4	FC ($nn = N_{class}$)	1/1
Softmax	Softmax operation	1/1

*Conv, BN, ReLU, MP, FC denote convolution, batch normalization, rectified linear unit, max pooling, and fully connected layers, respectively. Variable $w = [k1, k2]$ refers to the kernel size along the row and column dimensions of $R \times C$ feature matrix, $s = [s1, s2]$ refers to both the stride and pooling size of two dimensions of the input, where $k1 = max(\lceil \sqrt[3]{R} \rceil, 1)$, $k2 = max(\lceil \sqrt[3]{C} \rceil, 1)$, $s1 = max(\frac{\lceil \sqrt[3]{R} \rceil}{2}, 1)$, $s2 = max(\frac{\lceil \sqrt[3]{C} \rceil}{2}, 1)$, which change adaptively with the size of the input matrix, so that the outputs of the third layer can be highly compact representations. Variable nn denotes the number of neurons. N_{class} represents the number of gestures. The depth of input and output features is summarized in the rightmost column.

4) Data From Subject With Spastic Paraplegia: This subject performed 4 body postures, the same as the experiment for healthy subjects. Then the subject was required to walk on ground, with a self-defined comfortable speed. On average, for each body posture, 51.6 ± 6.3 s signals were recorded. Signals with 32.1 s duration was recorded for pathological gait data.

D. Missing Data

Participants 1–17 fully completed all experiments described above. For participants 18, 19, 20 and 24, data of the "leaning forward" body posture are not available. For participant 21, data of the "leaning forward" body posture and normal gait at low-speed are not available. For participant 22, data of the "leaning back" body posture are not available. For participant 23, data of the "leaning forward" and "half-squat" body postures are not available. we open source all available data.

III. METHODS

A. Feature Extraction

1) Body Posture Recognition: For signals of each modality, features were extracted via sliding windows. Both the window width and sliding step were set to 50 ms. For EMG signals, root mean square (RMS), waveform length (WL), slope sign changes (SSC), and zero crossing (ZC) were extracted from signals in each sensor channel within each sliding window. The extracted four types of features are the most widely used ones in previous studies on EMG analyses [15]. The overall length of the EMG feature vector in each window is 64 (16 sensors × 4 features). For IMU and IPS signals, the mean value of all samples in each channel within each sliding window was extracted. The overall length of the IMU and IPS feature vector in each window is 54 (6 IMU sensors × 9-axis information × 1 feature) and 460 (230 sensors × 2 feet × 1 feature), respectively. For the MoCap signals (i.e. the coordinates of each sensor in three-dimensional space), the MoCap sensor placed on the head of each subject was used as a new reference to calculate the new relative coordinates of each sensor, so that the MoCap data would not relate to subjects' location in the room but only depends on the body posture. Then, the mean value of re-referenced coordinates for each sensor within each sliding window was extracted as a feature. In addition to the coordinate features extracted from the re-referenced sensors, the velocity of each of the original 26 sensors was extracted as a new feature, by calculating the derivative of the coordinate signals. The velocity and re-referenced coordinate features were combined as representations of MoCap signals in gait analyses. The total length of MoCap feature vector in a sliding window is 153 ((26 sensors × 3-dimensional velocity) + (25 re-referenced sensors × 3-dimensional coordinates)).

2) Gait Analyses: For gait analyses, the feature extraction steps are similar as described above. One difference is that we used features in the current and previous neighbor windows together to describe the dynamic gait characteristics. After extracting features from gait data in each sliding window, we further stacked feature vectors in previous Q windows (a total of Q+1 windows) to form a aggregated feature vector. $Q = 5$ (equivalent to (Q+1)×50 ms=300 ms duration) was set for gait analyses. The parameter Q was selected empirically considering the dynamic gait patterns within previous 300 ms should be most relevant to the current pattern. Additionally, considering features in $Q = 5$ windows were integrated, we only extracted velocity features for each MoCap sensor in a window, to avoid an extremely high dimensionality of the MoCap feature vector.

Previous studies used different criteria to segment a gait cycle into different phases [16]. The most commonly used criteria is to segment a gait cycle into: (1) initial contact (the heel touches the ground first), (2) foot flat (the entire sole of the foot touches the ground in a flat angle), (3) heel off (the heel leaves the ground and toes touch the ground), and (4) toe off (the whole foot leaves the ground). However, most of these gait phases are designed for normal gaits, and are not applicable for abnormal gaits. For example, in abnormal gaits with legs raised higher or with dragging legs, one foot usually touches the ground directly in a flat angle. In other words, a cycle of certain abnormal gaits does not include initial contact and toe off phases. Previous studies have not paid enough attentions on the gait segmentation criteria applied to abnormal gaits. Therefore, in our work, we segment a gait cycle into two phases: (1) the stance phase, when the right foot is on the ground, and (2) the swing phase, when the right foot leaves the ground. The stance and swing phases exist in all gait cycles, both normal and abnormal ones. According to the definitions, the stance and swing phases can be easily and accurately segmented by calculating the total force on a foot (via IPS). In our work, we segmented signals in a window into the stance phase if the total force on right foot is higher than a threshold set as 50 N, and into swing phase otherwise. An additional criteria that each gait phase should last at least

