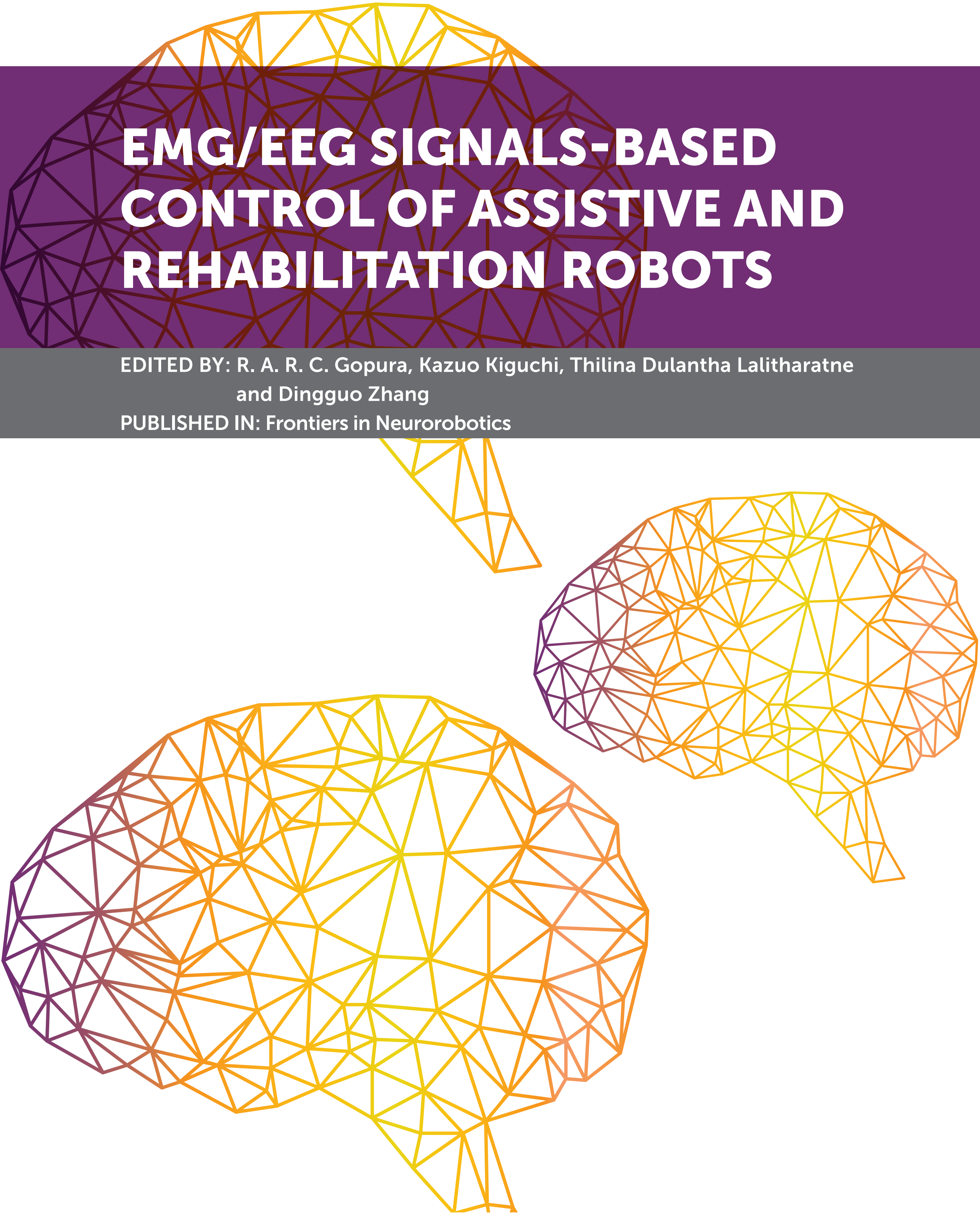


EMG/EEG SIGNALS-BASED CONTROL OF ASSISTIVE AND REHABILITATION ROBOTS

EDITED BY: R. A. R. C. Gopura, Kazuo Kiguchi, Thilina Dulantha Lalitharatne
and Dingguo Zhang
PUBLISHED IN: Frontiers in Neurorobotics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-592-0
DOI 10.3389/978-2-88974-592-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

EMG/EEG SIGNALS-BASED CONTROL OF ASSISTIVE AND REHABILITATION ROBOTS

Topic Editors:
R. A. R. C. Gopura, University of Moratuwa, Sri Lanka
Kazuo Kiguchi, Kyushu University, Japan
Thilina Dulantha Lalitharatne, Imperial College London, United Kingdom
Dingguo Zhang, University of Bath, United Kingdom

Citation: Gopura, R. A. R. C., Kiguchi, K., Lalitharatne, T. D., Zhang, D., eds. (2022). EMG/EEG Signals-based Control of Assistive and Rehabilitation Robots. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-592-0

Table of Contents

frontiers in Neurorobotics

EDITORIAL
published: 07 February 2022
doi: 10.3389/fnbot.2022.840321

Editorial: EMG/EEG Signals-Based Control of Assistive and Rehabilitation Robots

R. A. R. C. Gopura[1,2]*, **Thilina Dulantha Lalitharatne**[3], **Kazuo Kiguchi**[4] and **Dingguo Zhang**[5]

[1] Department of Mechanical Engineering, University of Moratuwa, Moratuwa, Sri Lanka, [2] Department of Medical Technology, University of Moratuwa, Moratuwa, Sri Lanka, [3] Dyson School of Design Engineering, Imperial College London, London, United Kingdom, [4] Department of Mechanical Engineering, Kyushu University, Fukuoka, Japan, [5] Department of Electronic and Electrical Engineering, University of Bath, Bath, United Kingdom

Keywords: EMG signal, EEG signal, assistive robot, rehabilitation robot, robot control (RC)

Editorial on the Research Topic

EMG/EEG Signals-Based Control of Assistive and Rehabilitation Robots

OPEN ACCESS

Edited and reviewed by:
Florian Röhrbein,
Technische Universität Chemnitz, Germany

****Correspondence:***
R. A. R. C. Gopura
gopura@gmail.com

Received: *21 December 2021*
Accepted: *12 January 2022*
Published: *07 February 2022*

Citation:
Gopura RARC, Lalitharatne TD, Kiguchi K and Zhang D (2022) Editorial: EMG/EEG Signals-Based Control of Assistive and Rehabilitation Robots. Front. Neurorobot. 16:840321. doi: 10.3389/fnbot.2022.840321

Assistive robots support persons with disabilities to improve their quality of life and independence. These robots make the life of a physically weak person more comfortable and useful to society. They sense, process sensory information, and execute actions that are helpful to the disabled person. Similarly, these robots carry out rehabilitation by understanding and augmenting the rehabilitation process through a robotic system. These robots are expected to assist different sensorimotor functions, therapeutic training, and assessment of the sensorimotor performance of the patients.

Assistive and rehabilitation robotic technologies have developed rapidly expanding the horizon of the field of robotics. These technologies are vital in the society in which around one billion among the growing global population experience some form of physical disability impacting everyday life. Assistive and rehabilitation robotic research, especially exoskeleton robotic research integrates advanced mechatronics and intelligent sensing to restore weak sensorimotor functions. Controlling these robots according to the requirement is one of the challenging tasks. Electromyography (EMG) signals have extensively been used to control assistive robots such as exoskeleton robots so that the wearer is free of actuating any additional device to control the robot. Additionally, many studies have been carried out recently to investigate the possibility of using Electroencephalography (EEG) signals to control robotic devices.

This Research Topic is aimed at creating a multidisciplinary forum of discussion on the recent advances in controlling assistive and rehabilitation robots presenting the diversity of the current approaches. It was expected to include Research Topics related to developments of controllers for upper limb/lower limb robotic exoskeletons, biomechanical investigations of robotic exoskeletons and other assistive robots, novel sensors for exoskeleton robots to generate human-like motions, EMG/EEG signals-based exoskeleton robot control, mathematical and physical algorithms for control of exoskeleton robots, assessment and benchmarking of exoskeleton robots' functionality, and clinical studies of assistive robot control. The Research Topic received overwhelming attention from the relevant researchers. Eleven articles have been selected to be published on this Research Topic after carrying out a rigorous peer-review process. Out of 11 articles, seven articles are based on the use of EMG signals for the control of assistive and rehabilitation robots. They are

- *Toward Hand Pattern Recognition in Assistive and Rehabilitation Robotics Using EMG and Kinematics* of Zhou et al.
- *Multi-Joint Angles Estimation of Forearm Motion Using a Regression Model* of Qin et al.
- *A Multi-Information Fusion Method for Gait Phase Classification in Lower Limb Rehabilitation Exoskeleton* of Zhang et al.
- *Face-Computer Interface (FCI): Intent Recognition Based on Facial Electromyography (fEMG) and Online Human-Computer Interface with Audiovisual Feedback* of Zhu et al.
- *Improved Motion Classification with an Integrated Multimodal Exoskeleton Interface* of Langlois et al.
- *Robust Torque Predictions from Electromyography Across Multiple Levels of Active Exoskeleton Assistance Despite Non-linear Reorganization of Locomotor Output* of George et al.
- *Control of Newly-Designed Wearable Robotic Hand Exoskeleton Based on Surface Electromyographic Signals* of Li et al.

Another three articles are based on the use of EEG signals.

- *Engagement Enhancement Based on Human-in-the-Loop Optimization for Neural Rehabilitation* of Wang et al.
- *Multi-Feature Input Deep Forest for EEG-Based Emotion Recognition* of Fang et al.
- *The Differences Between Motor Attempt and Motor Imagery in Brain-Computer Interface Accuracy and Event-Related Desynchronization of Patients with Hemiplegia* of Chen et al.

The article titled "*Evaluating Convolutional Neural Networks as a Method of EEG–EMG Fusion*" of Tryon and Trejos presented the use of EMG and EEG signals with a hybrid approach to control the assistive and rehabilitation robots.

Out of seven articles discussed on the EMG signals-based control of assistive and rehabilitation robots, in three articles EMG signals have been used for the control of upper-limb rehabilitation robots. Another two of the articles explained the uses of EMG signals for lower-limb exoskeleton robots and the rest of the articles described the classification of EMG signals to use for rehabilitation purposes.

The physiological and neurological differences among individuals can cause divergent responses to the same task, and the responses can further change considerably during training; both of these factors make engagement enhancement a challenge. This challenge can be overcome by training task optimization based on subjects' responses. In one article an engagement enhancement method based on human-in-the-loop optimization is proposed and the performance of the proposed method is demonstrated by the validation and comparison experiments. The results show that both subjects' surface EMG (sEMG) -based motor engagement and electroencephalography based neural engagement can be improved significantly and maintained at a high level.

Another article proposes a feature fusion and decision fusion method that combines EMG features and kinematic features for hand pattern recognition toward application in upper limb assistive and rehabilitation robotics. Ten normal subjects and five post-stroke patients participating in the experiments were tested with eight hand patterns of daily activities while EMG and kinematics were recorded simultaneously.

A prosthetic hand with high accuracy and robustness are necessary to improve the life quality of forearm amputees. The application of sEMG signals to control a prosthetic hand is challenging. The authors of one of the articles have proposed a time-domain Convolutional Neural Network (CNN) model for the regression prediction of joint angles in three degrees of freedom (DoFs) and 5-fold cross-validation was used to evaluate the correlation coefficient. The 3DOFs includes two wrist joint motion and one finger joint motion. In this study, CNN models were developed using combined EEG–EMG inputs to determine if they have potential as a method of EEG–EMG fusion that automatically extracts relevant information from both signals simultaneously. EEG and EMG signals were recorded during elbow flexion-extension and used to develop CNN models based on time-frequency and time-domain image inputs.

It is very important to have a natural, stable, and comfortable human-computer interface for controlling rehabilitation assistance robots to solve problems of patients who have lost limb control ability, such as upper limb amputation and high paraplegia. One article presents a complete limbs-free face-computer interface framework based on facial electromyography including offline analysis and online control of mechanical equipment.

One of the studies covered in this Research Topic explored the accuracy of the Brain-Computer-Interface (BCI) and event-related desynchronization between the two tasks. The articles identified the promise and the challenges faced by the field of assistive and rehabilitation robotic technologies and identified the critical need for additional prospective, in the design and control of relevant robots. Another article demonstrated a new wearable robotic hand exoskeleton with multi joints, higher degrees of freedom, and a larger range of motion. The exoskeleton hand comprises six linear actuators and can realize both independent movements of each digit and coordinative movement involving multiple fingers for grasp and pinch. The kinematic parameters of the hand exoskeleton were analyzed by using a motion capture system.

This Research Topic provides a snapshot of the current status of research on the *EMG/EEG Signals-Based Control of Assistive and Rehabilitation Robots*. All articles underlined both the promise and the challenges faced by the field of assistive and rehabilitation robots. The articles further recognized the need for additional perspective, in the control of these robotic devices. In summary, all the articles presented in the Research Topic have set a new trend in *EMG/EEG Signals-Based Control of Assistive and Rehabilitation Robots*.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 12 November 2020
doi: 10.3389/fnbot.2020.596019

Engagement Enhancement Based on Human-in-the-Loop Optimization for Neural Rehabilitation

Jiaxing Wang[1,2], *Weiqun Wang*[2]*, *Shixin Ren*[1,2], *Weiguo Shi*[1,2] *and Zeng-Guang Hou*[1,2,3]

[1] *School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China,* [2] *State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China,* [3] *Chinese Academy of Sciences Center for Excellence in Brain Science and Intelligence Technology, Beijing, China*

Enhancing patients' engagement is of great benefit for neural rehabilitation. However, physiological and neurological differences among individuals can cause divergent responses to the same task, and the responses can further change considerably during training; both of these factors make engagement enhancement a challenge. This challenge can be overcome by training task optimization based on subjects' responses. To this end, an engagement enhancement method based on human-in-the-loop optimization is proposed in this paper. Firstly, an interactive speed-tracking riding game is designed as the training task in which four reference speed curves (RSCs) are designed to construct the reference trajectory in each generation. Each RSC is modeled using a piecewise function, which is determined by the starting velocity, transient time, and end velocity. Based on the parameterized model, the difficulty of the training task, which is a key factor affecting the engagement, can be optimized. Then, the objective function is designed with consideration to the tracking accuracy and the surface electromyogram (sEMG)-based muscle activation, and the physical and physiological responses of the subjects can consequently be evaluated simultaneously. Moreover, a covariance matrix adaption evolution strategy, which is relatively tolerant of both measurement noises and human adaptation, is used to generate the optimal parameters of the RSCs periodically. By optimization of the RSCs persistently, the objective function can be maximized, and the subjects' engagement can be enhanced. Finally, the performance of the proposed method is demonstrated by the validation and comparison experiments. The results show that both subjects' sEMG-based motor engagement and electroencephalography based neural engagement can be improved significantly and maintained at a high level.

Keywords: human-in-the-loop optimization, EEG based neural engagement, sEMG based muscle activation, tracking accuracy, neural rehabilitation

Edited by:
Dingguo Zhang,
University of Bath, United Kingdom

Reviewed by:
Hong Zeng,
Southeast University, China
Gan Huang,
Université catholique de Louvain, Belgium
Mingming Zhang,
Southern University of Science and Technology, China

****Correspondence:***
Weiqun Wang
weiqun.wang@ia.ac.cn

Received: *18 August 2020*
Accepted: *22 September 2020*
Published: *12 November 2020*

Citation:
Wang J, Wang W, Ren S, Shi W and Hou Z-G (2020) Engagement Enhancement Based on Human-in-the-Loop Optimization for Neural Rehabilitation. Front. Neurorobot. 14:596019. doi: 10.3389/fnbot.2020.596019

1. INTRODUCTION

One of the most common sequela following stroke or cerebral injury is motor dysfunction, which seriously affects a person's quality of life. To regain their motor abilities, patients need to perform significant repetitive physical therapy, which is prone to boredom and often leads to low engagement. Previous studies have demonstrated that high levels of motivation and engagement are essential for obtaining relatively satisfactory rehabilitation outcomes (Tupper and Henley, 1987; Grant et al., 2004; Holden, 2005; Colombo et al., 2007). Developing a rehabilitation training method that can be used to reduce the boredom of the training tasks and promote engagement of the patients is therefore essential for post-stroke rehabilitation.

Engagement can be defined as a complex construct, which is driven by motivation and executed through active participation (Li et al., 2016). It was reported that motivating and empowering patients by providing them with the perception of control can improve patients' engagement, thus expediting the achievement of the patient's rehabilitation goals (Lenze et al., 2004; Dunn and Dougherty, 2005). Positive feedback can promote patient morale and engagement (Paolucci et al., 2012). Virtual reality (VR), which can be used to provide the task-specific training and intuitive multi-sensory feedbacks, has been therefore been widely applied in post-stroke rehabilitation.

The adaptive adjustment of the training task is often used for improving patient engagement. The challenge level of training tasks, which is one of the main sub-factors that contribute to engagement, can be adjusted to match a patients' motor abilities by use of training task adaptation (Csikszentmihalyi and Csikzentmihaly, 1990; Yannakakis and Hallam, 2009; Xu et al., 2017, 2018; Agarwal and Deshpande, 2019). In 2003, Krebs et al. proposed a performance-based progressive robotic therapy method (Krebs et al., 2003). In the Krebs's method, patients' active forces and motion-accuracy-based performance were used to customize the stiffness parameters of the robot controller and thus to maximize the recovery benefits (Krebs et al., 2003). Similarly, in 2014, an intelligent game engine was specifically designed for post-stroke rehabilitation, where the game parameters can be adjusted in real time according to patients' performance based on a Bayesian framework (Pirovano et al., 2014). Besides, interaction forces, muscle activity, or other physical or physiological parameters also have been used for training challenge adaption (Krebs et al., 2003; Novak et al., 2011; Luo et al., 2019).

However, due to the complexity of the training tasks and human-machine systems, the adaptive task adjustment-based engagement enhancement methods can hardly find an optimal design of the training tasks. This can be obtained via the optimization method, though this has rarely been studied. Besides, considering that physiological and neurological differences among individuals can cause divergent responses to the same task, and the responses can further change considerably during the training (Gordon and Ferris, 2007; Zelik et al., 2011; Jackson and Collins, 2015; Selinger et al., 2015; Quesada et al., 2016), subjects' physiological variations or responses also need to be considered during the training task optimization. Subjects' responses based training task optimization belongs to human-in-the-loop optimization (HILO).

To the best of our knowledge, HILO method-based training task optimization has rarely been studied. All the key steps of the HILO, including the training task modeling and design of the objective function and the optimization algorithm, can affect the optimization results. On one hand, the parameters used for modeling the training task should be sensitive to the engagement variation, based on which subjects' engagement can be improved through the parameter optimization. On the other hand, adding human responses to the engagement enhancement optimization loop also makes the optimization difficult to implement due to the time-varying dynamics of the subjects, such as the self-adaptation ability, the strong history dependence, and other complex neurocognitive factors (Gordon and Ferris, 2007; Selinger et al., 2015). Both the objective function and the optimization algorithm should therefore be insensitive to human dynamic variation and noises.

In this paper, an HILO-based engagement enhancement method is proposed. The original contributions of this study can be summarized as follows: ① an optimization-based engagement enhancement method is proposed, ② and the proposed HILO method is tolerant of both measurement noises and human adaptation.

Firstly, an interactive speed-tracking riding game is designed as the training task. In the task, subjects are asked to track the reference trajectory, which is constructed by four reference speed curves (RSCs), as accurately as possible. Each RSC is modeled using a piecewise function and determined by the starting velocity, transient time, and end velocity. By parameterizing the RSC, it is possible to optimize the difficulty of the training task, which is a key factor affecting a user's engagement level.

Then, the objective function is designed by consideration of the tracking accuracy (TA) and the muscle activation (MA), based on which subjects' physical and physiological responses can be evaluated simultaneously. By maximizing the subject's TA and MA concurrently, the difficulty of the training task can be optimized to match subject's current motor ability and physiological state.

Moreover, the covariance matrix adaptation evolution strategy (CMA-ES) is used to optimize the parameters of the RSCs (Hansen, 2006; Akimoto et al., 2012; Zhang et al., 2017; Maki et al., 2020). In the CMA-ES, neither objective function values nor their derivatives are used directly, and each generation is evaluated independently. It is therefore relatively tolerant of both measurement noises and human adaptation. By optimization of the RSCs persistently, the objective function can be maximized and subject engagement enhanced.

Finally, the performance of the proposed HILO method is demonstrated through a comparison experiment. The results show that both TA and MA can be improved significantly. Moreover, the subjects' neural engagement can also be improved significantly and maintained at a high level.

2. TASK MODELING AND OPTIMIZATION

An HILO method is designed to enhance the subjects' engagement in this study. Details of the HILO method are given in the following text.

2.1. Modeling the Training Task

Based on the previous study (Wang et al., 2019), an interactive speed-tracking riding game is designed as the training task, which

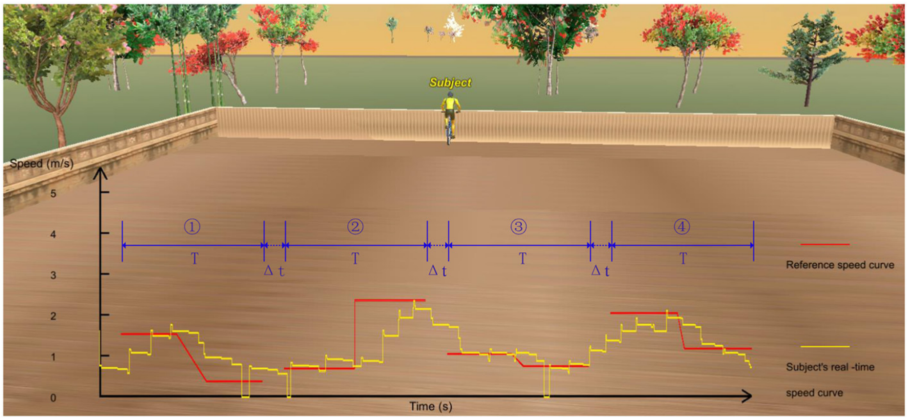


FIGURE 1 | The virtual scene of the designed speed-tracking riding game. The four red lines represent the four RSCs, which are used to construct the reference trajectory in each generation, and the yellow line represents subject' actual speed tracking trajectory.

can be seen from **Figure 1**. During the training, subjects need to try their best to track the reference trajectory.

To increase the complexity of the task and, meanwhile, facilitate optimization, four relatively simple RSCs were used to construct the reference trajectory in each generation, which can be seen from **Figure 1**. The T and Δt are the period of each RSC and the time interval between the two adjacent RSCs, respectively. In this study, the T and Δt were set to 12 and 3 s, respectively. Δt is designed to give the subjects enough time to adjust their riding speeds to better complete the subsequent tracking task.

Specifically, each RSC is determined by three parameters: starting velocity (v^s), transient time (t^{tra}), and end velocity (v^e). The definition of these three parameters is given in **Figure 2**. It can be seen that each RSC can be defined as a piecewise function of time, which is given by the following:

$$V^{\mathrm{ref}}(t) = \begin{cases} v^s & t \in [0, \frac{\mathrm{T}-t^{\mathrm{tra}}}{2}] \\ v^s + \frac{v^e - v^s}{t^{\mathrm{tra}}}(t - t^{\mathrm{tra}}) & t \in (\frac{\mathrm{T}-t^{\mathrm{tra}}}{2}, \frac{\mathrm{T}+t^{\mathrm{tra}}}{2}) \\ v^e & t \in [\frac{\mathrm{T}+t^{\mathrm{tra}}}{2}, \mathrm{T}] \end{cases} \qquad (1)$$

Based on the parameterized model, the difficulty of the training task, which is a key factor affecting the engagement, can be optimized. A wide range of possible RSCs can be obtained by Equation (1), and some examples of possible RSCs are given in **Figure 2B**.

In this study, constraints given in Equation (2) are used to avoid appearance of some weird RSCs, such as too high reference speeds and sharp change of the speed.

$$\begin{aligned} 0 \leq & v^s \leq 6, \; 0 \leq v^e \leq 6, \\ t^{\mathrm{tra}} & \geq \frac{|v^e - v^s|}{6} \end{aligned} \qquad (2)$$

where, the units of v^s (v^e) and t^{tra} are meters per second (m/s) and seconds (s), respectively.

2.2. Design of the Objective Function

Both electroencephalography (EEG) and surface electromyogram (sEMG)-based physiological responses, which can reflect subjects' engagement levels during the training, can be used to construct the objective function (Zimmerli et al., 2013; Tacchino et al., 2016). Compared with sEMG, the EEG signals are much weaker (microvolt level), and they can be easily contaminated by the environment noises or the subjects' physiological variation, such as emotional fluctuation. If the EEG based objective function is used for the HILO, the parameters to be optimized can hardly converge to the optima. In this paper, sEMG-based MA is thus chosen to measure subjects' physiological response. Besides, the subjects' physical response is evaluated by TA. By maximizing subject's TA and MA concurrently, the difficulty of the training task can be optimized to match the subject's current motor ability and physiological state. On one hand, a relatively high TA can be obtained when the tracking task is designed relatively easily. However, speed-tracking tasks that are too easy can easily lead to a phenomenon where a subject's MA is relatively low, which is not beneficial for the restoration of muscle strength. On the other hand, a relatively high MA can be obtained when the tracking task is designed relatively difficult. Tasks that are too difficult, however, can cause the subjects to become discouraged and unwilling to continue the training. Simultaneously maximizing TA and MA can result in a suitable challenging task for a specific subject, thus enhancing the subjects' engagement. In this paper, TA and MA are therefore used to construct the objective function.

Specifically, the TA is given by the following:

$$F_i^{\mathrm{TA}} = -\frac{||\boldsymbol{V}_i^{\mathrm{ref}} - \boldsymbol{V}_i^{\mathrm{act}}||_2}{\sqrt{\mathrm{N}}}, \quad i = 1, 2, 3, 4 \qquad (3)$$

where, $||*||_2$ means the calculation of the L2-norm. $\boldsymbol{V}_i^{\mathrm{ref}} \in \mathbb{R}^{\mathrm{N}}$ and $\boldsymbol{V}_i^{\mathrm{act}} \in \mathbb{R}^{\mathrm{N}}$ are the reference speed vector and subject's actual speed vector with 100 Hz sample rate acquired during tracking the ith RSC in each generation. The period of each RSC is 12 s, therefore, N is equal to 1,200. In this study, the subject's actual

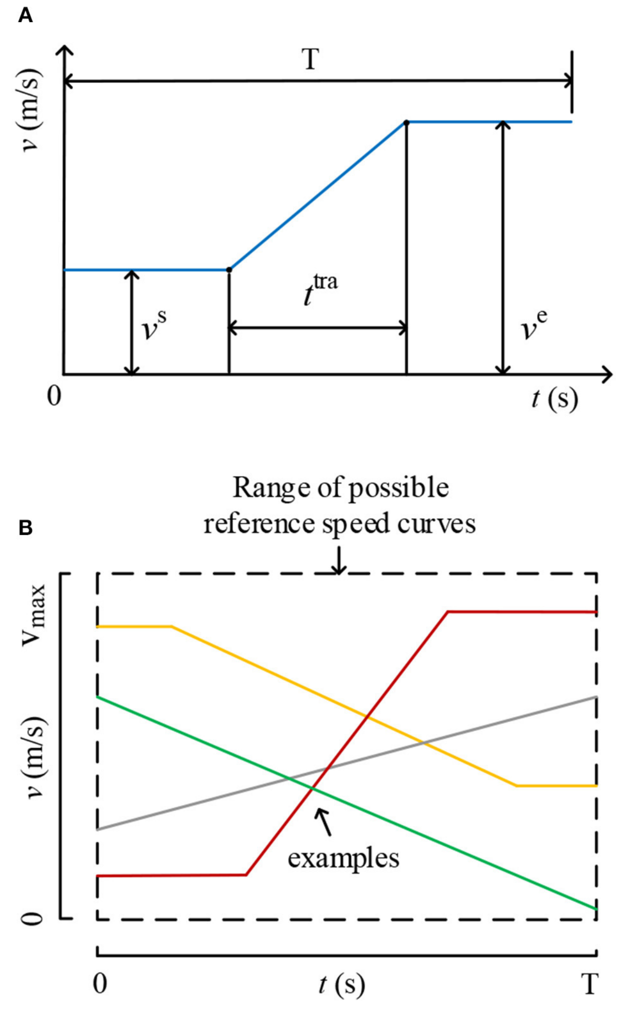


FIGURE 2 | (A) Parameterization of each RSC. **(B)** Examples of possible RSCs.

speeds are collected using a data acquisition card and transmitted to the computer via TCP/IP protocol.

As for the sEMG-based MA, it has been proved that, when subjects are focused on the training, the root mean square (RMS) of sEMG signals can become bigger (Zimmerli et al., 2013). In this paper, RMS is consequently used to indicate subjects' MA.

$$F_i^{\mathrm{MA}} = \frac{||\boldsymbol{S}_i^{\mathrm{EMG}}||_2}{\sqrt{\mathrm{M}}}, \quad i = 1, 2, 3, 4 \tag{4}$$

where, $\boldsymbol{S}_i^{\mathrm{EMG}} \in \mathbb{R}^{\mathrm{M}}$ represents the amplitude vector of the acquired sEMG signals with 400 Hz sample rate acquired during tracking the ith RSC in each generation, and M is equal to 4,800.

The muscles contributing to cycling motion, including rectus femoris (RF), hamstring, soleus, and gastrocnemius, are mainly considered. During the pre-experiment, it was found that the RF muscle had the highest activation during the cycling training, and it is therefore used to calculate the MA in this study. Delsys Trigno$^{\mathrm{TM}}$ device with a 1111.11 Hz sample rate was used to monitor muscle activities during cycling. The raw sEMG signals were first filtered by a band-pass butterworth filter (10–400 Hz) and a notch filter (50 Hz) to reduce the effects of noise and power line interference. Then, the subjects' average MA can be calculated using Equation (4).

Finally, the objective function, which is equal to the weight sum of the TA and MA, can be given as follows:

$$F_i^{\mathrm{OBJ}} = F_i^{\mathrm{TA}} + \alpha F_i^{\mathrm{MA}}, \quad i = 1, 2, 3, 4 \tag{5}$$

where, F_i^{TA} and F_i^{MA} represent the values of TA and MA of the ith sub-racking task in each generation, respectively. α is a scaling coefficient to weight F_i^{TA} and F_i^{MA}, and it is set to 1 in this study.

2.3. CMA-ES Based HILO

In this study, the optimization problem for engagement enhancement can be defined as follows.

Parameters to be optimized are the following:

$$\boldsymbol{m}_i = [v_i^{\mathrm{s}},\ t_i^{\mathrm{tra}},\ v_i^{\mathrm{e}}], \quad i = 1, 2, 3, 4 \tag{6}$$

The objective function to be maximized is the following:

$$F_i^{\mathrm{OBJ}} = F_i^{\mathrm{TA}} + F_i^{\mathrm{MA}}, \quad i = 1, 2, 3, 4 \tag{7}$$

Constraints to be satisfied are the following:

$$\begin{aligned} &0 \le v_i^{\mathrm{s}} \le 6,\ 0 \le v_i^{\mathrm{e}} \le 6, \\ &t_i^{\mathrm{tra}} \ge \frac{|v^{\mathrm{e}} - v^{\mathrm{s}}|}{6}, \quad i = 1, 2, 3, 4 \end{aligned} \tag{8}$$

It can be seen that the optimization problem of this paper is strongly non-linear, and it can be easily disturbed by the time-varying dynamics of the subjects. Therefore, CMA-ES, which is relatively tolerant of both measurement noises and human adaptation, is applied to optimize the training task in this paper. No gradient calculation is involved in the CMA-ES, which makes this method robust and feasible even for a non-continuous problem. With each iteration, new task-setting parameters are generated stochastically using a multivariate normal distribution, and the distribution parameters, including the mean vector, the covariance matrix, and the evolution paths, are updated with successful candidate solutions and their objective value ranking. In this paper, the algorithm of the CMA-ES (Hansen, 2006; Maki et al., 2020) based HILO is given in Algorithm 1.

In each generation, four groups of the RSC parameter settings, $(\boldsymbol{m}_i)_{i=1,2,3,4}$, are generated stochastically using a multivariate normal distribution $\mathcal{N}(\bar{\boldsymbol{m}}, \sigma^2\boldsymbol{C})$, to form the tracking trajectory of the current generation.

$$\boldsymbol{m}_i = \bar{\boldsymbol{m}} + \sigma\mathcal{N}(0, \boldsymbol{C}), \quad i = 1, 2, 3, 4 \tag{9}$$

where, $\bar{\boldsymbol{m}} = [\bar{v^{\mathrm{s}}}, \bar{t^{\mathrm{tra}}}, \bar{v^{\mathrm{e}}}]$. Specifically, $\bar{\boldsymbol{m}}$ is the mean vector of the parameters to be optimized, and it determines the search space of the $\boldsymbol{m}_i$. σ is the step parameter, which determines the size and intensity of the search range. $\boldsymbol{C}$ is the covariance matrix, which

Algorithm 1: CMA-ES based HILO.

```
$n$: number of parameters to be optimized, defaults to 3.
$\bar{\boldsymbol{m}} \in \mathbb{R}^n$, mean vector, initialized with [2.5, 4, 5.5].
$\boldsymbol{p_c}, \boldsymbol{p_\sigma} \in \mathbb{R}^n$, evolution paths, initialized with 0.
$\sigma$: step size, initialized with 2.
$\boldsymbol{C} \in \mathbb{R}^{n \times n}$: covariance matrix, initialized with $\boldsymbol{I}$.
$\lambda$: population size of each generation, defaults to 4.
$\lambda^{\text{opt}}$: number of candidate population, defaults to 2.
$\boldsymbol{w} \in \mathbb{R}^{\lambda^{\text{opt}}}, \mu_w$: weight constants for $\bar{\boldsymbol{m}}$, $\sigma$ and $\boldsymbol{C}$ update.
    $\boldsymbol{w}(i) = \frac{log(\lambda^{\text{opt}} + 1/2) - log(i)}{\sum_{i=1}^{\lambda^{\text{opt}}} (log(\lambda^{\text{opt}} + 1/2) - log(i))}, i = 1, .., \lambda^{\text{opt}}$.
    $\mu_w = 1/\sum_{i=1}^{\lambda} \boldsymbol{w}(i)^2$.
$\mu_n$: approximated norm of the expected value of n-dimension normal distribution.
    $\mu_n = n^{1/2}(1 - 1/4n + 1/21n^2)$.
$c_c, c_\sigma$: cumulation factors for evolution.
    $c_c = (4 + \mu_w/n)/(n + 4 + 2\mu_w/n)$.
    $c_\sigma = (\mu_w + 2)/(n + \mu_w + 5)$.
$d_\sigma$: damping factor for $\sigma$ update.
    $d_\sigma = 1 + 2max(0, sqrt((\mu_w - 1)/(n + 1)) - 1) + c_\sigma$.
$c_1, c_\mu$: learning rate for covariance update.
    $c_1 = 2/((n + 1.3)^2 + \mu_w)$.
    $c_\mu = min(1 - c_1, 2(\mu_w - 2 + 1/\mu_w)/((n + 2)^2 + \mu_w))$.
for each generation do
    for $i = 1 \rightarrow \lambda$ do
        $\boldsymbol{m}_i = \bar{\boldsymbol{m}} + \sigma\mathcal{N}(0, \boldsymbol{C})$.
        Generate the RSC and $\boldsymbol{V}_i^{\text{ref}}$ with $\boldsymbol{m}_i$ using Equation (1).
    end for
    Tracking, acquire $\boldsymbol{V}_i^{\text{act}}$ and $\boldsymbol{S}_i^{EMG}$, where $i = 1, ..., \lambda$.
    for $i = 1 \rightarrow \lambda$ do
        Compute $F_i^{\text{TA}}$ with $\boldsymbol{V}_i^{\text{ref}}$ and $\boldsymbol{V}_i^{\text{act}}$ using Equation (3).
        Compute $F_i^{\text{MA}}$ with $\boldsymbol{S}_i^{\text{EMG}}$ using Equation (4).
        Compute $F_i^{\text{OBJ}}$ with $F_i^{\text{TA}}$ and $F_i^{\text{MA}}$ using Equation (7).
    end for
    Get $\boldsymbol{I}^{\text{opt}}$: the indices of the top $\lambda^{\text{opt}}$ values of $F^{\text{OBJ}}$ in descending order.
    $\boldsymbol{m}^{\text{opt}} = \{\boldsymbol{m}_i\}_{i \in \boldsymbol{I}^{\text{opt}}}$.
    $\boldsymbol{m_d} = (\sum_{i=1 \rightarrow \lambda^{\text{opt}}} (\boldsymbol{m}^{\text{opt}}(i))\boldsymbol{w}(i) - \bar{\boldsymbol{m}})/\sigma$
    Update $\boldsymbol{p_\sigma} = (1 - c_\sigma)\boldsymbol{p_\sigma} + (c_\sigma(2 - c_\sigma)\mu_w)^{1/2}\frac{1}{\sqrt{\boldsymbol{C}}}\boldsymbol{m_d}$
    Update $\boldsymbol{p_c} = (1 - c_\sigma)\boldsymbol{p_c} + (c_c(2 - c_\sigma)\mu_w)^{1/2}\boldsymbol{m_d}$
    Update $\bar{\boldsymbol{m}} = \sum_{i=1 \rightarrow \lambda^{\text{opt}}} (\boldsymbol{m}^{\text{opt}}(i))\boldsymbol{w}(i)$.
    Update $\sigma = \sigma \exp[\frac{c_\sigma}{d_\sigma}(\frac{||\boldsymbol{p_\sigma}||}{\mu_w} - 1)]$
    Update $\boldsymbol{C} = \boldsymbol{C} + c_1(\boldsymbol{p_c}\boldsymbol{p_c}^T - \boldsymbol{C}) + c_\mu(\sigma^2\boldsymbol{m_d}^2 - \boldsymbol{C})$
end for
```

determines the shape of the distribution. In this study, $\bar{\boldsymbol{m}}$, σ and $\boldsymbol{C}$ are initialized with [2.5, 4, 5.5], 2 and $\boldsymbol{I}$, respectively.