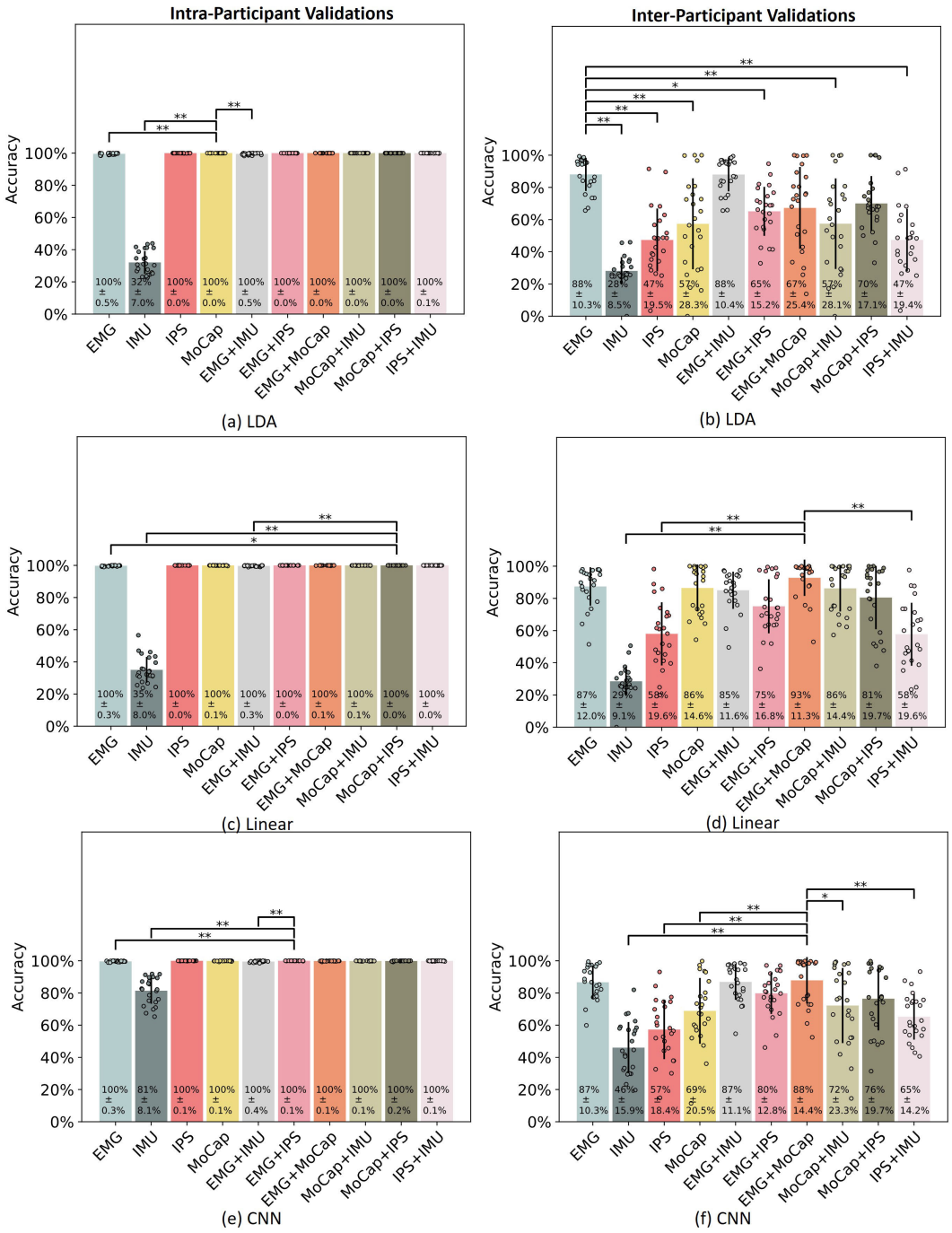

Fig. 2. Results of body posture recognition (healthy subjects). Symbols "*" and "**" denote $\mathbf{0.005} \leq p < \mathbf{0.05}$ and $p < \mathbf{0.005}$, respectively. To make results clear, only significant differences compared with the best modality (or combination) are presented.

200 ms was adopted to filter outlier phases. The insole pressure is an acceptable modality used as gold standard in wearable sensing [17]. We also provided RGB videos and MoCap data which future studies could use to define their own gait cycles according to different needs in different applications.

B. Decoding Models

To perform baseline analyses and verify the validity of our data, three models namely linear discriminant analysis (LDA), linear classification model, and convolutional neural networks (CNN) were employed.

For LDA and linear models, the model input is a one-dimensional feature vector. For CNN model, the model input is a $R \times C$ feature matrix, where $R = Q + 1$ refers to the number of windows considered in each example ($R = 1$ and $R = 6$ for body posture recognition and gait segmentation, respectively), C refers to the number of features in each window. When integrating multiple modalities, features of different modalities were concatenated along the second dimension (increasing the value of C). The architecture of CNN model is presented in Table II. To train a CNN model, stochastic gradient descent with momentum (SGDM)