When the tracking task in each generation is finished, the average TA and MA can be calculated according to the subjects' responses. The value of the objective function can consequently be calculated by Equation (7). Then, according to the value ranking of $(F_i^{\text{OBJ}})_{i=1,2,3,4}$, λ^{opt} parameter settings, $\boldsymbol{m}^{\text{opt}}$, can be obtained, and these are used to update the two evolution paths, $\boldsymbol{p_\sigma}$ and $\boldsymbol{p_c}$. Finally, based on the $\boldsymbol{m}^{\text{opt}}$, $\boldsymbol{p_\sigma}$, and $\boldsymbol{p_c}$, $\bar{\boldsymbol{m}}$, σ, and $\boldsymbol{C}$, which are used to generate the tracking trajectory of the next generation, can also be updated. The tracking trajectory in each generation can therefore be updated continuously by using the current multivariate normal distribution $\mathcal{N}(\bar{\boldsymbol{m}}, \sigma^2\boldsymbol{C})$. It can be seen that, by using the proposed CMA-ES based HILO method, the training task can be optimized automatically and constantly to achieve engagement enhancement.

2.4. Neural Engagement Evaluation Method

Since the purpose of the proposed optimization method is to enhance and maintain subjects' engagement during the rehabilitation training, the subjects' neural engagement levels were also evaluated in this study.

Neural engagement, which is an essential factor in promoting neural reorganization and compensation, is considered to be proportional to the level of concentration (attention) during the rehabilitation training (Park et al., 2014; Li et al., 2016). Previous researches have demonstrated that EEG signals in the theta and beta bands can be used to quantitatively represent subjects' attention states (Mann et al., 1992; Harmony et al., 1996). Good performance and high attention level have been proven to be related to the decrease of the theta rhythm power and the increase of the beta rhythm power (Kropotov, 2009; Gürkök et al., 2011; Arns et al., 2012; Loo and Makeig, 2012; Marshall et al., 2013). The EEG-based theta to beta power ratio (TBR) was thus used to measure subjects' neural engagement, which can be given by the following:

$$\begin{aligned} TBR &= -\frac{E(\text{theta})}{E(\text{beta})} \\ E_{\text{n}} &= \frac{\sum_{i=1}^{5} TBR(i)}{5} \end{aligned} \tag{10}$$

where E(theta) and E(beta) represent the energy of theta and beta bands in the latest 3 s, respectively. TBR was calculated every 3 s. E_{n}, which is equal to the mean of the latest 5 TBR values, was used to indicate subjects' attention and neural engagement. A high E_{n} represents a high level of neural engagement.

By considering that EEG activities in the frontal and temporal lobes are most related to human engagement levels (Barkley et al., 1992; Mann et al., 1992), EEG signals acquired from these two brain regions can be used to compute E_n, which can be seen from **Figure 3**. However, EEG signals, especially collected during cycling, can be easily contaminated by ocular artifacts (OAs) and EMG (Frølich et al., 2015; Kline et al., 2015; Zink et al., 2016). Many studies focused on eliminating the artifacts have been conducted, but the results are still not satisfying.

For the term of OAs, blinking or moving the eyes can produce large electrical potential, which will spread across scalp and contaminate the EEG signals. EEG signals in the forehead (FP1 and FP2) are most susceptible to OAs (Babu and Prasad, 2011). For the term of EMG artifacts, subject movement (riding) can introduce some muscle artifacts to EEG signals inevitably and the muscle artifacts are mainly distributed at the outer electrode

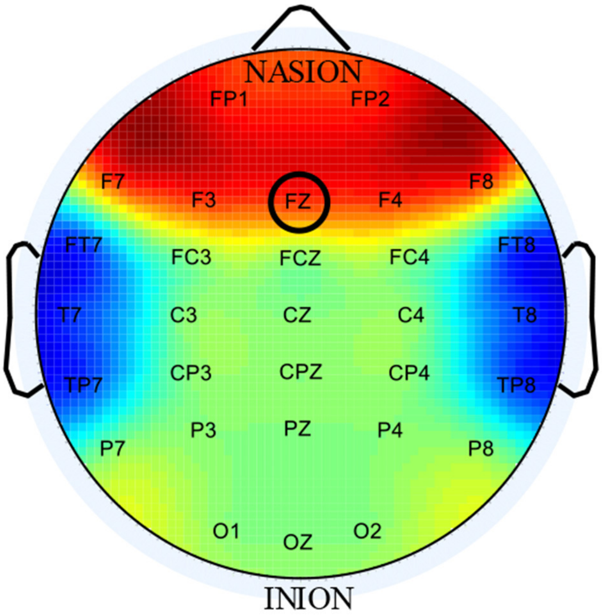


FIGURE 3 | International 10–20 system. The red and blue areas represent the frontal and temporal regions, respectively.

sites, such as the temporary region (Muthukumaraswamy, 2013). To reduce the effect of artifacts on EEG signals, therefore, only signals acquired from FZ electrode are used to indicate the subject's neural engagement, which can be seen from **Figure 3**.

NeuroScan system with 256 Hz sample rate was used to acquire subjects' EEG signals. Baseline drift, which is mainly caused by spontaneous brain waves, was avoided by the removing mean method. Then the theta (3–8 Hz) and beta (12–30 Hz) bands were extracted by fast Fourier transform, and subjects' neural engagement can be calculated by Equation (10) finally.

3. EXPERIMENT AND RESULTS

A contrast experiment was conducted to validate the feasibility of the proposed HILO method for engagement enhancement. The experiment was approved by the ethics committee of the Institute of Automation, Chinese Academy of Sciences. All the recruited subjects were informed of the experiment contents and signed the consent forms before the experiment.

3.1. Experiment Design

The interactive speed-tracking riding game was used as the training task for both the control group (CG) and the experiment group (EG). More specifically, during the training, subjects should track the reference trajectory, which is constructed by four RSCs, as accurately as possible. For the CG, the proposed HILO based engagement enhancement method was not used, which was used for the EG. For the CG, the RSCs displayed on the screen were thus given randomly under the constraints of Equation (8). But for the EG, the RSCs can be optimized continuously by the HILO.

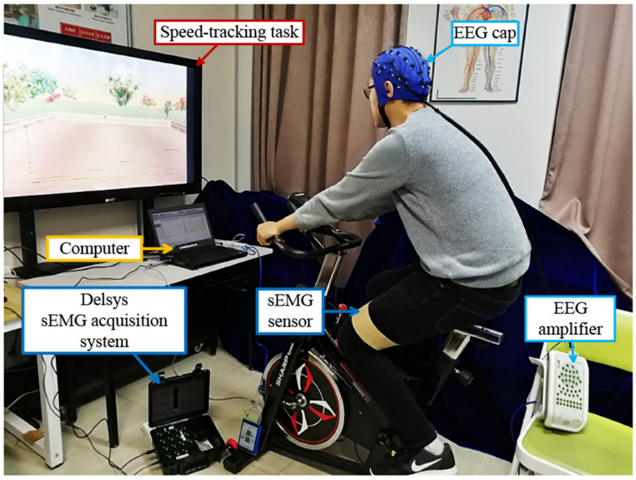


FIGURE 4 | One of the experiment scenes during the training process.

A total of 10 healthy subjects (eight men and two women aged from 24 to 29 years old), numbered from S1 to S10, were recruited to participate in the experiment. None of them knew the design process or the purpose of this study. They participated in the experiments for both CG and EG. Each experiment took about 25 min, as is similar to the commonly used period of each post-stroke rehabilitation session. The interval between the two experiments was about 20 min to give subjects enough time to rest and thus minimize the influence of the previous experiment on the next experiment results.

During the previous 2 days before the experiment, the subjects were required to not engage in any vigorous exercises to prevent muscle fatigue and avoid affecting the collected sEMG data. To reduce possible bias, we shuffled the sequence of the experiments for the CG and EG. Subjects were able to choose which experiment to conduct first. Before the experiment, one Delysis sensor was placed on the subjects' RF muscle to acquire their sEMG signals during training, which are used to calculate their MA. Besides, an EEG cap needs to be worn to acquire subjects' EEG signals, which is used for neural engagement evaluation. All subjects received the same task instructions. They were supposed to try their best to track the reference trajectory. One of the experiment scenes during the training process is given in **Figure 4**. Besides, during the training, they should keep their upper body motionless to reduce muscle artifacts caused by movement.

3.2. Analysis of TA and MA

For the EG, one subject's reference/actual speed curve variations are given in **Figure 5**, and his TA and MA during the training are given in **Figure 6**.

It can be seen from **Figures 5**, **6** that, at the beginning of the experiment, the shape of the four RSCs varied greatly, with the maximum speed reaching 6 m/s. However, the maximum cycling speed that the subject can reach was about 4 m/s. The subject could not follow the RSCs, which led to a low TA. From the

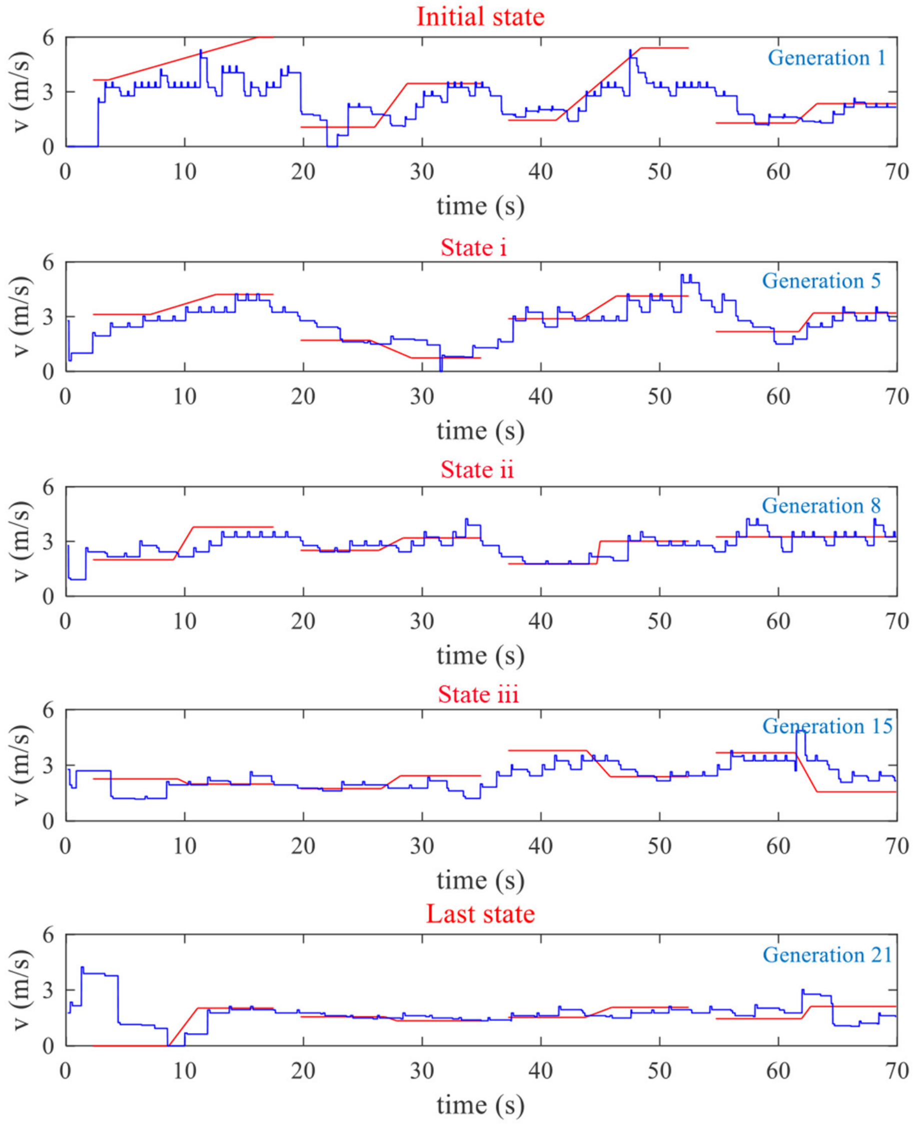


FIGURE 5 | One subject's reference/actual speed curves in different generations during the training for the EG. RSCs and actual speed curves are represented by red lines and blue lines, respectively.

initial state to state i, TA was mainly optimized to ensure that the RSCs could be tracked by the subject. The purpose of the process from state i to state ii was to improve MA as much as possible under the premise of a high TA. In generation 8, both the TA and MA were acceptable. The purpose of the process from state ii to iii was thus to maintain the subject's high TA and MA. By the 15th generation, the subject was exhausted due to a long time of training. It can be seen that from state iii to the last state, the reference speed gradually decreased to ensure that the subjects could still track the RSCs well.

Boxplots of the average TA and MA of all subjects in different generations are shown in **Figures 7**, **8**, respectively. In each box, the central line represents the median value, the dot represents the mean value, the edges of the box are the 25th and 75th percentiles. Moreover, the Wilcoxon signed-rank test results also indicate that there are significant differences between the CG and EG for both TA and MA (TA: p-value = 2.14e-04 < 0.0001; MA: p-value = 2.13e-04 < 0.0001).

It can be seen from **Figure 7** that, at the beginning of the experiment, the TA for both EG and CG were relatively low since the subjects cannot track the randomly generated RSCs accurately. However, for the EG, TA can be improved obviously due to the proposed HILO. Besides, as the experiment went on, subjects became fatigued gradually, which resulted in a further decrease in the TA for the CG. This phenomenon can cause the subjects discouraged and unwilling to continue the training. However, for the EG, the difficulty of the training task can be adaptively reduced to maintain a relatively high

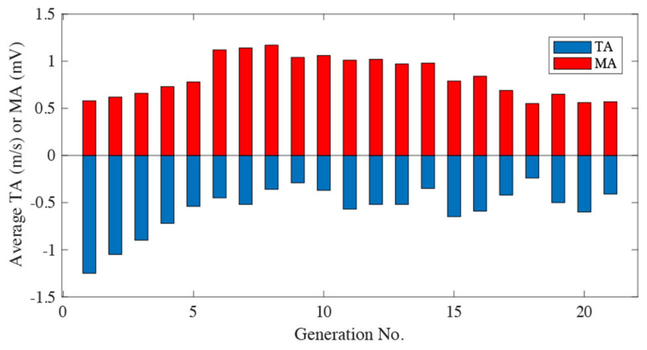


FIGURE 6 | The subjects' corresponding TA and MA variations during the experiment in the EG.

TA. The proposed HILO method can thus result in a suitable challenging task for a specific subject to improve the enthusiasm of the subjects.

It can be seen from **Figure 8** that the difference of the MA between the EG and CG was not obvious in the early stage of the experiment. One possible reason is that, in the early stage of the experiment, the main purpose of the optimization was to improve the TA due to the subjects' relatively bad tracking performance, during which the MA didn't change much for the EG.

Besides, for the CG, the ranges of both TA and MA in each generation fluctuated larger than that for the EG, especially in the later state of each experiment. One of the possible reasons is that, for the CG, the RSCs of each generation were given randomly regardless of subjects' motor ability or physiological status. The TA and MA therefore fluctuated with the variation of the given RSCs.

3.3. EEG-Based Neural Engagement Evaluation

One of the subjects' EEG-based engagement variation curves and fitting curves based on a first-order linear function are given in **Figure 9**. The fitting curves' slopes represent the variation trends of the subject's neural engagement during training. It can be seen from **Figure 9** that, for the EG, with the progress of tracking task, the values of E_n gradually increased, and these are decreased for the CG. It denotes that the neural engagement of the subject for the EG showed different degrees of improvement by using the proposed method. However, for the CG, neural engagement can be increased to some extent in the early stage (Wang et al., 2019) but dropped obviously after that.

The mean values of the E_n for the 10 subjects, and the results of the significant test about the neural engagement between CG and EG by using Wilcoxon signed-rank tests are given in **Figure 10**. Compared to the neural engagement in the CG, subjects' neural engagement in the EG can be improved significantly.

To clearly show the brain activity variation during the speed-tracking task, one subject's time-frequency spectra, which were obtained by short-time Fourier transformation of the EEG signals (Wang et al., 2018), are given in **Figure 11**. From the figure we can see that, for the EG, the energy of the beta rhythm (12–30 Hz) increased gradually, and the energy of the theta rhythm (3–8 Hz) decreased gradually after around 13 min. For the CG, there was a little fluctuation of the EEG spectrum in different frequency bands. Since good performance and high neural engagement are related to a phenomenon of decreased theta rhythm power and increased beta rhythm power, the feasibility of the proposed HILO method in engagement enhancement can be further proved by **Figure 11**.

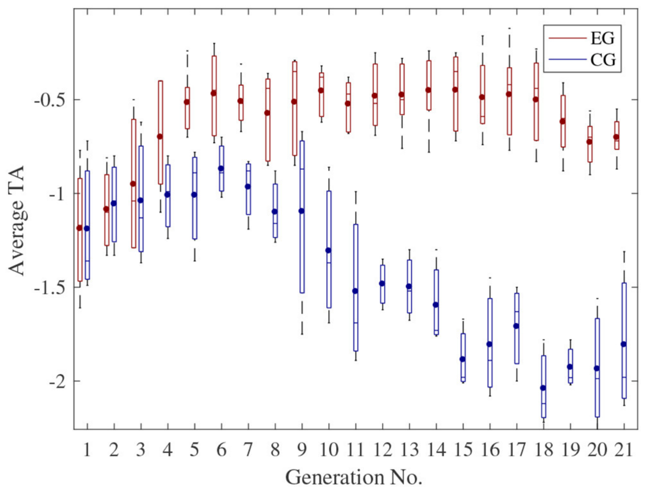


FIGURE 7 | Boxplot of the average TA of all the subjects.