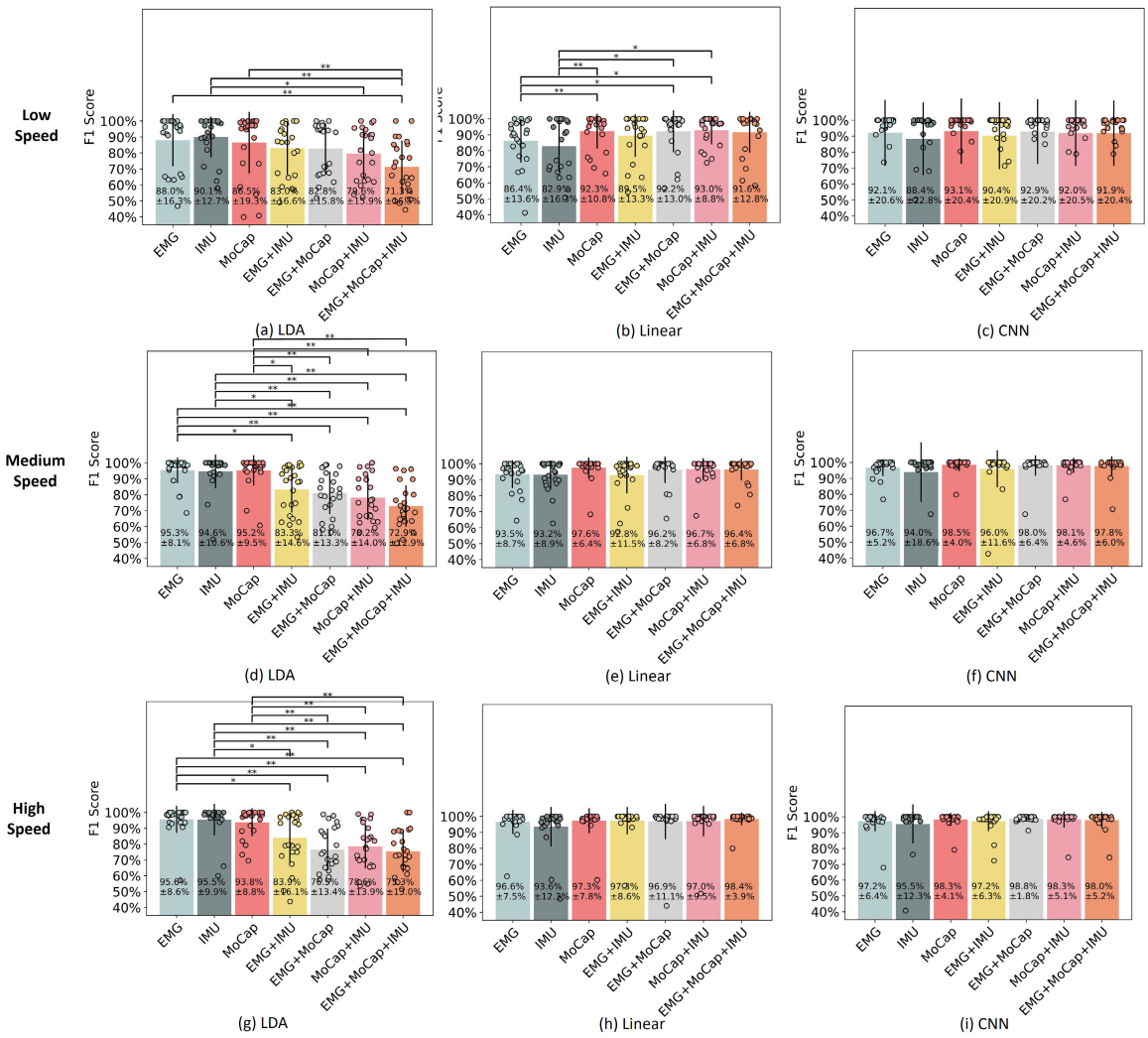

Fig. 3. Results of normal gait segmentation in intra-participant validations (healthy subjects). Symbols "*" and "**" denote significant differences with $\mathbf{0.005} \leq p < \mathbf{0.05}$ and $p < \mathbf{0.005}$, respectively. In each subfigure, all pair-wise comparisons were performed, and significant differences were not observed for all group pairs without symbols "*" or "**". IMU data of subject 15 is relatively more noisy but here we present results on all subjects as an overall baseline (the same for other results).

optimizer [18] was applied with a batch size of 16. The parameter optimization stopped after 10000 iteration steps or 30 epochs (whichever is lower, depending on the sample size in different validations). 10% data from the training set was allocated as the validation set. The network with the best validation performance was saved. For gait segmentation using all models, once a switch between two phases was detected, a cold down mechanism was adopted to keep the gait phase unchanged for 200 ms. Key functions of our open-sourced codes are presented in Table VI in Appendix.

C. Validation Methodologies

We employed both intra- and inter-participant validations. For intra-participant validations, 70% of the data from each participant were allocated into the training set with the remaining 30% allocated into the testing set. For inter-participant validations, data from each participant were used as testing data separately, with data from all other participants as the training set.

For body posture recognition, classification accuracy was used as evaluation metric. For gait segmentation, F_1 score of each event was used as the evaluation metric [19]. The events of switch from stance to swing and switch from swing to stance were demoted as $Stance2swing$ and $Swing2stance$, respectively. F_1 is defined as the harmonic mean of Precision P and Recall R, i.e. $F_1 = \frac{2PR}{P+R}$, where $P = \frac{TP}{TP+FP}$, $R = \frac{TP}{TP+FN}$, TP, FP, and FN refer to true positives, false positives, and false negatives. Specifically, TP are the first correctly detected event lying within a $\pm Tol$ ms window near each ground truth event, where Tol is a tolerance parameter that equals to $0.2\times$ the median period of a gait cycle. Accordingly, the Tol parameter varies with gaits of different speeds. Most of the Tol parameters are about 200 ms for gaits with speed within 3–4 km/h, which is the same as the previous study with a similar speed [19]. FP are the detected events lying outside the tolerance window, or additional detected events within the tolerance window after the first correctly detected event. FN are those ground truth events that are not detected within the tolerance window.

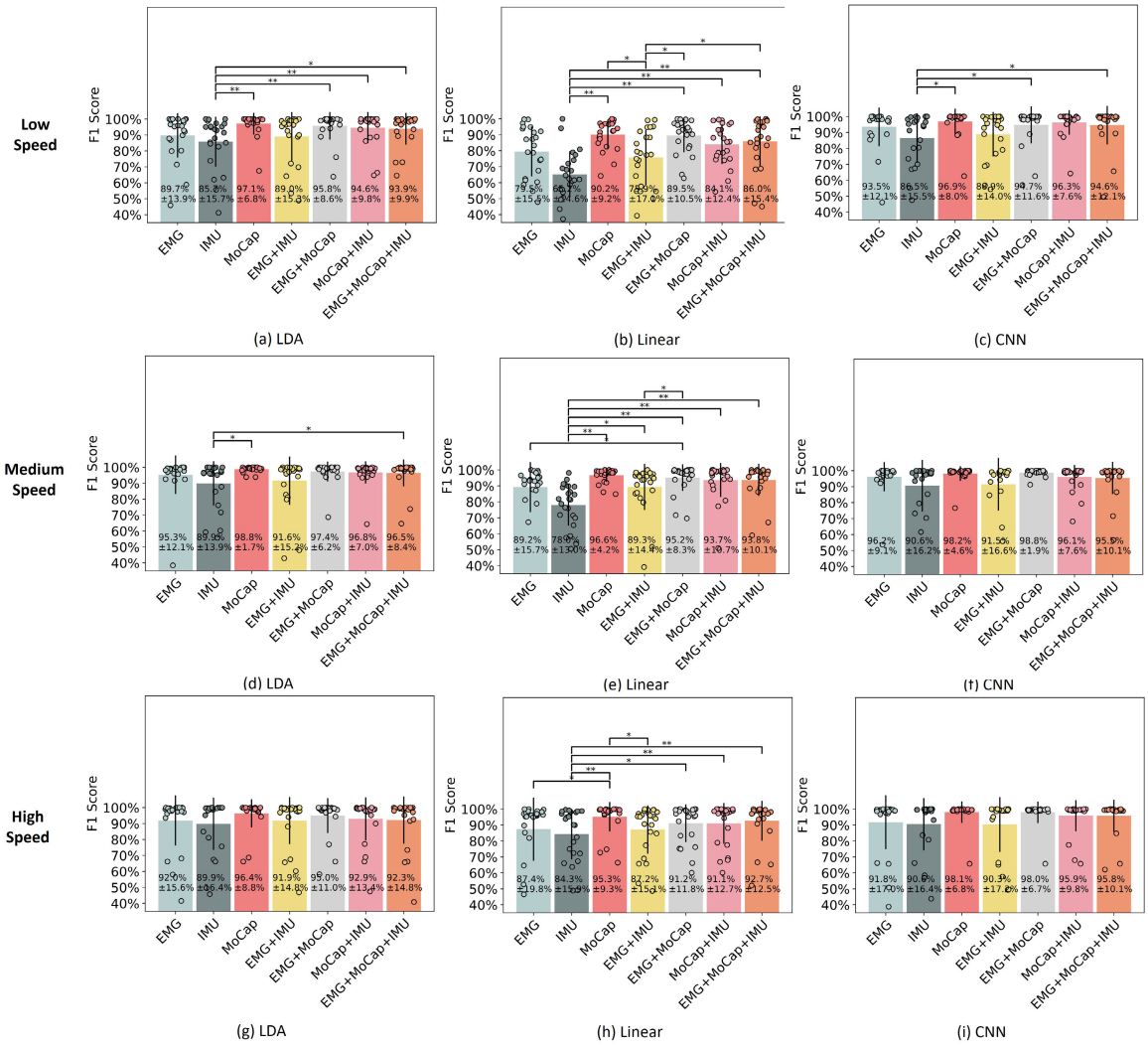

Fig. 4. Results of normal gait segmentation in inter-participant validations (healthy subjects). Symbols "*" and "**" denote significant differences with $\mathbf{0.005} \leq p < \mathbf{0.05}$ and $p < \mathbf{0.005}$, respectively. In each subfigure, all pair-wise comparisons were performed, and significant differences were not observed for all group pairs without symbols "*" or "**".

D. Gait Deviation Index

To evaluate the similarity between the abnormal gaits and the normal ones, GDI was calculated for each gait mode. GDI is a multivariate measure of the pathology of gaits. The GDI was calculated based on 9 Plug-in gait kinematic variables extracted from MoCap data. These variables are: pelvic tilt, pelvic obliquity, pelvic rotation, hip flexion/extension, hip ab/adduction, hip rotation, Knee flexion/extension, ankle dorsi/plantar flexion, and foot progression. The details of GDI can be found in [13]. To normalize the effect of speed, before calculating these variables, data in each gait cycle were resampled to 51 samples (2% increments for each new sample in a cycle, the same as [13]). Leave-one-participant out GDI evaluation was applied. For each evaluation participant, we first allocated data of normal gaits from other participants as the modelling set. Then GDI of the normal gaits, gaits with dragging legs, and gaits with legs raised higher from the evaluation participant were measured by comparing to the modelling set separately.

E. Statistical Analyses

To compare the performance of different modalities (with $\geq$ 3 groups), the Friedman test was applied to verify the overall inter-group significance. If the overall inter-group significance was detected, the Nemenyi post-hoc test, a multi-comparison test, was applied to identify pair-wise group differences. Significance was claimed if $p < 0.05$ was obtained.

IV. Results

A. Body Posture Recognition

Fig. 2 presents the results of body posture recognition on healthy subjects using different models and modalities in different validation methods. In intra-participant validations, all models achieved excellent performances with accuracy of $\sim$ 100% achieved using almost all modalities (and modality combinations), except IMU. The accuracy in inter-participant validations dropped for all models and all modalities. In inter-participant validations, EMG achieved