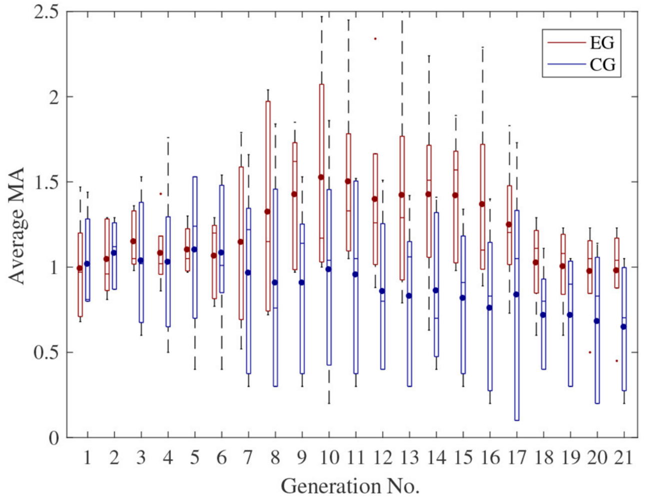


FIGURE 8 | Boxplot of the average MA of all the subjects in different generations and different groups.

4. DISCUSSION

To maximize engagement during therapy and prevent frustration, it is essential to design rehabilitation exercises in such a way where they challenge patients at a difficulty level neither too simple nor too difficult (Choi et al., 2011; Metzger

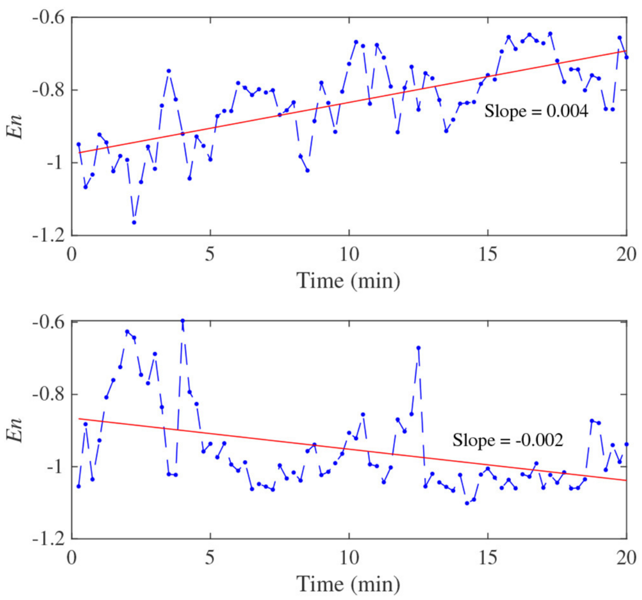


FIGURE 9 | One of the subjects' EEG-based engagement variation curves and fitting curves based on a first-order linear functions. A higher E_n represents a higher engagement. The up and down figures are results for EG and CG, represently.

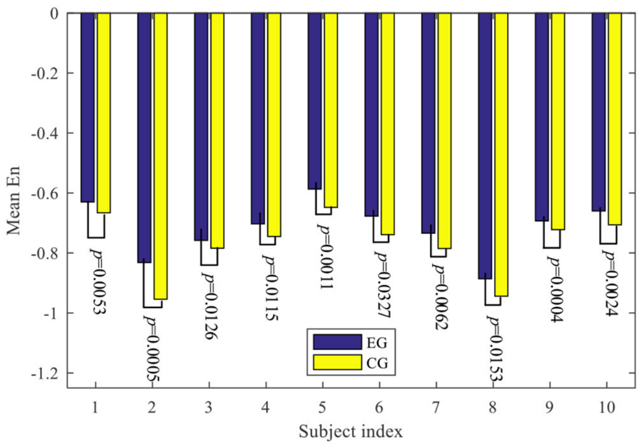


FIGURE 10 | The mean values of the E_n for the 10 subjects, and the results of the significant test about the neural engagement between CG and EG.

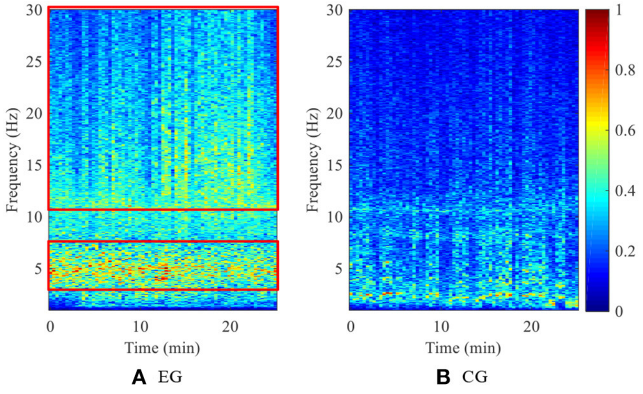


FIGURE 11 | One of the subjects' time-frequency spectra during the whole training. **(A)** EG. **(B)** CG.

et al., 2014). The ability to select and maintain an engaging and challenging training difficulty level in post-stroke rehabilitation, however, remains an open challenge. In this paper, we presented an HILO based training task optimization method by which the difficulty levels of the training task can be optimized continuously to well match the subject's current motor ability and physiological state.

Several strategies have been proposed for online decision making to modify task parameters and modulate its difficulty. For example, in Metzger et al. (2014), the difficulty of the training task is adjusted based on the completion of the task to maintain the training performance of patients in a certain range. Besides, interaction forces, muscle activity, or other kinematic or physiological parameters have also been used for training challenge adaption (Krebs et al., 2003; Novak et al., 2011; Luo et al., 2019). However, due to the complexity of the training tasks and human-machine systems, the adaptive task adjustment based engagement enhancement methods can hardly find an optimal design of the training tasks, which can be found by the optimization method.

In this paper, according to subjects' current physiological state and task performances, i.e., MA and TA, the training task parameters can be optimized continuously, to make sure that the current task parameter settings are nearly optimal for engagement enhancement. The proposed optimization method can be termed as "greedy" optimization since only the subject's performance in the latest generation rather than overall superimposed performance is considered during the optimization. In this way, the system can quickly converge to the "greedy" optimal state to improve the immediate engagement. However, during the experiment, it was found that the system can fall into a local optimal situation sometimes, which should be improved in the future.

In clinical settings, selection of the training difficulty and its adaptation over the course of therapy is often determined by the experience of trained therapists and their subjective perception of a patient's abilities (Metzger et al., 2014). Our proposed method can effectively avoid the mismatch between the difficulty of the task set manually and the patients' abilities. Moreover, by considering that active engagement of the human motor and neural system is essential for functional rehabilitation, the proposed method is promising for transfer to the rehabilitation of post-stroke patients. In the future, more experiments are to be conducted to further validate the feasibility of the proposed method for enhancement of the post-stroke patients' engagement and improvement of the rehabilitation outcomes.

5. CONCLUSION

In this paper, an HILO-based engagement enhancement method is proposed to enhance subjects' engagement. Firstly, subjects are asked to track the reference trajectory, which is constructed by four RSCs, as accurately as possible. After finishing the tracking task of each generation, the value of the designed objective function, which is equal to the sum of the TA and MA, can be calculated according to subjects' responses. Then, CMA-ES is used to generate the optimal parameters of the RSCs periodically. By optimization of the reference trajectory continuously, the objective function can be maximized and subject engagement enhanced. Finally, the feasibility of the proposed HILO method in engagement enhancement is validated through the comparison experiment on 10 subjects. Experiment results show that both TA and MA can be improved significantly ($p < 0.0001$). Moreover, all the recruited subjects' EEG based neural engagement can also be improved significantly ($p < 0.01$) and maintained at a high level by using the proposed method.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The experiment was approved by the ethics committee of the Institute of Automation, Chinese Academy of Sciences. The ethics approval number is IA-201947. All the recruited subjects were informed of the experiment contents and signed the consent forms before the experiment.

AUTHOR CONTRIBUTIONS

JW, WW, and Z-GH response for study design. JW carried out the research. JW and WW wrote part of the manuscript. JW, SR, and WS analyzed the results and prepared the figures and tables. All the authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (Grants U1913601 and 91848110), National Key R&D Program of China (Grant 2018YFB1307804), Beijing Natural Science Foundation (Grant 4202074), and the Strategic Priority Research Program of Chinese Academy of Science (Grant XDB32000000).

REFERENCES

Agarwal, P., and Deshpande, A. D. (2019). A framework for adaptation of training task, assistance and feedback for optimizing motor (re)-learning with a robotic exoskeleton. *IEEE Robot. Autom. Lett.* 4, 808–815. doi: 10.1109/LRA.2019.2891431

Akimoto, Y., Nagata, Y., Ono, I., and Kobayashi, S. (2012). Theoretical foundation for CMA-ES from information geometric perspective. *Algorithmica* 64, 698–716. doi: 10.1007/s00453-011-9564-8

Arns, M., Conners, C. K., and Kraemer, H. C. (2012). A decade of EEG theta/beta ratio research in ADHD: a meta-analysis. *J. Atten. Disord.* 17, 374–383. doi: 10.1177/1087054712460087

Babu, P. A., and Prasad, K. (2011). "Removal of ocular artifacts from EEG signals using adaptive threshold PCA and wavelet transforms," in *2011 International Conference on Communication Systems and Network Technologies* (Xi'an: IEEE), 572–575. doi: 10.1109/CSNT.2011.122

Barkley, R. A., Grodzinsky, G., and DuPaul, G. J. (1992). Frontal lobe functions in attention deficit disorder with and without hyperactivity: a review and research report. *J. Abnorm. Child Psychol.* 20, 163–188. doi: 10.1007/BF00916547

Choi, Y., Gordon, J., Park, H., and Schweighofer, N. (2011). Feasibility of the adaptive and automatic presentation of tasks (ADAPT) system for rehabilitation of upper extremity function post-stroke. *J. Neuroeng. Rehabil.* 8:42. doi: 10.1186/1743-0003-8-42

Colombo, R., Pisano, F., Mazzone, A., Delconte, C., Micera, S., Carrozza, M. C., et al. (2007). Design strategies to improve patient motivation during robot-aided rehabilitation. *J. Neuroeng. Rehabil.* 4:3. doi: 10.1186/1743-0003-4-3

Csikszentmihalyi, M., and Csikzentmihaly, M. (1990). *Flow: The Psychology of Optimal Experience*, Vol. 1990. New York, NY: Harper & Row New York.

Dunn, D. S., and Dougherty, S. B. (2005). Prospects for a positive psychology of rehabilitation. *Rehabil. Psychol.* 50:305. doi: 10.1037/0090-5550.50.3.305

Frølich, L., Winkler, I., Müller, K.-R., and Samek, W. (2015). "Investigating effects of different artefact types on motor imagery BCI," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Milan: IEEE), 1942–1945. doi: 10.1109/EMBC.2015.7318764

Gordon, K. E., and Ferris, D. P. (2007). Learning to walk with a robotic ankle exoskeleton. *J. Biomech.* 40, 2636–2644. doi: 10.1016/j.jbiomech.2006.12.006

Grant, J. S., Glandon, G. L., Elliott, T. R., Giger, J. N., and Weaver, M. (2004). Caregiving problems and feelings experienced by family caregivers of stroke survivors the first month after discharge. *Int. J. Rehabil. Res.* 27, 105–111. doi: 10.1097/01.mrr.0000127639.47494.e3

Gürkök, H., Hakvoort, G., and Poel, M. (2011). "Evaluating user experience in a selection based brain-computer interface game a comparative study," in *International Conference on Entertainment Computing* (Vancouver, BC: Springer), 77–88. doi: 10.1007/978-3-642-24500-8_9

Hansen, N. (2006). "The CMA evolution strategy: a comparing review," in *Towards a New Evolutionary Computation*, eds J. A. Lozano, P. Larrañaga, I. Inza, and E. Bengoetxea (Berlin; Heidelberg: Springer), 75–102. doi: 10.1007/3-540-32494-1_4

Harmony, T., Fernández, T., Silva, J., Bernal, J., Díaz-Comas, L., Reyes, A., et al. (1996). EEG delta activity: an indicator of attention to internal processing during performance of mental tasks. *Int. J. Psychophysiol.* 24, 161–171. doi: 10.1016/S0167-8760(96)00053-0

Holden, M. K. (2005). Virtual environments for motor rehabilitation. *Cyberpsychol. Behav.* 8, 187–211. doi: 10.1089/cpb.2005.8.187

Jackson, R. W., and Collins, S. H. (2015). An experimental comparison of the relative benefits of work and torque assistance in ankle exoskeletons. *J. Appl. Physiol.* 119, 541–557. doi: 10.1152/japplphysiol.01133.2014

Kline, J. E., Huang, H. J., Snyder, K. L., and Ferris, D. P. (2015). Isolating gait-related movement artifacts in electroencephalography during human walking. *J. Neural Eng.* 12:046022. doi: 10.1088/1741-2560/12/4/046022

Krebs, H. I., Palazzolo, J. J., Dipietro, L., Ferraro, M., Krol, J., Rannekleiv, K., et al. (2003). Rehabilitation robotics: performance-based progressive robot-assisted therapy. *Auton. Robots* 15, 7–20. doi: 10.1023/A:1024494031121

Kropotov, J. D. (2009). "Theta beta ratio as inattention index," in *Quantitative EEG, Event-Related Potentials and Neurotherapy*, ed J. Kropotov (Cambridge, MA: Academic Press), 399–400.

Lenze, E. J., Munin, M. C., Quear, T., Dew, M. A., Rogers, J. C., Begley, A. E., et al. (2004). The Pittsburgh rehabilitation participation scale: reliability and validity of a clinician-rated measure of participation in acute rehabilitation. *Archiv. Phys. Med. Rehabil.* 85, 380–384. doi: 10.1016/j.apmr.2003.06.001

Li, C., Rusak, Z., Horvath, I., Kooijman, A., and Ji, L. (2016). Implementation and validation of engagement monitoring in an engagement enhancing rehabilitation system. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 726–738. doi: 10.1109/TNSRE.2016.2591183

Loo, S. K., and Makeig, S. (2012). Clinical utility of EEG in attention-deficit/hyperactivity disorder: a research update. *Neurotherapeutics* 9, 569–587. doi: 10.1007/s13311-012-0131-z

Luo, L., Peng, L., Wang, C., and Hou, Z.-G. (2019). A greedy assist-as-needed controller for upper limb rehabilitation. *IEEE Trans. Neural Netw. Learn. Syst.* 30, 3433–3443. doi: 10.1109/TNNLS.2019.2892157

Maki, A., Sakamoto, N., Akimoto, Y., Nishikawa, H., and Umeda, N. (2020). Application of optimal control theory based on the evolution strategy (CMA-ES) to automatic berthing. *J. Mar. Sci. Technol.* 25, 221–233. doi: 10.1007/s00773-019-00642-3

Mann, C. A., Lubar, J. F., Zimmerman, A. W., Miller, C. A., and Muenchen, R. A. (1992). Quantitative analysis of EEG in boys with attention-deficit-hyperactivity disorder: controlled study with clinical implications. *Pediatr. Neurol.* 8, 30–36. doi: 10.1016/0887-8994(92)90049-5

Marshall, D., Coyle, D., Wilson, S., and Callaghan, M. (2013). Games, gameplay, and BCI: the state of the art. *IEEE Trans. Comput. Intell. AI Games* 5, 82–99. doi: 10.1109/TCIAIG.2013.2263555

Metzger, J.-C., Lambercy, O., Califfi, A., Dinacci, D., Petrillo, C., Rossi, P., et al. (2014). Assessment-driven selection and adaptation of exercise difficulty in robot-assisted therapy: a pilot study with a hand rehabilitation robot. *J. Neuroeng. Rehabil.* 11:154. doi: 10.1186/1743-0003-11-154

Muthukumaraswamy, S. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Front. Hum. Neurosci.* 7:138. doi: 10.3389/fnhum.2013.00138

Novak, D., Mihelj, M., Ziherl, J., Olensek, A., and Munih, M. (2011). Psychophysiological measurements in a biocooperative feedback loop for upper extremity rehabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 19, 400–410. doi: 10.1109/TNSRE.2011.2160357

Paolucci, S., Di Vita, A., Massicci, R., Traballesi, M., Bureca, I., Matano, A., et al. (2012). Impact of participation on rehabilitation results: a multivariate study. *Eur. J. Phys. Rehabil. Med.* 48, 455–466.

Park, W., Kwon, G. H., Kim, D.-H., Kim, Y.-H., Kim, S.-P., and Kim, L. (2014). Assessment of cognitive engagement in stroke patients from single-trial EEG during motor rehabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 351–362. doi: 10.1109/TNSRE.2014.2356472

Pirovano, M., Mainetti, R., Baud-Bovy, G., Lanzi, P. L., and Borghese, N. A. (2014). Intelligent game engine for rehabilitation (IGER). *IEEE Trans. Comput. Intell. AI Games* 8, 43–55. doi: 10.1109/TCIAIG.2014.2368392

Quesada, R. E., Caputo, J. M., and Collins, S. H. (2016). Increasing ankle push-off work with a powered prosthesis does not necessarily reduce metabolic rate for transtibial amputees. *J. Biomech.* 49, 3452–3459. doi: 10.1016/j.jbiomech.2016.09.015

Selinger, J. C., O'Connor, S. M., Wong, J. D., and Donelan, J. M. (2015). Humans can continuously optimize energetic cost during walking. *Curr. Biol.* 25, 2452–2456. doi: 10.1016/j.cub.2015.08.016

Tacchino, G., Gandolla, M., Coelli, S., Barbieri, R., Pedrocchi, A., and Bianchi, A. M. (2016). EEG analysis during active and assisted repetitive movements: evidence for differences in neural engagement. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 761–771. doi: 10.1109/TNSRE.2016.2597157

Tupper, A., and Henley, S. (1987). Predictive factors in stroke outcome and implications for intervention. *Int. J. Rehabil. Res.* 10, 119–121. doi: 10.1097/00004356-198712005-00024

Wang, J., Wang, W., Hou, Z.-G., Liang, X., Ren, S., and Peng, L. (2018). "Brain functional connectivity analysis and crucial channel selection using channel-wise CNN," in *International Conference on Neural Information Processing* (Siem Reap: Springer), 40–49. doi: 10.1007/978-3-030-04212-7_4

Wang, J., Wang, W., Hou, Z.-G., Shi, W., Liang, X., Ren, S., et al. (2019). "BCI and multimodal feedback based attention regulation for lower limb rehabilitation," in *2019 International Joint Conference on Neural Networks (IJCNN)* (Budapest: IEEE), 1–7. doi: 10.1109/IJCNN.2019.8851945

Xu, G., Gao, X., Chen, S., Wang, Q., Zhu, B., and Li, J. (2017). A novel approach for robot-assisted upper-limb rehabilitation: progressive resistance training as a paradigm. *Int. J. Adv. Robot. Syst.* 14:1729881417736670. doi: 10.1177/1729881417736670

Xu, G., Gao, X., Pan, L., Chen, S., Wang, Q., Zhu, B., et al. (2018). Anxiety detection and training task adaptation in robot-assisted active stroke rehabilitation. *Int. J. Adv. Robot. Syst.* 15, 1–18. doi: 10.1177/1729881418806433

Yannakakis, G. N., and Hallam, J. (2009). Real-time game adaptation for optimizing player satisfaction. *IEEE Trans. Comput. Intell. AI Games* 1, 121–133. doi: 10.1109/TCIAIG.2009.2024533

Zelik, K. E., Collins, S. H., Adamczyk, P. G., Segal, A. D., Klute, G. K., Morgenroth, D. C., et al. (2011). Systematic variation of prosthetic foot spring affects center-of-mass mechanics and metabolic cost during walking. *IEEE Trans. Neural Syst. Rehabil. Eng.* 19, 411–419. doi: 10.1109/TNSRE.2011.2159018

Zhang, J., Fiers, P., Witte, K. A., Jackson, R. W., Poggensee, K. L., Atkeson, C. G., et al. (2017). Human-in-the-loop optimization of exoskeleton assistance during walking. *Science* 356, 1280–1284. doi: 10.1126/science.aal5054

Zimmerli, L., Jacky, M., Lünenburger, L., Riener, R., and Bolliger, M. (2013). Increasing patient engagement during virtual reality-based motor rehabilitation. *Archiv. Phys. Med. Rehabil.* 94, 1737–1746. doi: 10.1016/j.apmr.2013.01.029

Zink, R., Hunyadi, B., Van Huffel, S., and De Vos, M. (2016). Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks. *J. Neural Eng.* 13:046017. doi: 10.1088/1741-2560/13/4/046017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 11 January 2021
doi: 10.3389/fnbot.2020.617531

Multi-Feature Input Deep Forest for EEG-Based Emotion Recognition

Yinfeng Fang[1], **Haiyang Yang**[1], **Xuguang Zhang**[1]*, **Han Liu**[2] and **Bo Tao**[3]

[1] School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China, [2] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, [3] State Key Lab of Digital Manufacturing Equipment & Technology, School of Mechanical Science and Engineering, Huazhong University of Science & Technology, Wuhan, China

Due to the rapid development of human–computer interaction, affective computing has attracted more and more attention in recent years. In emotion recognition, Electroencephalogram (EEG) signals are easier to be recorded than other physiological experiments and are not easily camouflaged. Because of the high dimensional nature of EEG data and the diversity of human emotions, it is difficult to extract effective EEG features and recognize the emotion patterns. This paper proposes a multi-feature deep forest (MFDF) model to identify human emotions. The EEG signals are firstly divided into several EEG frequency bands and then extract the power spectral density (PSD) and differential entropy (DE) from each frequency band and the original signal as features. A five-class emotion model is used to mark five emotions, including neutral, angry, sad, happy, and pleasant. With either original features or dimension reduced features as input, the deep forest is constructed to classify the five emotions. These experiments are conducted on a public dataset for emotion analysis using physiological signals (DEAP). The experimental results are compared with traditional classifiers, including K Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM). The MFDF achieves the average recognition accuracy of 71.05%, which is 3.40%, 8.54%, and 19.53% higher than RF, KNN, and SVM, respectively. Besides, the accuracies with the input of features after dimension reduction and raw EEG signal are only 51.30 and 26.71%, respectively. The result of this study shows that the method can effectively contribute to EEG-based emotion classification tasks.

Keywords: electroencephalogram (EEG), machine learning, feature exaction and selection, deep forest, emotion feelings-as-information

Edited by:
Dingguo Zhang,
University of Bath, United Kingdom

Reviewed by:
Uchenna Ogenyi,
University of Portsmouth,
United Kingdom
Gongfa Li,
Wuhan University of Science and
Technology, China
Gaoxiang Ouyang,
Beijing Normal University, China

***Correspondence:**
Xuguang Zhang
zhangxg@hdu.edu.cn

Received: *15 October 2020*
Accepted: *27 November 2020*
Published: *11 January 2021*

Citation:
Fang Y, Yang H, Zhang X, Liu H and Tao B (2021) Multi-Feature Input Deep Forest for EEG-Based Emotion Recognition.
Front. Neurorobot. 14:617531.
doi: 10.3389/fnbot.2020.617531

1. INTRODUCTION

Emotions occupy a very important position in human communication and personal decision-making. Although the existence of emotion is well-known, human knows very little about the mechanism behind it. Traditionally, human–computer interaction (HCI) for emotion recognition is carried out by using voice and facial expression signals (Fan et al., 2003; Sidney et al., 2005; Zeng et al., 2008). But these external signals have a certain degree of camouflage. Using voice and facial expression signals as the basis for emotion recognition is therefore not convincing. EEG physiological signals are directly produced by the central nervous system of human body, and the central nervous system is closely related to human emotions. Zheng et al. (2017) proved that the neural characteristics and stable EEG patterns are related to positive, neutral, and negative emotions. It indirectly proves that the use of EEG signals for emotion recognition is reliable.

Emotion recognition becomes a hot research topic regarding the development of basic emotional research theories and applications of emotional brain–computer interactions (aBCIs) (Nijboer et al., 2009; Garcia-Molina et al., 2013), such as emotion recognition with human brain-activity sensors is used for the treatment of patients with mental disorders (Mehmood et al., 2017).

In the field of EEG-based emotion recognition, traditional classifiers, such as K-Nearest Neighbor (KNN), Random Forest (RF), Gaussian Naive Bayes (GNB), Linear Discriminant Analysis (LDA), and Support Vector Machine (SVM), have been widely used. Heraz and Frasson (2007) used KNN to classify the intensity of each emotion (i.e., pleasure, arousal, and dominance) into two classes (high or low). With the database collected by themselves, the experimental accuracy reaches 73.5, 74.6, and 74%, respectively. He et al. (2017) proposed a feature extraction method based on multiple empirical mode decomposition (MEMD). The emotional state is identified as high/low arousal and high/low valence for the recorded eight-channel EEG signals on the DEAP database. The accuracy of SVM for arousal and valence is 67.9 and 70.9%, respectively. Veeramallu et al. (2019) utilized empirical mode decomposition (EMD) to classify automatic emotion classification based on EEG. This method uses the random forest classifier to classify positive, neutral, and negative emotions on the SJTU emotional EEG database (SEED) and obtains the highest recognition accuracy of 89.59, 91.45, and 93.87%, respectively. With the development of neural networks, deep learning based on neural networks and convolutional neural networks (CNN) have widely used in emotion recognition. Yang et al. (2018b) design a novel emotion recognition system which combines recurrence quantification analysis (RQA) with channel-frequency convolutional neural network (CFCNN). With the database collected by themselves, they classify the three specific emotions: happiness, sadness, and fear. The average recognition accuracy is 92.24%. Mehmood and Lee (2015) use KNN and SVM to classify four emotions: scared, sad, happy, and calm. With the database tested by themselves, the accuracy of emotion in valence and arousal dimensions is 32 and 37%, respectively by SVM, and the highest accuracy of KNN is 61%. Zheng et al. (2017) used discriminative Graph regularized Extreme Learning Machine (GELM) to perform LALV, HALV, LAHV, and HAHV (low arousal/low valence, high arousal/low valence, low arousal/high valence, and high arousal/high valence), four-classes classification experiments on valence-arousal (VA) space on the DEAP database, and achieve average recognition accuracy of 69.67%. The above results show that the accuracy of valence and arousal (two classes) have achieved good results. However, for the classification of 3 or 4 emotions, the recognition accuracy is generally not acceptable and needs to be improved. Therefore, how to improve the recognition accuracy of emotion classification for more classes and how to recognize the relationship between the generation of emotion and the corresponding physiological mechanism are the problems that need to be solved urgently in the field of BCI. This study aims to apply the deep forest technology to recognize human emotion from EEG signals, and improve the accuracy.

The layout of the paper is as follows. Section 2 introduces the development of multi-Grained Cascade Forest (gcForest) and its application in image classification and emotion recognition. In section 3, the DEAP dataset, the preprocess of EEG emotion recognition, the method of extracting features, the realization of the MFDF method, and the confirmation of hyper-parameters of deep forest are introduced. Section 4 demonstrates the experiment and experimental results, and section 5 analyzes and discusses the experimental results. Section 6 gives some conclusions and future work.

2. RELATED WORK

GcForest is a highly competitive decision tree integration method for deep neural networks (Zhou and Feng, 2017). The gcForest consists of multi-grained scanning and cascade Forest. It employs a cascade structure to realize layer-by-layer processing. Through multi granularity scanning, they increase the diversity of features to enhance the cascade forest. In addition, the gcForest is a deep model based on decision trees, and the training process does not rely on back-propagation and gradient adjustment. Compared with deep neural networks, gcForest has fewer hyper-parameters and achieves excellent performance across various domains by using even the same parameter setting. In the study of Cao et al. (2019b), the rotation-based deep forest (RBDF) is proposed for the classification of hyper-spectral images (HSIs). Experimental results based on three HSIs demonstrate that the proposed method achieves the state-of-the-art classification performance. Cao et al. (2019a) propose a new deep model–densely connected deep random forest (DCDRF) to classify the HSIs. Experimental results prove that the proposed method can achieve a better classification performance than the conventional deep-learning-based methods. A deep multigrained cascade forest (dgcForest) was proposed by Liu et al. (2019). Experimental results testify that their proposed algorithm presents a good performance on the hyper-spectral image (HSI). Zhou et al. (2019) proposed a deep-forest-based method for hashing learning. The experimental results show that the proposed method has better performance with shorter binary codes than other corresponding hashing methods. In conclusion, in image detection, voice detection, and other fields, the gcForest has been applied and achieved excellent results.

Each cascade layer of gcForest is composed of random forest. Random forest is an algorithm that integrates multiple decision trees based on the idea of ensemble learning. Its basic unit is the decision tree, and these decision trees are independent of each other and have no relationship (Ho, 1995; Breiman, 2001). The integrated learning feature of random forest enables it to obtain better results even if each tree does not have high-precision decision-making. Random forest is utilized by Memar and Faradji (2017) to classify the sleep stage based on EEG signals, which is one of the most critical steps in effective diagnosis and treatment of sleep-related disorders. Random forest consists of decision trees. Decision tree is a shortcut mode of attribute classification (Janikow, 1998). Additionally, decision tree is a kind of white box method, and it is more convincing than other

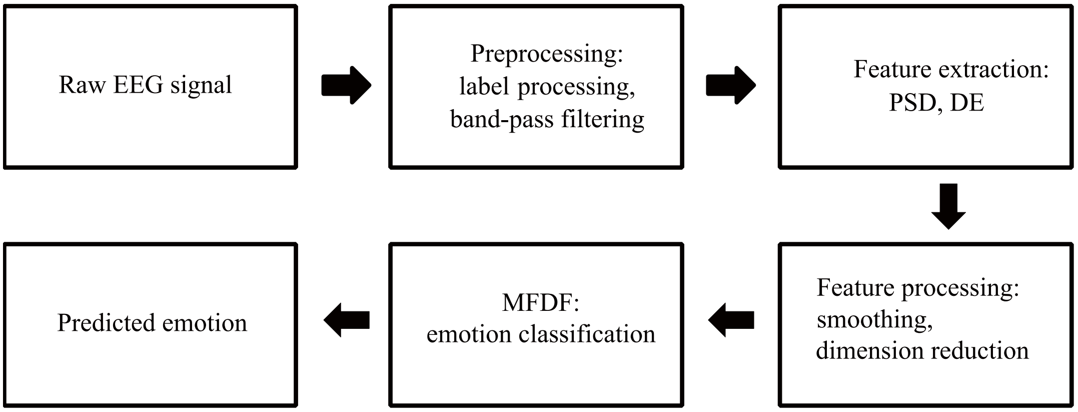


FIGURE 1 | The flow chart of EEG-based emotion recognition.

classifiers. In dealing with the undefined problem of emotion recognition, the results obtained by the decision tree help us understand the physiological mechanism behind the data that generates emotion. Since decision tree is proposed, it has been widely used in the treatment of diseases with EEG signals. According to previous studies, the induction of decision trees from data has been applied in various medical domains, and Hetmerova et al. (1997) argued that it is interesting to use decision tree to extract useful rules for disease judgments. For example, Rajaguru and Prabhakar (2017) proposed a soft decision tree classifier in the EEG seizure classification, and Sukanesh and Harikumar (2008) proposed hierarchical aggregation functions decision trees to classify epilepsy risk classification based on EEG signals. Based on the above studies, it is feasible to use gcForest based on decision tree for emotion recognition.

Multi-grained scanning is usually used to process the original data. For example, Cheng et al. (2020) propose a method for multi-channel EEG-based emotion recognition using deep forest. On the DEAP database, the average accuracy reaches 97.69 and 97.53% for valence and arousal, respectively. Yao et al. (2019) used deep forest with multi-scale window (MSWDF) to identify EEG emotions, and the average recognition accuracy in the classification of pleasure, relaxation, sadness (three classes) is 84.90%. These studies directly use multi-grained scanning on the original data to conduct experiments instead of using the feature extraction method. This paper proposes an emotion recognition algorithm model, the multi-feature deep forest (MFDF), on the basis of gcForest. The algorithm extracts effective features from the original data and inputs the features into the deep forest for emotion classification and recognition. The experimental results show that the average accuracy of MFDF reaches 71.05%, and the highest accuracy can reach 87.10% on the DEAP database. The experimental results prove our model is valid.

3. METHOD

The MFDF algorithm is shown in **Figure 1**. Firstly, the original data is preprocessed, including emotion label processing and frequency band division. Secondly, affective computing is performed on the data to obtain two types of features: PSD and DE. Thirdly, data smoothing and dimensionality reduction are performed on the features. Finally, the original data is converted into feature vectors, and the feature vectors are input to the deep forest for emotion recognition and classification.

TABLE 1 | Data organization for one subject.

Identification	Size	Content
Data	40× 40 × 8,064	Video/trial × channel × data
Label	40 × 4	Video/trial × label (valence, arousal, dominance, liking)

3.1. Introduction to DEAP Dataset

The public dataset DEAP (Koelstra et al., 2011) is utilized to validate our proposed approach in this study. The dataset contains 32 channel EEG signals and eight-channel peripheral physiological signals recorded from 32 subjects watching 40 music videos. Each video is rated to the levels of arousal, valence, liking, and dominance of each subject. The rating scores are closely related to emotions. In the current study, only the EEG signals are used for emotional recognition. The EEG signals are sampled at the frequency of 128Hz, and then are preprocessed by a bandpass filter with a bandwidth ranging from 4.0 to 45.0 Hz. The recorded EEG data contains 60 s video-induced EEG data and 3 s baseline data without watching video. The EEG data from each subject are organized as shown in **Table 1**.

3.2. Data Pre-processing

For the study of EEG signals, five frequency bands are normally separated with different frequency ranges: delta (1–4 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (31–47 Hz). Except for the delta signal, which is generally generated during deep sleep, the theta, alpha, beta, and gamma signals are closely related to emotions (Aftanas et al., 2001; Balconi and

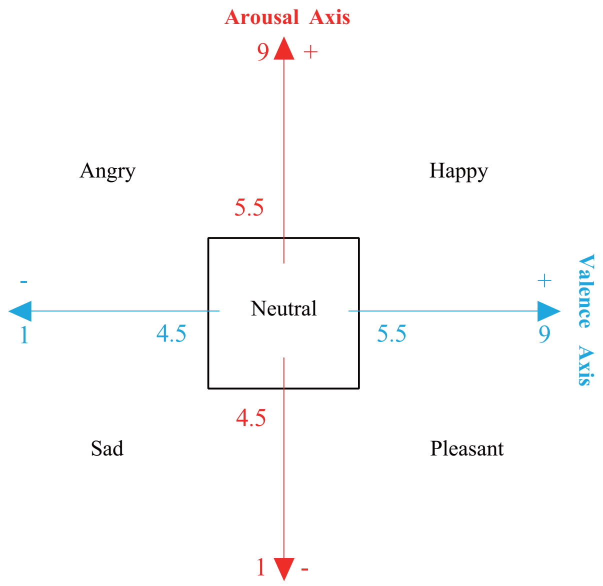


FIGURE 2 | The emotion model for relabeling five emotions, including neutral, angry, happy, sad, and pleasant.

Lucchiari, 2008; Balconi and Mazza, 2009); the theta, alpha, beta, and gamma signals are thus extracted in the current study.

This study marks angry, happy, sad, pleasant, and neutral according to the following rules.

Angry: Valence < 4.5 and Arousal > 5.5
Happy: Valence > 5.5 and Arousal > 5.5
Sad: Valence < 4.5 and Arousal < 4.5
Pleasant: Valence > 5.5 and Arousal < 4.5
Neutral: $4.5 \leq$ Valence ≤ 5.5 and $4.5 \leq$ Arousal ≤ 5.5

Related studies tend to classify the level of arousal and valence as two-class or three-class classification problems. Few studies provide emotional labels according to the scores of valence and arousal. Lan et al. (2016) mark the EEG signals with the labels of pleasant, happy, frightened, and angry. Zheng et al. (2017) proposed an emotion representation model based on the valence-arousal (VA) level and mark four quadrants of the VA space with four types of emotions. On the basis of these studies, the current study labels the data with five emotional classes, with an additional emotion type neutral, as shown in **Figure 2**. The data of subject S3, S12, S13, S14, S23, S26, S27, S30, and S31 are not used in this study for evaluation because some labels are absent after remarking by the proposed five-class emotion model.

3.3. Feature Extraction

Several features are extracted from the EEG signals for analysis, including power spectral density (PSD) and differential entropy (DE). Naderi and Mahdavi-Nasab (2010) found that the power spectral density (PSD) estimation of the Welch method can provide very strong features, and it is also a good representation of the EEG signal. DE has proven to be the most accurate and stable EEG feature that reflects the change of vigilance (Duan et al., 2013; Shi et al., 2013).