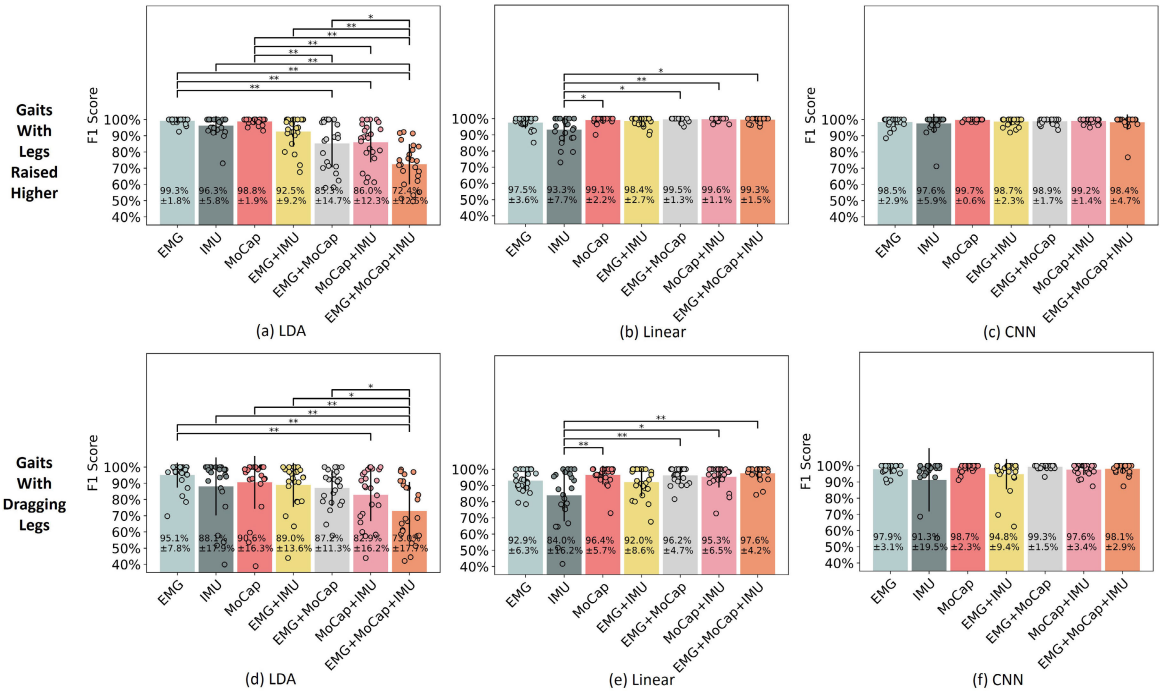

Fig. 5. Results of abnormal gait segmentation in intra-participant validations (healthy subjects). Symbols "*" and "**" denote significant differences with $\mathbf{0.005} \leq p < \mathbf{0.05}$ and $p < \mathbf{0.005}$, respectively. In each subfigure, all pair-wise comparisons were performed, and significant differences were not observed for all group pairs without symbols "*" or "**".

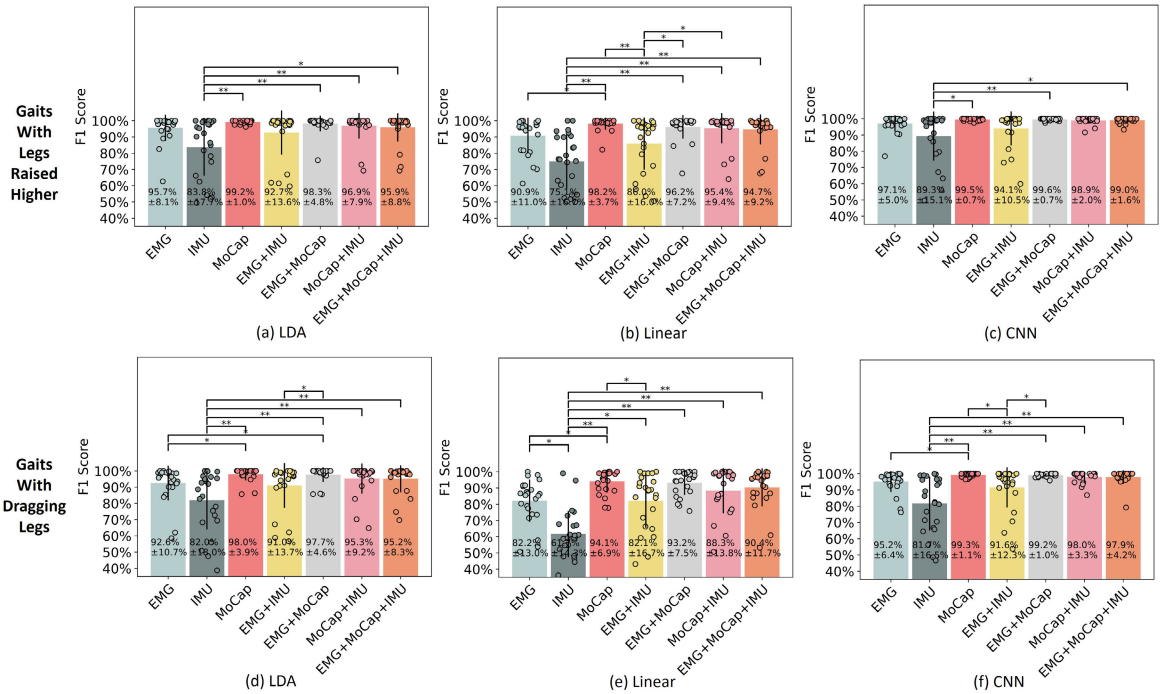

Fig. 6. Results of abnormal gait segmentation in inter-participant validations (healthy subjects). Symbols "*" and "**" denote significant differences with $\mathbf{0.005} \leq p < \mathbf{0.05}$ and $p < \mathbf{0.005}$, respectively. In each subfigure, all pair-wise comparisons were performed, and significant differences were not observed for all group pairs without symbols "*" or "**".

TABLE III

ACCURACY (%) OF BODY POSTURE RECOGNITION FOR THE PATIENT WITH SPASTIC PARAPLEGIA

	EMG	IMU	IPS	MoCap	EMG+IMU	EMG+IPS	EMG+MoCap	IMU+MoCap	IPS+MoCap	IMU+IPS
LDA	93.3	23.7	100	100	92.5	100	100	100	100	100
Linear	92.6	29.3	100	100	89.9	100	100	100	100	100
CNN	91.0	52.3	100	100	88.0	100	100	99.9	100	100

TABLE IV
F1 SCORE (%) OF GAIT SEGMENTATION FOR THE PATIENT WITH SPASTIC PARAPLEGIA

	EMG	IMU	MoCap	EMG+IMU	EMG+MoCap	IMU+MoCap	EMG+MoCap+IMU
LDA	73.9	79.7	61.6	66.6	88.5	83.9	72.1
Linear	88.6	92.6	85.9	78.9	84.7	98.1	71.2
CNN	93.8	97.1	91.2	91.2	93.8	97.1	91.2

TABLE V
GDI OF DIFFERENT GAITS

Gait Type	GDI
Noraml gait from healthy subjects	100.12±11.26
Gait with dragging legs from healthy subjects	94.07±8.24
Gait with legs raised higher from healthy subjects	83.16±8.00
Pathological gait from subject with spastic paraplegia	72.14

the highest unimodal recognition accuracy, demonstrating the reliability and robustness of EMG in real world body posture recognition applications.

B. Segmentation of Normal Gaits

Fig. 3 presents the results of normal gait segmentation on healthy subjects using different models and modalities in intra-participant validations. In intra-participant validations, the F_1 score of LDA tended to decrease with multimodal fusion compared with unimodal methods. For linear and CNN models, F_1 scores keep in a relatively stable level with different modalities. The gait segmentation performances on low speed gaits are relatively lower compared with the counterparts of medium and high speeds.

Fig. 4 presents the results of normal gait segmentation on healthy subjects in inter-participant validations. In inter-participant validations, LDA did not show degraded performance with multimodal data fusion, compared with intra-participant validations. The high inter-participant F_1 scores of $> 96\%$, $> 98\%$, and $> 98\%$ achieved by CNN in segmenting gaits with low, medium and high speeds, respectively, demonstrated the high transferability of the learned gait pattern on new participants.

C. Segmentation of Abnormal Gaits

Fig. 5 presents the results of abnormal gait segmentation on healthy subjects using different models and modalities in intra-participant validations. In intra-participant validations, LDA achieved significantly lower F_1 scores using multimodal segmentation compared with the unimodal method, showing similar characteristics as intra-participant normal gait segmentation. For linear and CNN models, multimodal fusion achieved relatively high F_1 scores compared with LDA.

Fig. 6 presents the results of abnormal gait segmentation on healthy subjects in inter-participant validations. In inter-participant validations, multimodal fusion would not lead to performance degradation of LDA, demonstrating that LDA could learn to find complex inter-participant multimodal patterns. Using CNN models, F_1 scores as high as 99.6% and 99.3% can be achieved on segmenting gaits with legs raised higher and gaits with dragging legs, respectively, without using any training data from the target participant. The high F_1 scores of inter-participant segmentation of abnormal gaits demonstrate the high potentials of using our data to develop a pre-trained but highly generalizable model in new application scenarios.

D. Results on the Subject With Spastic Paraplegia

Table III presents the accuracy of posture recognition on the spastic paraplegia patient. IPS and MoCap performed the best with 100% accuracy achieved using all models, while unimodal IMU contributed to the lowest accuracy, demonstrating that IMU could not provide enough information in a still body state. Table IV presents the results of pathological gait segmentation. Overall, CNN model achieved the highest F_1 score in most cases.

E. Results on GDI of Different Types of Gait

Table V presents the GDI of different types of gait. The normal gait from healthy subjects are with the highest DGI of 100.12. According to [13], a GDI value of ∼100 can be viewed as normal gaits. Lower GDI values refer to deviation of the gait away from the normal gait. Overall, for imitated abnormal gaits, both gait with dragging legs and gait with legs raised higher lead to lower GDI compared with the noraml gait (both with significant differences). The pathological gait from the spastic paraplegia patient leads to the lowest GDI of 72.14.

V. DISCUSSION

In our analyses, we presented results in both body posture recognition and normal/abnormal gait segmentation, using LDA, linear and CNN models via both unimodal and multimodal data. The achieved high accuracy and F_1 score both verified the validity of our collected data and the excellent performances of our developed models.

For intra-participant body posture recognition, IMU data contributed to the lower accuracy, probably due to the fact that the inertial information in a still state cannot provide useful information. While in inter-participant validations, EMG data could contribute to the highest unimodal body posture recognition accuracy using all of LDA, linear and CNN models. The excellent performance using EMG is probably due to that, EMG data describe different functions/activations of several key muscles in different postures. MoCap data which only describe the shape of human body might not be discriminant between the two similar body postures. However, the activated muscles might change largely with slightly different postures. Additionally, the inter-subject differences of EMG are mainly reflected in the different local anatomical structures of the body. In the macroscopic level, e.g. EMG collected from