The PSD can be defined as follows,

$$P_i(f) = \frac{1}{L}|\sum_{n=1}^{L-1} \mathrm{x_i}[\mathrm{n}]\mathrm{e}^{\mathrm{j}2\pi\,\mathrm{fn}}|^2, \tag{1}$$

where L is the length of the signal, and $P_i(f)$ is the fast Fourier transform of the signal $\mathrm{x_i}[\mathrm{n}]$. The PSD feature $[P(\mathrm{x_i})]$ of the signal $\mathrm{x_i}[\mathrm{n}]$ can then be obtained by Equation (2).

$$P(\mathrm{x_i}) = \frac{1}{\mathrm{K}}(\sum_{\mathrm{i}=1}^{\mathrm{K}} \mathrm{P_i}(\mathrm{f})), \tag{2}$$

where, K is number of frequency points used to calculate the discrete Fourier transforms.

Feature DE can be defined as follows,

$$\begin{aligned} h(\mathrm{x_i}) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{\frac{-(\mathrm{x}-\mu)^2}{2\sigma^2}} \log \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{\frac{-(\mathrm{x}-\mu)^2}{2\sigma^2}} \mathrm{dx} \\ &= \frac{1}{2} log(2\pi e\sigma^2), \end{aligned} \tag{3}$$

where, $\mathrm{x_i}[\mathrm{n}]$ is assumed to satisfy the Gaussian distribution of $\mathrm{N}(\mu, \sigma^2)$ (Shi et al., 2013).

Sliding window technology is applied for feature extraction in this study. The Hanning window with a length of 1 s and an increment of 1 s increment without overlap is taken to segment the EEG signal for feature extraction.

In order to remove the noise which has nothing to do with the emotional states, this paper uses Savitzky-Golay method with span of 5 and degree of 3 to smooth the data. Dimensionality reduction could reduce the computational burden and increase the stability of the computation (Duan et al., 2013). Moreover, it is a practical solution to avoid "dimension disaster" (Duan et al., 2013). Popular dimensionality reduction methods include principal component analysis (PCA), minimum-redundancy-maximum-correlation (MRMR), and so on (Peng et al., 2005). Although PCA can reduce the feature dimensions, it cannot preserve the original domain information, such as channel and frequency after the transformation. Hence, this paper chooses the MRMR algorithm to select a feature subset from an initial feature set (Zheng et al., 2017). The MRMR algorithm utilizes mutual information as the relevance measure with the max-dependency criterion and minimal redundancy criterion. The max-relevance criterion searches for features satisfying with the mean value of all the mutual information values between the individual feature $\mathrm{x_i}$ and class c as follows,

$$\max D\,(\mathrm{S}, \mathrm{c})\,, \mathrm{D} = \frac{1}{|\mathrm{S}|} \sum_{\mathrm{x_i} \in \mathrm{S}} \mathrm{I}\,(\mathrm{x_i}; \mathrm{c}) \tag{4}$$

When two features are highly dependent on the same class, if one of the features is removed, the overall class distinction ability does not change much. The following minimal redundancy condition can thus be added to select for mutually exclusive features,

$$\min\,(\mathrm{S})\,, \mathrm{R} = \frac{1}{|\mathrm{S}|^2} \sum_{\mathrm{x_i}, \mathrm{x_j} \in \mathrm{S}} \mathrm{I}\left(\mathrm{x_i}; \mathrm{x_j}\right) \tag{5}$$

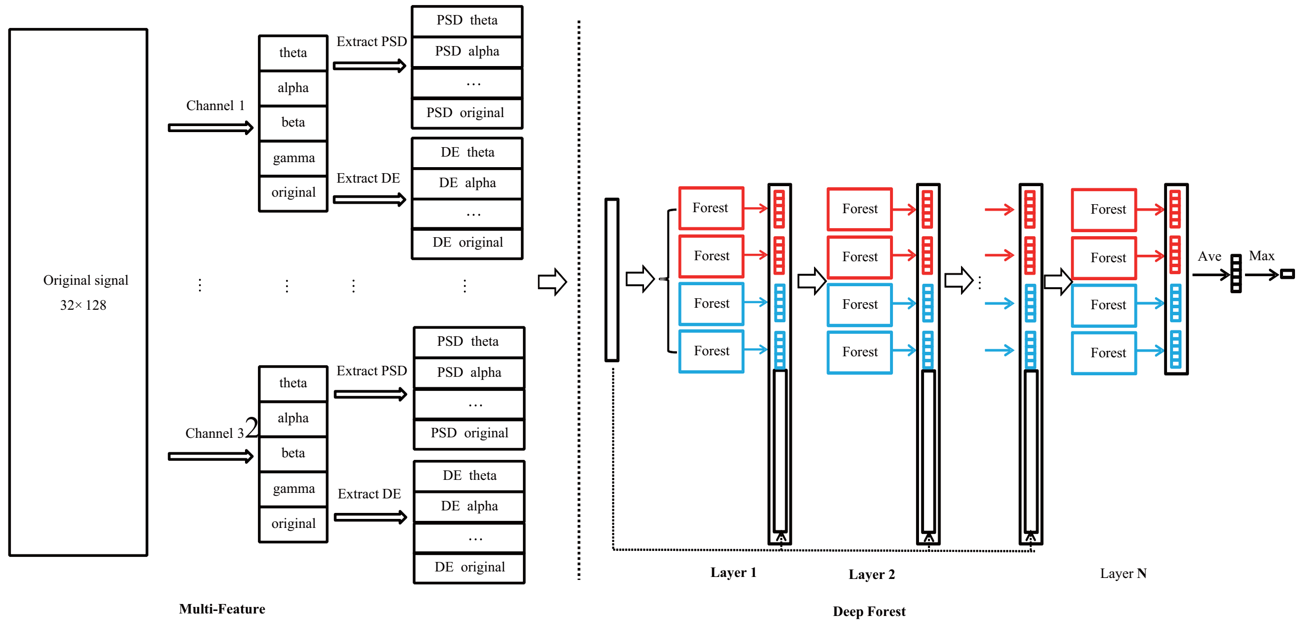


FIGURE 3 | The structure of multi-feature deep forest. It includes a multi-feature extraction part (left panel) and a deep forest part (right panel). In multi-feature extraction stage, PSD and DE features are extracted from the EEG signal. In deep forest part, each layer of the cascade consists of two types of random forests with different colored blocks, and the number of layers is optimized during the training stage.

The above two constraints are termed as "minimum-redundancy-maximum-correlation" (MRMR). We define the operator $\varnothing(D, R)$ to combine D and R, and the simplest definition can be expressed as.

$$\max \varnothing(D, R), \varnothing = D - R \tag{6}$$

3.4. Multi-Feature Deep Forest Method

As an extension of random forest, deep forest is different from general random forest. Random forest is based on decision trees and uses the idea of ensemble learning to classify data. Deep forest adopts cascade structure which combines the characteristics of the neural network to further improve the recognition of random forest, and the cascade layer can automatically adjust the optimal number of classification layers (Xu et al., 2019). Deep forest automatically optimizes the structure of deep forest by comparing the classification performance of adjacent layers.

The structure of the proposed MFDF method is demonstrated in **Figure 3**. It includes two parts. One is multi-feature extraction, and the other is deep forest architecture. In the multi-feature extraction stage, PSD and DE features are extracted from different wave bands of EEG signals or original signals for each EEG channel. The size of the extracted feature is 320×1 (32 channels and 10 types of feature for each channel). The architecture of the deep forest can be found in the right panel of **Figure 3**. In each layer, four random forests are included. Two of random forest set the number of tree node's split feature by the number of square root for the total number of features, and rest two forests set it by the logarithm for the total number of features (Zhou and Feng, 2017; Yao et al., 2019; Cheng et al., 2020). A Iterative Dichotmizer 3 (ID3) decision tree is utilized in this study (Zhou and Feng, 2017; Yao et al., 2019; Cheng et al., 2020).

In the deep forest, the output vector of the current layer is taken as enhancement features that is further used to combine original data as the new input of the next layer, as seen in the right panel of **Figure 3**. The output vector is actually a vector of classification probability for each class. The classification probability is statistically calculated from the output of each decision tree. In the current study, five classes of emotions are investigated. Each layer would thus generate 20 enhancement features by four random forests. The number of layers is not fixed in this study, and it is determined by the algorithm via evaluate whether additional layer could improve the classification performance.

3.5. Confirmation of Hyper-Parameters of Deep Forest

It is well-known that the number of trees of a random forest would influence its classification performance, and a large number of trees would increase computing burden. To determine the number of trees in a forest, subject's (S5) data are used to evaluate how the number of trees influence the classification accuracy. **Figure 4** shows the accuracy change along the increase of the number of trees. It can be found that with the increase of tree number, the accuracy shows a clear improvement. When the number of trees increases to 200, accuracy enhancement becomes stable. In order to balance computing resources and

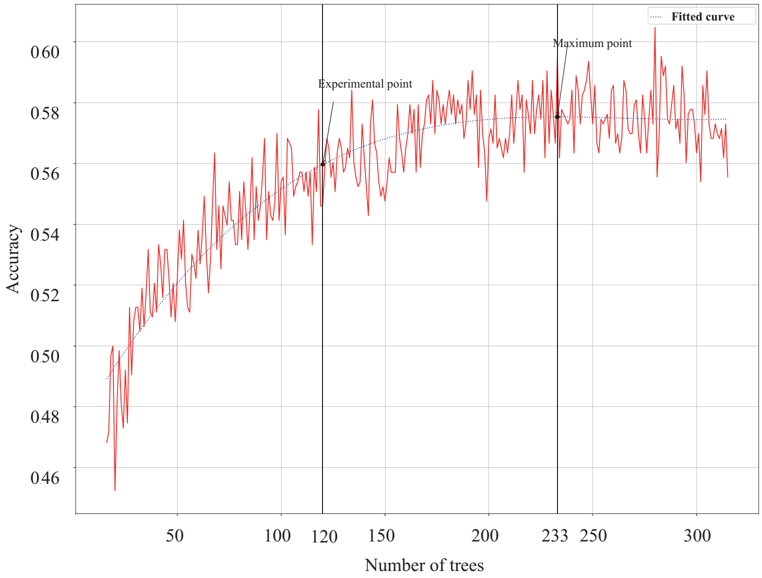


FIGURE 4 | The change of classification accuracy along with the number of trees in each forest, and the highest accuracy point and the 97% turning point are marked on the fitted curve.

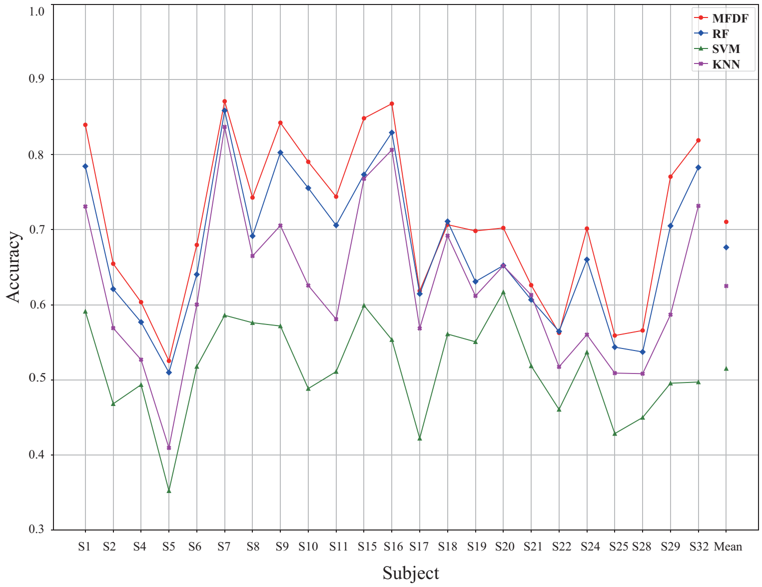


FIGURE 5 | The classification accuracy across different subjects under the input of PSD and DE feature without dimension reduction, in which the classifiers of MFDF, RF, SVM, and KNN are compared.

classification accuracy, this study sets the number of trees in each random forest to 120 trees, which is the 97% turning point of the highest accuracy. Other parameters of random forest are set as follows. The maximum depth of decision tree is set to 14. The minimum number of samples required to split middle nodes is set to five. This paper also uses out-of-bag samples to estimate the generalization accuracy. To tackle the problem of data imbalance, the samples in a small group would be more frequently chosen during the training stage.

4. EXPERIMENTAL RESULT

The proposed MFDF is compared with three traditional classifiers (i.e., RF, SVM, and KNN), and 5-fold cross-validation is applied to obtain the classification accuracy. For SVM classifier, linear kernel is applied with the penalty coefficient at 0.8. The K coefficient is set to 5 for KNN classifier. In the random forest, 100 decision trees are included, and ID3 algorithm is used for training. The average accuracy and the accuracy for different subjects are demonstrated in **Figure 5**. It can be found that the average accuracy is around 71, 68, 52, and 63% for MFDF, RF, SVM, and kNN, respectively. The average recognition accuracy of the proposed model is 19.53% higher than SVM, 3.4% higher than RF, and 8.54% higher than KNN in **Table 2**. Besides, the proposed MFDF attains the highest accuracy up to 86% for subject s7 and s16.

The reliability and effectiveness of the MFDF are also evaluated through investigating different types of data as input (i.e., raw data, features, and features after dimension reduction). The experimental results with dimension reduced feature as input are demonstrated in **Figure 6**. The average recognition accuracy of the proposed method is 51.30%. Although it is much lower than that without dimensionality reduction, this accuracy is still higher than that of the compared classifiers. **Table 3** shows that the average recognition rate of the model is 5.76% higher than SVM, 1.19% higher than RF, and 11.21% higher than KNN.

This study also takes raw data without feature extraction as the input for deep forest and investigates the accuracy. The results are demonstrated in **Figure 7**. The average accuracy without dimensionality reduction processing is 71.05%, and the average accuracy for dimensionality reduction processing is 51.30%. The average accuracy of the original data is 26.71%. It can be found that using original feature as input obtains the highest classification accuracy, which indicates that deep forest can deal with features rather than original data more successfully in the current case study.

In addition, for real-time application, the computing cost is very important. **Table 4** demonstrates the computational cost for feature extraction with or without dimension reduction, and the corresponding accuracy of subject S1. It can be found that feature extraction takes a large amount of time, which is much longer than the time cost of classification. The total time of MFDF, SVM, KNN, and RF for feature extraction and classification of a sample is 38, 25.56, 25.22, and 25.65 ms, respectively. Although MFDF takes more time than the other three traditional classifiers, it achieves the relative higher classification accuracy.

TABLE 2 | A comparison of classification accuracy by four classifiers with original feature as input.

	Highest recognition accuracy (%)	Lowest recognition accuracy (%)	Average recognition accuracy (%)
MFDF	87.90	52.54	71.05 ± 10.61
RF	85.87	51.00	67.65 ± 9.72
SVM	61.71	35.23	51.52 ± 6.41
KNN	83.69	40.96	62.51 ± 10.30

TABLE 3 | A comparison of classification accuracy by four classifiers with dimension reduced feature as input.

	Highest recognition accuracy (%)	Lowest recognition accuracy (%)	Average recognition accuracy (%)
MFDF	62.50	38.17	51.30 ± 7.02
RF	62.67	40.17	50.11 ± 6.73
SVM	59.33	34.67	45.54 ± 5.75
KNN	53.83	32.33	40.09 ± 5.54

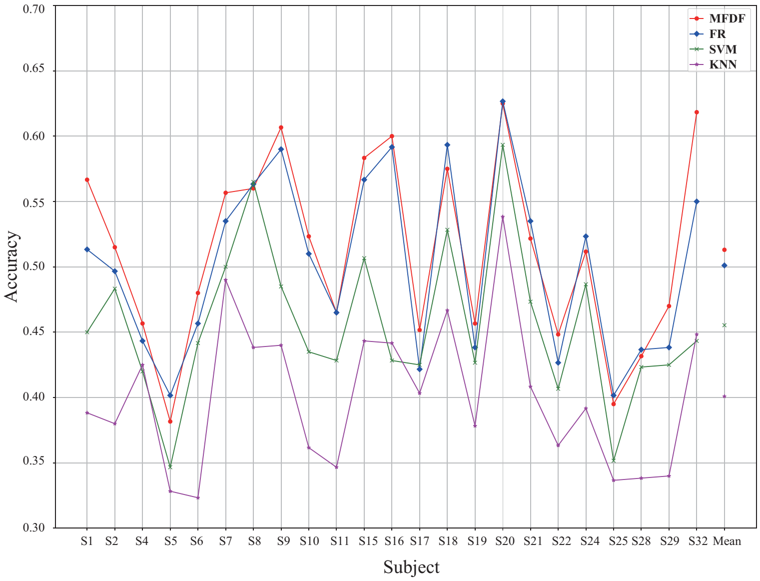


FIGURE 6 | The classification accuracy across different subjects under the input of PSD and DE feature with dimension reduction across different subjects, in which the classifiers of MFDF, RF, SVM, and KNN are compared.

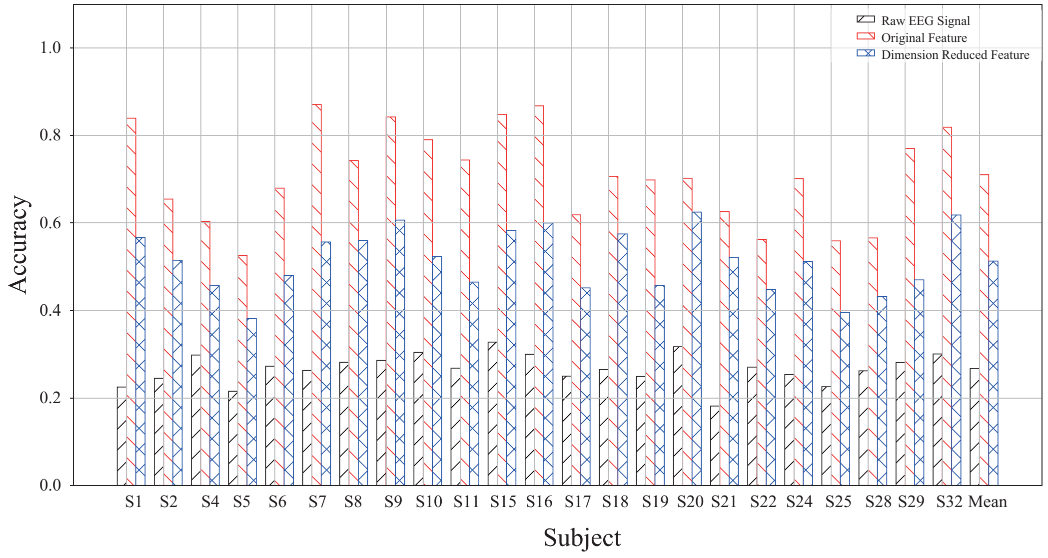


FIGURE 7 | A comparison of MFDF with different input of raw EEG signal, original features, and dimension reduced features.

TABLE 4 | The computational cost for feature extraction and classification per sample.

	Features extraction (ms)	Classification (ms)	Accuracy (%)
MFDF(1)	25	13	71.05
MFDF(2)	44	11	51.30
SVM	25	0.56	51.52
RF	25	0.65	67.65
KNN	25	0.22	62.51

The reported accuracy is from subject S1. MFDF(1) indicates the results without dimensionality reduction, and MFDF(2) is the ones after dimensionality reduction. The compared SVM, RF, and kNN classifier takes original feature as input.

TABLE 5 | The reported accuracy by the literatures with DEAP database.

Study	Results
Koelstra et al. (2011)	62.0, 57.6% for valence and arousal (two-classes) with all 32 participants.
Xie et al. (2018)	79.06 and 77.19% for valence and arousal (two-classes) with all 32 participants.
Yang et al. (2018a)	90.80 and 91.03% for valence and arousal (two-classes) with all 32 participants.
Chung and Yoon (2012)	66.6, 66.4% for valence and arousal (two-classes), 53.4, 51.0% for valence and arousal (three-classes) with all 32 participants.
Yao et al. (2019)	84.90% for pleasure, relax, sadness (three-classes)with all 32 participants
Liu and Sourina (2013)	63.04% for arousal-dominance recognition (four-classes) with the selected 10 participants.
Zheng et al. (2017)	69.67% for quadrants of VA space (four-classes) with all 32 participants.
Huang et al. (2019)	73.76% for relax, depression, excitement, fear (four-classes) with all 32 participants
Our method	71.05% for angry, happy, sad, pleasant, and neutral (five-classes) with the selected 23 participants

5. DISCUSSION

EEG-based emotion recognition is a hot-spot in recent years. Many researchers have proposed effective classification models to improve emotion recognition accuracy (Huang et al., 2019; Yao et al., 2019). The extraction of distinguish and consistent EEG features are critical for classification systems. PSD and DE are two classic features set for EEG-based emotion recognition (Naderi and Mahdavi-Nasab, 2010; Duan et al., 2013; Shi et al., 2013), which are selected as the input feature for the proposed MFDF method. This study also finds that the implementation of dimensionality reduction would negatively influence the classification accuracy, which is inconsistent with the results obtained by Zheng et al. (2017) who find that dimensionality reduction does not affect the performance of our model greatly. Additionally, it is also found that deep forest cannot achieve acceptable classification accuracy with raw EEG signal as input, although recent studies show a tendency to use original data as input for deep forest. For instance, Cheng et al. (2020) utilized the raw multi-channel EEG data as 2D input and achieved more than 97% classification for a two-class classification problem of the state of valence and arousal; it is not clear, however, whether the extraction of feature can further improve the classification accuracy. In sum, the result of the current study indicates that it is rational to use PSD and DE feature as input for deep forest.

Table 5 lists related studies that used the DEAP dataset for pattern recognition. The baseline accuracy for DEAP is only

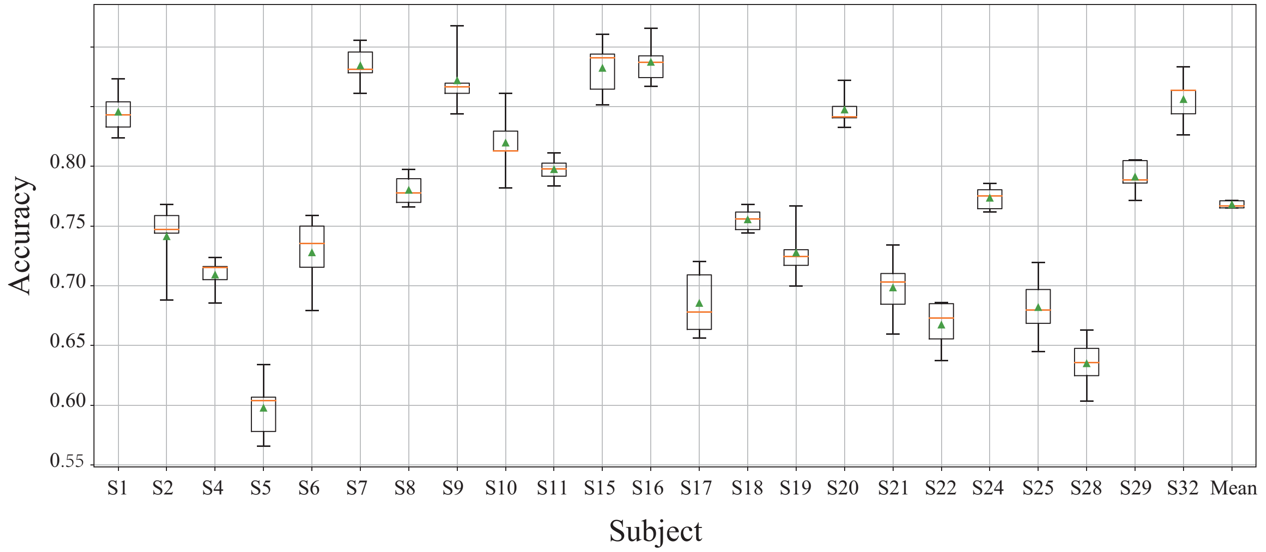


FIGURE 8 | The accuracy for the classification of 4 types of emotions by MFDF. The error bar is obtained by 5-fold cross validation.

about 60% for two-class classification problems, and this was published together with DEAP by Koelstra et al. (2011). In the two-class classification problem, Yang et al. (2018a) obtained the accuracy of 91% via the parallel convolutional recurrent neural network, which is the highest classification accuracy within these studies on two-class classification. For three-class emotion classification, Yao et al. (2019) used deep forest with multi-scale window (MSWDF) to identify the emotions of pleasure, relaxation and sadness and achieved an accuracy of 84.90%. The proposed method achieved an accuracy of 76.8% for four emotions (shown in **Figure 8**, which is higher than the ensemble convolutional neural network (ECNN) approach that obtained the accuracy of 73.76% (Huang et al., 2019). Besides, the current study takes five emotions for classification, including a neutral emotion, and achieved an average accuracy of 71.05%.

6. CONCLUSION

This paper proposes an emotion recognition method based on gcForest (MFDF) for emotion recognition, which takes human-crafted features (i.e., PSD and DE) as the input for deep forest. The proposed methods demonstrate a competitive performance (71% accuracy for five types of emotions) in the comparison with traditional classifiers. The results indicate that using feature as input can obtain much higher accuracy than using raw EEG signals. In the future, deep forest will be further optimized through the combination of raw EEG data together with EEG features as input for deep forest. The issue of cross-subject emotion recognition will be investigated to establish a more general classification model.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YF: original motivation and idea and language revising. HY: coding and drafted manuscript. XZ: project supervisor. HL: method of deep forest algorithm. BT: final check. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by basic public welfare research program of Zhejiang Province (no. LQ20F020016), the National Natural Science Foundation of China (no. 61771418), and public funds of the State Key Laboratory of digital manufacturing equipment and technology of China (no. DMETKF2020027).

REFERENCES

Aftanas, L., Varlamov, A., Pavlov, S., Makhnev, V., and Reva, N. (2001). Affective picture processing: event-related synchronization within individually defined human theta band is modulated by valence dimension. *Neurosci. Lett*. 303, 115–118. doi: 10.1016/S0304-3940(01)01703-7

Balconi, M., and Lucchiari, C. (2008). Consciousness and arousal effects on emotional face processing as revealed by brain oscillations. A gamma band analysis. *Int. J. Psychophysiol*. 67, 41–46. doi: 10.1016/j.ijpsycho.2007.10.002

Balconi, M., and Mazza, G. (2009). Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues: ERS/ERD and coherence measures of alpha band. *Int. J. Psychophysiol*. 74, 158–165. doi: 10.1016/j.ijpsycho.2009.08.006

Breiman, L. (2001). Random forests. *Mach. Learn*. 45, 5–32. doi: 10.1023/A:1010933404324

Cao, X., Li, R., Ge, Y., Wu, B., and Jiao, L. (2019a). Densely connected deep random forest for hyperspectral imagery classification. *Int. J. Rem. Sens*. 40, 3606–3622. doi: 10.1080/01431161.2018.1547932

Cao, X., Wen, L., Ge, Y., Zhao, J., and Jiao, L. (2019b). Rotation-based deep forest for hyperspectral imagery classification. *IEEE Geosci. Rem. Sens. Lett*. 16, 1105–1109. doi: 10.1109/LGRS.2019.2892117

Cheng, J., Chen, M., Li, C., Liu, Y., Song, R., Liu, A., et al. (2020). Emotion recognition from multi-channel eeg via deep forest. *IEEE J. Biomed. Health Inform*. doi: 10.1109/JBHI.2020.2995767

Chung, S. Y., and Yoon, H. J. (2012). "Affective classification using bayesian classifier and supervised learning," in *2012 12th International Conference on Control, Automation and Systems* (Jeju-do: IEEE), 1768–1771.

Duan, R.-N., Zhu, J.-Y., and Lu, B.-L. (2013). "Differential entropy feature for eeg-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)* (San Diego, CA: IEEE), 81–84. doi: 10.1109/NER.2013.6695876

Fan, C., Johnson, M., Messom, C., and Sarrafzadeh, A. (2003). "Machine vision for an intelligent tutor," in *Proceedings of the International Conference on Computational Intelligence, Robotics and Autonomous Systems* (Singapore: Citeseer).

Garcia-Molina, G., Tsoneva, T., and Nijholt, A. (2013). Emotional brain-computer interfaces. *Int. J. Auton. Adapt. Commun. Syst*. 6, 9–25. doi: 10.1504/IJAACS.2013.050687

He, Y., Ai, Q., and Chen, K. (2017). "A memd method of human emotion recognition based on valence-arousal model," in *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Vol. 2 (Hangzhou: IEEE), 399–402. doi: 10.1109/IHMSC.2017.201

Heraz, A., and Frasson, C. (2007). Predicting the three major dimensions of the learner's emotions from brainwaves. *Int. J. Comput. Sci*. 2, 187–193. doi: 10.5281/zenodo.1076762

Hetmerova, A., Kraus, J., and Stepankova, O. (1997). "Generation of decision trees from EEG data," in *Information Technology Applications in Biomedicine, Itab 97, IEEE Engineering in Medicine & Biology Society Region 8 International Conference* (Prague). doi: 10.1109/ITAB.1997.649413

Ho, T. K. (1995). "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, Vol. 1 (Montreal, QC: IEEE), 278–282.

Huang, H., Hu, Z., Wang, W., and Wu, M. (2019). Multimodal emotion recognition based on ensemble convolutional neural network. *IEEE Access* 8, 3265–3271. doi: 10.1109/ACCESS.2019.2962085

Janikow, C. Z. (1998). Fuzzy decision trees: issues and methods. *IEEE Trans. Syst. Man Cybern. B Cybern*. 28, 1–14. doi: 10.1109/3477.658573

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2011). Deap: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput*. 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Lan, Z., Sourina, O., Wang, L., and Liu, Y. (2016). Real-time EEG-based emotion monitoring using stable features. *Vis. Comput*. 32, 347–358. doi: 10.1007/s00371-015-1183-y

Liu, X., Wang, R., Cai, Z., Cai, Y., and Yin, X. (2019). Deep multigrained cascade forest for hyperspectral image classification. *IEEE Trans. Geosci. Rem. Sens*. 57, 8169–8183. doi: 10.1109/TGRS.2019.2918587

Liu, Y., and Sourina, O. (2013). "Real-time fractal-based valence level recognition from EEG," in *Transactions on Computational Science XVIII* (Berlin: Springer), 101–120. doi: 10.1007/978-3-642-38803-3_6

Mehmood, R. M., Du, R., and Lee, H. J. (2017). Optimal feature selection and deep learning ensembles method for emotion recognition from human brain eeg sensors. *IEEE Access* 5, 14797–14806. doi: 10.1109/ACCESS.2017.2724555

Mehmood, R. M., and Lee, H. J. (2015). "Emotion classification of eeg brain signal using SVM and KNN," in *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (Turin: IEEE), 1–5. doi: 10.1109/ICMEW.2015.7169786

Memar, P., and Faradji, F. (2017). A novel multi-class EEG-based sleep stage classification system. *IEEE Trans. Neural Syst. Rehabil. Eng*. 26, 84–95. doi: 10.1109/TNSRE.2017.2776149

Naderi, M. A., and Mahdavi-Nasab, H. (2010). "Analysis and classification of EEG signals using spectral analysis and recurrent neural networks," in *2010 17th Iranian Conference of Biomedical Engineering (ICBME)* (Isfahan: IEEE), 1–4. doi: 10.1109/ICBME.2010.5704931

Nijboer, F., Morin, F. O., Carmien, S. P., Koene, R. A., Leon, E., and Hoffmann, U. (2009). "Affective brain-computer interfaces: psychophysiological markers of emotion in healthy persons and in persons with amyotrophic lateral sclerosis," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* (Amsterdam: IEEE), 1–11. doi: 10.1109/ACII.2009.5349479

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell*. 27, 1226–1238. doi: 10.1109/TPAMI.2005.159

Rajaguru, H., and Prabhakar, S. K. (2017). "Sparse PCA and soft decision tree classifiers for epilepsy classification from EEG signals," in *2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA)*, Vol. 1 (Coimbatore: IEEE), 581–584. doi: 10.1109/ICECA.2017.8203604

Shi, L.-C., Jiao, Y.-Y., and Lu, B.-L. (2013). "Differential entropy feature for EEG-based vigilance estimation," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Osaka: IEEE), 6627–6630.

Sidney, K. D., Craig, S. D., Gholson, B., Franklin, S., Picard, R., and Graesser, A. C. (2005). "Integrating affect sensors in an intelligent tutoring system," in *Affective Interactions: The Computer in the Affective Loop Workshop* (Massachusetts), 7–13.

Sukanesh, R., and Harikumar, R. (2008). Fuzzy techniques and hierarchical aggregation functions decision trees for the classification of epilepsy risk levels from eeg signals. In TENCON 2008–2008 IEEE Region 10 Conference, 1–6. IEEE. doi: 10.1109/TENCON.2008.4766545

Veeramallu, G. K. P., Anupalli, Y., Kumar Jilumudi, S., and Bhattacharyya, A. (2019). "EEG based automatic emotion recognition using emd and random forest classifier," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (Kanpur: IEEE), 1–6. doi: 10.1109/ICCCNT45670.2019.8944903

Xie, O., Liu, Z.-T., and Ding, X.-W. (2018). "Electroencephalogram emotion recognition based on a stacking classification model," in *2018 37th Chinese Control Conference (CCC)* (Wuhan: IEEE), 5544–5548. doi: 10.23919/ChiCC.2018.8483496

Xu, S., Tang, Q., Jin, L., and Pan, Z. (2019). A cascade ensemble learning model for human activity recognition with smartphones. *Sensors* 19:2307. doi: 10.3390/s19102307

Yang, Y., Wu, Q., Qiu, M., Wang, Y., and Chen, X. (2018a). "Emotion recognition from multi-channel eeg through parallel convolutional recurrent neural network," in *2018 International Joint Conference on Neural Networks (IJCNN)* (Rio de Janeiro: IEEE), 1–7. doi: 10.1109/IJCNN.2018.8489331

Yang, Y.-X., Gao, Z.-K., Wang, X.-M., Li, Y.-L., Han, J.-W., Marwan, N., et al. (2018b). A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG. *Chaos* 28:085724. doi: 10.1063/1.5023857

Yao, H., He, H., Wang, S., and Xie, Z. (2019). "EEG-based emotion recognition using multi-scale window deep forest," in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)* (Xiamen: IEEE), 381–386. doi: 10.1109/SSCI44817.2019.9003164

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2008). A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell*. 31, 39–58. doi: 10.1109/TPAMI.2008.52

Zheng, W.-L., Zhu, J.-Y., and Lu, B.-L. (2017). Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans. Affect. Comput*. 10, 417–429. doi: 10.1109/TAFFC.2017.2712143

Zhou, M., Zeng, X., and Chen, A. (2019). Deep forest hashing for image retrieval. *Pattern Recogn*. 95, 114–127. doi: 10.1016/j.patcog.2019.06.005

Zhou, Z.-H., and Feng, J. (2017). Deep forest. *arXiv* 1702.08835.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 13 May 2021
doi: 10.3389/fnbot.2021.659876

Toward Hand Pattern Recognition in Assistive and Rehabilitation Robotics Using EMG and Kinematics

Hui Zhou[1]*, *Qianqian Zhang*[1], *Mengjun Zhang*[1], *Sameer Shahnewaz*[1], *Shaocong Wei*[1], *Jingzhi Ruan*[1], *Xinyan Zhang*[2]* and *Lingling Zhang*[2]

[1] School of Automation, Nanjing University of Science and Technology, Nanjing, China, [2] Affiliated Nanjing Brain Hospital, Nanjing Medical University, Nanjing, China

Wearable hand robots are becoming an attractive means in the facilitating of assistance with daily living and hand rehabilitation exercises for patients after stroke. Pattern recognition is a crucial step toward the development of wearable hand robots. Electromyography (EMG) is a commonly used biological signal for hand pattern recognition. However, the EMG based pattern recognition performance in assistive and rehabilitation robotics post stroke remains unsatisfactory. Moreover, low cost kinematic sensors such as Leap Motion is recently used for pattern recognition in various applications. This study proposes feature fusion and decision fusion method that combines EMG features and kinematic features for hand pattern recognition toward application in upper limb assistive and rehabilitation robotics. Ten normal subjects and five post stroke patients participating in the experiments were tested with eight hand patterns of daily activities while EMG and kinematics were recorded simultaneously. Results showed that average hand pattern recognition accuracy for post stroke patients was 83% for EMG features only, 84.71% for kinematic features only, 96.43% for feature fusion of EMG and kinematics, 91.18% for decision fusion of EMG and kinematics. The feature fusion and decision fusion was robust as three different levels of noise was given to the classifiers resulting in small decrease of classification accuracy. Different channel combination comparisons showed the fusion classifiers would be robust despite failure of specific EMG channels which means that the system has promising potential in the field of assistive and rehabilitation robotics. Future work will be conducted with real-time pattern classification on stroke survivors.