TABLE VI
IMPORTANT FUNCTIONS IN OUR CODES

Function Name	Description
main_gait_treadmill_normal.m	Intra-subject segmentation of normal gait from healthy subjects using LDA/linear models
main_gait_treadmill_normal_inter_subject.m	Inter-subject segmentation of normal gait from healthy subjects using LDA/linear models
main_gait_treadmill_normal_cnn.m	Intra-subject segmentation of normal gait from healthy subjects using CNN
main_gait_treadmill_normal_inter_subject_cnn.m	Inter-subject segmentation of normal gait from healthy subjects using CNN
main_gait_treadmill_abnormal.m	Intra-subject segmentation of imitated abnormal gait from healthy subjects using LDA/linear models
main_gait_treadmill_abnormal_inter_subject.m	Inter-subject segmentation of imitated abnormal gait from healthy subjects using LDA/linear models
main_gait_treadmill_abnormal_cnn.m	Intra-subject segmentation of imitated abnormal gait from healthy subjects using CNN
main_gait_treadmill_abnormal_inter_subject_cnn.m	Inter-subject segmentation of imitated abnormal gait from healthy subjects using CNN
main_pathological_gait.m	Segmentation of pathological gait from the subject with spastic paraplegia using LDA/linear models
main_pathological_gait_cnn.m	Segmentation of pathological gait from the subject with spastic paraplegia using CNN
main_posture_recognition.m	Intra-subject recognition of body postures collected from healthy subjects using LDA/linear models
main_posture_recognition_inter_subject.m	Inter-subject recognition of body postures collected from healthy subjects using LDA/linear models
main_posture_recognition_cnn.m	Intra-subject recognition of body postures collected from healthy subjects using CNN
main_posture_recognition_inter_subject_m	Inter-subject recognition of body postures collected from healthy subjects using CNN
demo_skeleton_construct.m	A demo to construct joint skeleton from MoCap data

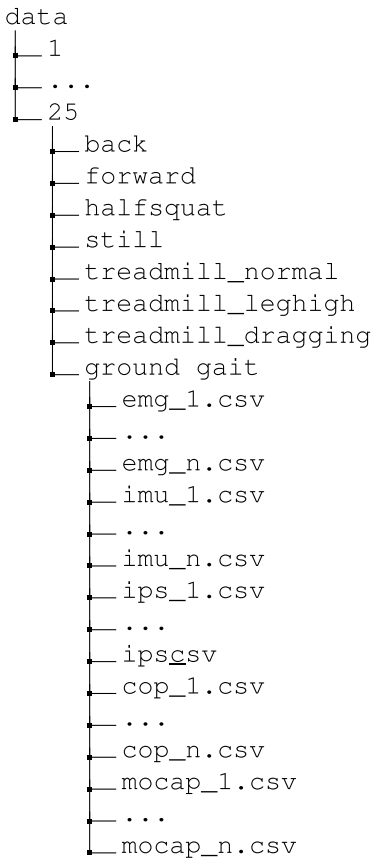

Fig. 7. File structure of our dataset. Number 1–25 refers to the index of a subject. Folders "back", "forward", "halfsquat" and "still" save data under 4 body postures. Folder "treadmill_normal" saves data of normal gait from healthy subjects. Folders "treadmill_leghigh" and "treadmill_dragging" save data of imitated abnormal gait. Folder "ground_gait" saved data of pathological gait on ground (only available for the subject with spastic paraplegia).

the full body, the overall EMG patterns among different individuals tend to be similar. However, the pattern of EMG may be largely different for those participants with significantly lower/higher muscle activation levels. We expect future follow-up studies could provide more insights using our data.

For gaits segmentation, LDA tended to achieve degraded performances in intra-participant validations using multimodal data compared with unimodal data. This finding is likely due to that fact in intra-participant validations, the size of training data from only one participant is small. Additionally, LDA-based methods to find an effective feature subspace via matrix singular value decomposition are highly likely to be affected by the "dimension disaster" problem with a small number of training examples but a high-dimensional feature vector. This problem of LDA models can be solved in inter-participant gait segmentation with more training data from more participants. Moreover, CNN models achieved relatively better performances compared with LDA and linear in segmenting both normal and abnormal gaits, mainly due to the fact the gait data showed highly dynamic patterns, with the high-dimensional data distributed on a manifold in the feature space. Neural networks are known to better handle highly non-linear complex data by flattening the manifold-shaped data in deeper layers [20]. A standard CNN model without extensively searching the optimal architecture could contribute to an excellent performance, showing the promising prospects of neural networks and deep learning models in future studies.

Abnormal motion analyses are neglected by most previous studies, but are essential in applications related to rehabilitation training. Establishing a dataset for research purposes is extremely time-consuming. Our dataset can save a huge amount of time for researchers, especially in pilot studies. To the best of our knowledge, the provided dataset is the first one to facilitate pilot studies in motion analyses within neural rehabilitation areas. Here we also summarized all possible research directions that can benefit from our dataset:

1) Abnormal body posture recognition/monitoring for fall prevention [21], applied in both rehabilitation training and health monitoring of elderly/isolated people [22], [23].

2) Biomechanical modelling of body balance and postures in both still and moving states [24].

3) Abnormal gait segmentation in rehabilitation training, which can provide feedback to tune the parameters of FES/EES systems in different gait phases.

4) Normal gait segmentation using data from different speeds to develop speed-robust models.

5) Multimodal gait biometric identification [25]. In addition to rehabilitation training, gaits are also important biometrics that can be used for human identification. The provided dataset can facilitate gait biometric identification using wearable sensors which are more user-centric compared with vision-based methods.

6) Inter-modality signal correlation/transfer. Since we provided multimodal motion signals, future studies could investigate the correlation between different modalities. Previous study also employed data disentanglement to transfer signals in one modality into another [26], based on the inherent correlation and variance between different modalities.

VI. Conclusion

In this work, we provide the *MovePort* dataset, consisting of multimodal data collected via EMG, IMU, IPS and MoCap sensors. Data were collected under diverse experimental paradigms. The provided open access dataset facilitates future studies in diverse research directions, including body posture recognition, normal/abnormal gait analyses in rehabilitation, and more. The high inter-participant performances demonstrate both the high signal quality and the high potentials of developing a model on our data and then applying to new participant in new research. The *MovePort* dataset and source codes in our analyses are available online. Please fetch our data at figshare (https://doi.org/10.6084/m9.figshare.25202183.v1) and codes at github (https://github.com/Open-EMG/MovePortToolbox).

Appendix
Supplementary Figures and Tables

See Table VI and Fig. 7.

References

[1] L.-W. Chou et al., "Effects of noise electrical stimulation on proprioception, force control, and corticomuscular functional connectivity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2518–2524, 2023.

[2] B. Zhong et al., "A cable-driven exoskeleton with personalized assistance improves the gait metrics of people in subacute stroke," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2560–2569, 2023.

[3] S. Bala, V. Y. Vishnu, and D. Joshi, "Muscle synergy-based functional electrical stimulation reduces muscular fatigue in post-stroke patients: A systematic comparison," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2858–2871, 2023.

[4] A. Rowald et al., "Activity-dependent spinal cord neuromodulation rapidly restores trunk and leg motor functions after complete paralysis," *Nature Med.*, vol. 28, no. 2, pp. 260–271, Feb. 2022.

[5] J.-B. Mignardot et al., "A multidirectional gravity-assist algorithm that enhances locomotor control in patients with stroke or spinal cord injury," *Sci. Transl. Med.*, vol. 9, no. 399, Jul. 2017, Art. no. eaah3621.

[6] E. Bardi, M. Gandolla, F. Braghin, F. Resta, A. L. G. Pedrocchi, and E. Ambrosini, "Upper limb soft robotic wearable devices: A systematic review," *J. NeuroEng. Rehabil.*, vol. 19, no. 1, p. 87, Aug. 2022.

[7] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, "The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1511–1521, Oct. 2012.

[8] C. Schreiber and F. Moissenet, "A multimodal dataset of human gait at different walking speeds established on injury-free adult participants," *Sci. Data*, vol. 6, no. 1, p. 111, Jul. 2019.

[9] L. Moreira, J. Figueiredo, P. Fonseca, J. P. Vilas-Boas, and C. P. Santos, "Lower limb kinematic, kinetic, and EMG data from young healthy humans during walking at controlled speeds," *Sci. Data*, vol. 8, no. 1, p. 103, Apr. 2021.

[10] Y. Luo, S. M. Coppola, P. C. Dixon, S. Li, J. T. Dennerlein, and B. Hu, "A database of human gait performance on irregular and uneven surfaces collected by wearable sensors," *Sci. Data*, vol. 7, no. 1, p. 219, Jul. 2020.

[11] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *J. Appl. Physiol.*, vol. 88, no. 6, pp. 2045–2053, Jun. 2000.

[12] C. Chatzaki et al., "The smart-insole dataset: Gait analysis using wearable sensors with a focus on elderly and Parkinson's patients," *Sensors*, vol. 21, no. 8, p. 2821, Apr. 2021.

[13] M. H. Schwartz and A. Rozumalski, "The gait deviation index: A new comprehensive index of gait pathology," *Gait Posture*, vol. 28, no. 3, pp. 351–357, Oct. 2008.

[14] M. Plooij et al., "Neglected physical human-robot interaction may explain variable outcomes in gait neurorehabilitation research," *Sci. Robot.*, vol. 6, no. 58, Sep. 2021, Art. no. eabf1888.

[15] X. Jiang et al., "Open access dataset, toolbox and benchmark processing results of high-density surface electromyogram recordings," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1035–1046, 2021.

[16] H. T. T. Vu et al., "A review of gait phase detection algorithms for lower limb prostheses," *Sensors*, vol. 20, no. 14, p. 3972, Jul. 2020.

[17] H. Prasanth et al., "Wearable sensor-based real-time gait detection: A systematic review," *Sensors*, vol. 21, no. 8, p. 2727, Apr. 2021.

[18] Y. Liu et al., "An improved analysis of stochastic gradient descent with momentum," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, H. Larochelle et al., Eds., Red Hook, NY, USA: Curran Associates, 2020, pp. 18261–18271.

[19] F. A. Garcia, J. C. Pérez-Ibarra, M. H. Terra, and A. A. G. Siqueira, "Adaptive algorithm for gait segmentation using a single IMU in the thigh pocket," *IEEE Sensors J.*, vol. 22, no. 13, pp. 13251–13261, Jul. 2022.

[20] P. P. Brahma, D. Wu, and Y. She, "Why deep learning works: A manifold disentanglement perspective," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 10, pp. 1997–2008, Oct. 2016.

[21] A. Choi et al., "Deep learning-based near-fall detection algorithm for fall risk monitoring system using a single inertial measurement unit," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2385–2394, 2022.

[22] M. Yu, Y. Yu, A. Rhuma, S. M. R. Naqvi, L. Wang, and J. A. Chambers, "An online one class support vector machine-based person-specific fall detection system for monitoring an elderly individual in a room environment," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 6, pp. 1002–1014, Nov. 2013.

[23] A. Chaudhary, R. Mishra, H. P. Gupta, and K. K. Shukla, "Jointly prediction of activities, locations, and starting times for isolated elderly people," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 5, pp. 2288–2295, May 2023.

[24] D. Winter et al., "An integrated EMG/biomechanical model of upper body balance and posture during human gait," in *Natural and Artificial Control of Hearing and Balance* (Progress in Brain Research), vol. 97, J. Allum et al., Eds., Amsterdam, The Netherlands: Elsevier, 1993, pp. 359–367.

[25] Y. Shi et al., "Robust gait recognition based on deep CNNs with camera and radar sensor fusion," *IEEE Internet Things J.*, vol. 10, no. 2, pp. 10817–10832, Jun. 2023.

[26] X. Gu, Y. Guo, F. Deligianni, B. Lo, and G.-Z. Yang, "Cross-subject and cross-modal transfer for generalized abnormal gait pattern recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 546–560, Feb. 2021.