Keywords: EMG, kinematics, hand pattern recognition, sensor fusion, machine learning

Edited by:
Dingguo Zhang,
University of Bath, United Kingdom

Reviewed by:
Jiayuan He,
University of Waterloo, Canada
Siqi Cai,
South China University of Technology, China
Zhan Li,
University of Electronic Science and Technology of China, China

****Correspondence:***
Hui Zhou
zhouhui@njust.edu.cn
Xinyan Zhang
zhzh_0003@163.com

Received: *28 January 2021*
Accepted: *09 April 2021*
Published: *13 May 2021*

Citation:
Zhou H, Zhang Q, Zhang M, Shahnewaz S, Wei S, Ruan J, Zhang X and Zhang L (2021) Toward Hand Pattern Recognition in Assistive and Rehabilitation Robotics Using EMG and Kinematics. Front. Neurorobot. 15:659876. doi: 10.3389/fnbot.2021.659876

INTRODUCTION

Stroke is currently the major cause of disability worldwide and more than 17 million people are estimated to suffer from stroke globally each year (Feigin et al., 2014). 80% acute stroke patients have upper limb motor impairment, and 50% of such post-stroke patients face reduced arm function problems even after 4 years (Bernhardt and Mehrholz, 2019). They experienced loss of sensation, capability, movement, and coordination leading to difficulties surrounding activities of daily living (ADL). Rehabilitation and assistive robotics represent promising treatment methods for post-stroke patients' upper limb recovery and further assist of ADL (Mehrholz et al., 2018).

Since hand function recovery is one of the most important and challenging aspects post stroke, robot based neurorehabilitation and assisting of ADL is a research area of great social and clinical significance.

Electromyogram (EMG) has been extensively used in driving the rehabilitation and assistive robotics for stroke. The extracted information from EMG can be used as a trigger (Dipietro et al., 2005) or as a proportional control strategy in hand rehabilitation robotics (Lenzi et al., 2011; Song et al., 2013). Furthermore, current rehabilitation and ADL assistive devices only use simple extension-flexion mode, but more tasks such as hand opening, grip, lateral pinch, and fist etc. are of imperative need in stroke survivors as well. However, the hand pattern recognition of ADL tasks remain challenging (Lu et al., 2019). Some studies reported the classification results of ADL tasks with very mixed results. The test results from normal subjects in those experiments were promising (Chen et al., 2017; Castiblanco et al., 2020) but for some stroke subjects the accuracy was subpar at best (Lee et al., 2011; Geng et al., 2013; Lu et al., 2019). Hence, more research is needed to develop a practical and effective pattern recognition paradigm that can be used in the rehabilitation and assistive robotics of post stroke patients.

Motion capture technology has been used in many rehabilitation robots to record kinematics data during the task and assess the recovery process post stroke (Rose et al., 2018; Vermillion et al., 2019). Currently, the commonly used motion capture equipment are marker based multicamera motion capture system, such as Vicon and Motion Analysis, etc. However, these devices are expensive and the establishment of markers on the subjects are time consuming. On the contrary, some low cost and non-marker based motion capture technologies have been developed recently, such as Leap motion (Leap Motion Inc., San Francisco, CA, USA) and Kinect (Microsoft, USA). Compared with Kinect, the Leap Motion controller can provide more precise information of the hand and finger movements (Nizamis et al., 2018). Furthermore, Leap motion is widely used as an effective human computer interaction method that can recognize patterns with high accuracy (Lu et al., 2016; Mantecón et al., 2019; Nogales and Benalcazar, 2019; Li et al., 2020). Besides, some Leap motion based hand function assessment system have been developed to evaluate the recovery progress of post stroke patients (Cohen, 2020). Leap motion is fascinating in the area of rehabilitation since it can record hand kinematics data with ease of use, low cost and acceptable precision (Bachmann et al., 2018).

EMG and kinematic recording devices have been jointly used in the state of the art hand rehabilitation robotics for the generation of subject specific task kinematics and assessment of recovery progress post stroke (Rose et al., 2018; Vermillion et al., 2019). For applications in hand rehabilitation and ADL assisting robots, the joint use of EMG and Leap motion controller is an affordable method to obtain muscle activation and kinematics simultaneously from subjects. To our great knowledge, few papers have been reported about the use of EMG and Leap motion sensors together in the area of rehabilitation robotics. In a recent publication, EMG signals and joint trajectories were simultaneously recorded by MYO armband and Leap motion controller (Arteaga et al., 2020). In their work, EMG was used to classify five patterns and three different activation levels, while Leap motion was used to calculate joint trajectories. Besides, it was reported that EMG and Leap motion were used together to recognize hand pattern of six ADL tasks (Ricardez et al., 2018). Integrated myoelectric potential from eight channels were extracted and fused with joint angles to achieve an average classification of 87.3%. It is desirable to develop pattern recognition system which is robust to EMG channel failure and noises. The objective of this study is to develop a pattern recognition paradigm with EMG and kinematics for future use in rehabilitation and assistive robotics with high accuracy and ease use. The contribution of the presented paper will be:

1. Design and implementation of feature fusion and decision fusion of EMG and Kinematics based hand pattern recognition system of daily activities based on EMG and kinematics for post stroke patients.
2. Checking robustness of feature fusion and decision fusion classifiers by adding different levels of noise and failure of EMG channels.

In the following sessions, the methods and results of hand pattern recognition using feature fusion and decision fusion of EMG and kinematics are described and discussed. In session II, the recruitment of subjects, the design of daily activities and experimental protocol, experiment instrument and proposed classification framework are briefly explained. In session III, the hand pattern recognition results are compared between EMG, kinematics, feature fusion and decision fusion of EMG and kinematics. Furthermore, classification results of different EMG channel combination and fusion with kinematics are considered. A discussion is provided in session IV followed by conclusion.

METHODS

Subjects

Ten healthy normal subjects (six male and four females, 20–24 years old) and five post stroke patients (three male and two female, 50–81 years old) participated in the experiment and normal subjects were students of Nanjing University of Science and Technology. All normal subjects had no hand movement disorders and were familiar with patterns and data collection

TABLE 1 | Details of five post stroke patients.

Subject	Age	Gender	Months since onset	Affected side	FMA	FMA-C
S1	76	Female	1	Left	44	14
S2	50	Male	2	Right	43	12
S3	58	Male	1	Right	64	14
S4	54	Male	4	Right	64	13
S5	81	Female	1.5	Left	63	13

FMA, Fugl-Meyer Assessment Upper Extremity, a score between 0 (no function) and 66 (intact); FMA-C, Part C (Hand) of FMA, a score between 0 (no function) and 14 (intact).

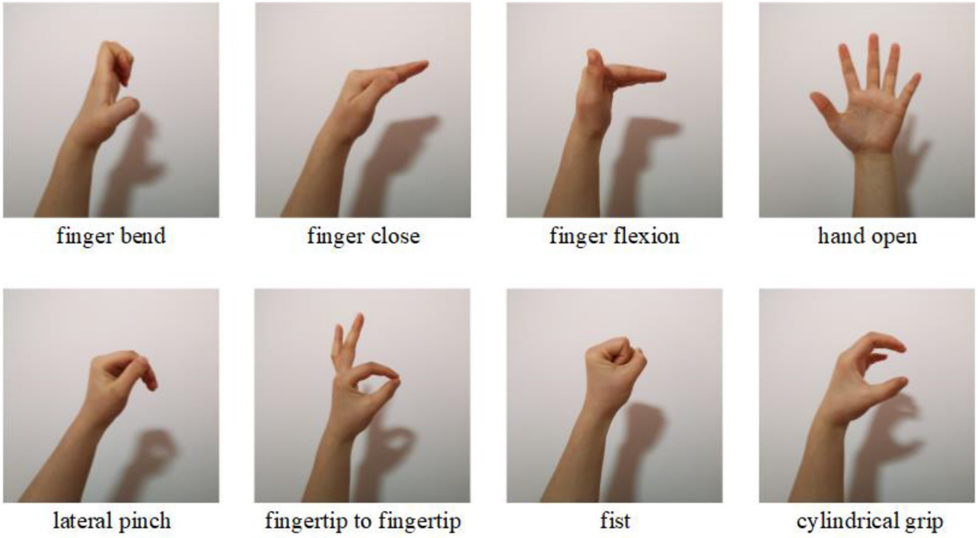


FIGURE 1 | The eight different hand patterns tested in the experiment.

procedures before the start of the experiment. Three post stroke patients' right side was affected and two patients' left side was affected after stroke. Hand corresponding to the affected side of post stroke patients were chosen for pattern recognition procedure. Impairment severity range of the post stroke patients were covered with score of FMA (Fugl-Meyer Assessment Upper Extremity) and FMA-C (Hand part of FMA) (Lu et al., 2019). Detailed information of the post stroke patients are given in **Table 1**. All subjects were informed about the purpose and experimental procedure of the study. The recruitment of subjects and the experimental protocols were approved by the Ethics Committee for Human Research, Nanjing Brain Hospital Affiliated to Nanjing Medical University.

Experiment Protocol

In the experiment, all the subjects were asked to sit on a chair and placed their hands loosely on the table surface. The subject were instructed to perform eight static hand patterns sequentially with their left hand for normal subjects and affected side hand for post stroke patients (finger bend, finger close, finger flexion, etc., see **Figure 1**). Each pattern was held for 5 s, followed by 10 s of rest to prevent muscle fatigue. The process was repeated 10 times for normal subjects and five times for post stroke patients.

Instrumentation and Data Acquisition

In the experiment signals from sensors of EMG and Leap motion were recorded at the same time. EMG signal were measured using the Trigno wireless EMG system (Delsys Inc., Natick, MA, USA) at sampling frequency of 2000 Hz (He et al., 2015). In the study, the muscles of abductor pollicis brevis, flexor carpi radialis, extensor digitorum and extensor carpi ulnaris of the left hand were chosen as the sites for myoelectrical recording, see **Figure 2**. For each subject, the skin of the recording muscles were cleaned with 75% alcohol prior to the start of the experiment, and then the EMG sensors were attached to the muscles with medical grade double sided adhesive tape. Additionally, a bandage (Kindmax Inc., Irvine, CA, USA) was used to fix the sensor to the upper limb to minimize EMG sensor movement and vibration during the tasks. Besides, an infrared motion sensor (Leap Motion Inc., San Francisco, California, USA) was also used to record hand kinematics at sampling rate of 20 Hz. In the study, the subjects were asked to place their left hand directly above the Leap motion controller, ∼30–50 cm, see **Figure 2**. In addition, the participants were required to place their palm of left side (for normal subjects) or palm of affected side (for post stroke patients) facing the device.

Feature Extraction

EMG Feature Extraction

The recorded EMG data were preprocessed via a second-order Butterworth band pass filter of 20–500 Hz. EMG signals of middle 3s were used in the analysis. A 250-ms-lengths sliding window with an overlap of 50 ms was used to segment the raw signal and features were extracted within the window. The following time domain features were used: mean absolute value (MAV), ZC (zero crossing), slope sign changes (SSC), difference absolute mean value (DAMV), and variance (VAR), (Englehart and Hudgins, 2003).

1) MAV

MAV detects the muscle contraction level. MAV is defined as the sum of the absolute value of the EMG signal, a moving time window of N samples is used to calculate it with the following algorithm:

$$MAV = \frac{1}{N}\sum\nolimits_{\mathrm{i}=1}^{N}|x_i|$$

2) Zero Crossings

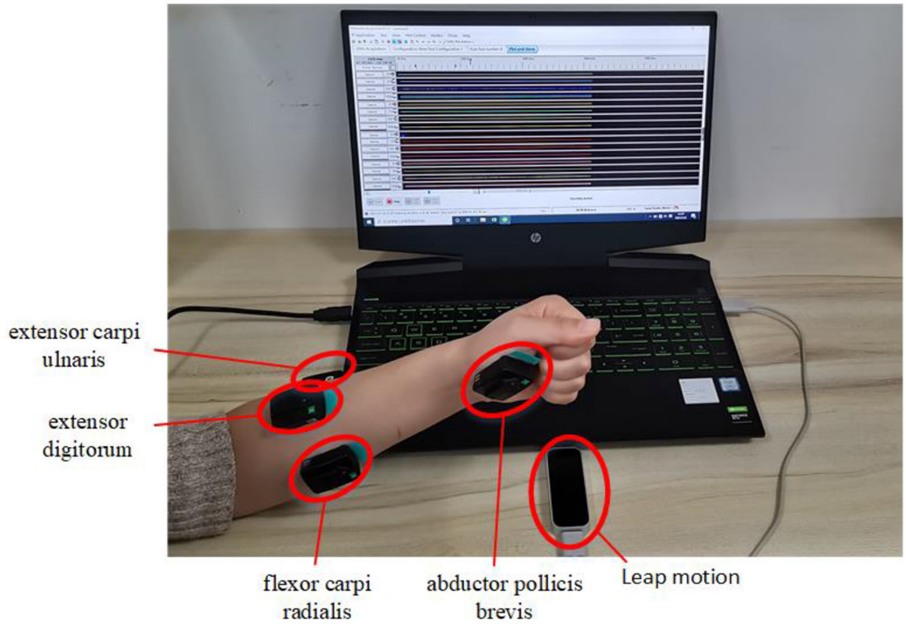


FIGURE 2 | Demonstration of the experimental setup of EMG and kinematic sensors. EMG signals are recorded from muscles of abductor pollicis brevis, flexor carpi radialis, extensor digitorum, and extensor carpi ulnaris of the left hand and kinematics data are recorded from Leap motion sensor.

Zero crossings refers to calculating the number of times the signal waveform passes through the zero point so that it can estimate the frequency domain characteristics.

3) Slope Sign Changes

This parameter calculates the number of sign changes of the signal slope. Similarly, it needs a threshold to reduce the interference caused by noise to the slope sign change.

4) DAMV

DAMV reflects the vibration characteristics of the EMG signal. It is calculated as follows:

$$DAMV = \frac{1}{N}\sum\nolimits_{i=1}^{N-1} |x_{i+1} - x_i|$$

5) VAR

VAR is a measure of the power of the EMG signal. It can be calculated as:

$$VAR = \frac{1}{N-1}\sum_{i=1}^{N} {x_i}^2.$$

Leap Motion Feature Extraction

Similar to EMG processing, Leap motion signal of middle 3s were used in the analysis. A 250-ms-lengths sliding window with an overlap of 50 ms was used to segment leap motion data. For each window, the average of the five frames is extracted as the kinematics data. Each frame of leap motion controller contains instant hand skeleton data of the recorded scene. In the experiment, Leap motion controller was recording at 20 frames per second. Extracted joint angle and finger-to-palm distance were used as kinematic features for recognition of different patterns. A 19-dimensional feature sequence is then established in each frame of kinematic data, see **Figure 3**. Among them, T1 and T2 are the joint angles of the thumb, I1–I3 are the joint angles of index finger, M1–M3 are the joint angles of middle finger, R1–R3 are the joint angles of ring finger, P1–P3 are joint angles of the little thumb, and distances from each of the five fingertip to palm are given as D1–D5.

1) Joint Angle

The joint angle refers to the angle between the joints of the palm, that is, the angle formed between Metacarpal and Proximal, Proximal and Intermediate, and Intermediate and Distal of each finger. The calculation formula is:

$$\theta = \arccos \frac{\vec{u} \cdot \vec{v}}{|\vec{u}| \cdot |\vec{v}|}$$

$\vec{u}$ and $\vec{v}$ are the two adjacent parts linked to the joint of a finger. The calculated joint angle range is [0°,180°]. When the palm is fully opened, the angle value of each joint reaches the maximum. Since the Thumb finger does not have Metacarpal, there are only two joint angles, and the remaining fingers have three joint angles each. Thus, a total of 14 joint angle feature values were used for one hand, see **Figure 3**.

2) Finger-to-Palm Distance

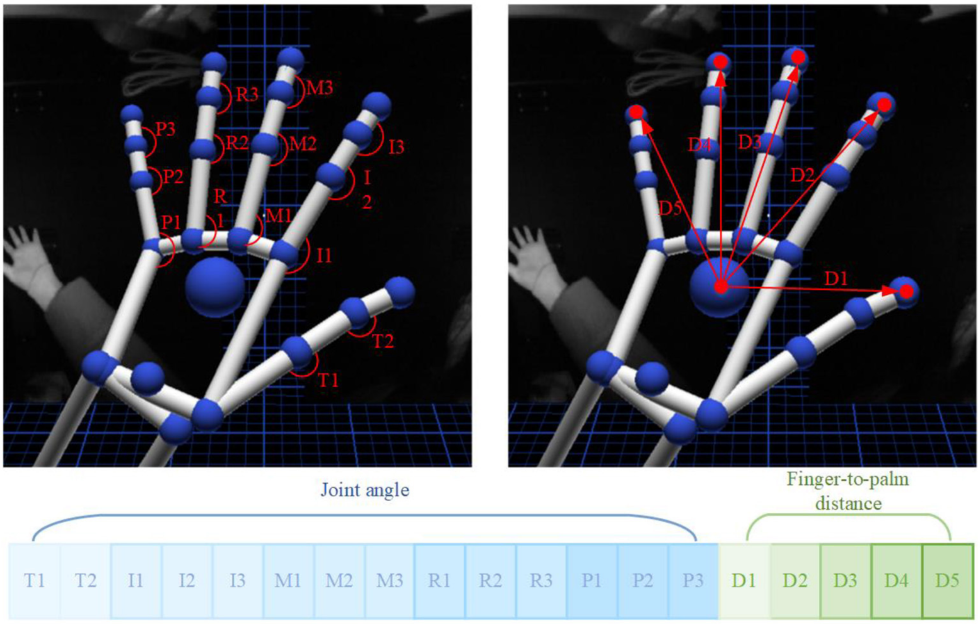


FIGURE 3 | Joint angle and finger-to-palm distance kinematic features were extracted from Leap motion controller.

The distance from the fingertip to the palm is calculated as the Euclidean distance between the fingertip of each finger and the palm. The calculation formula is:

$$D_i = ||\mathrm{P_i} - O||,_{i\, =\, 1,2,3,4,5}$$

Point Pi is the position coordinate of the fingertip of each finger, and point O is the position coordinate of the palm. When the palm is fully opened the distance from the fingertip to the palm reaches the maximum. A total of five fingertip distance were used as distances features, see **Figure 3**.

Proposed Classification Framework

Feature fusion and decision fusion of EMG and kinematics were used to achieve hand pattern recognition, respectively. Among them, the feature sequences extracted from EMG data and Leap Motion data were fused together as inputs to the classifier, which is called feature fusion shown in **Figure 4A**. For decision fusion, the Leap Motion feature and the EMG feature were first input to the classifier separately, then the probabilities output of the two classifiers were superimposed to finally obtain the classification result, see **Figure 4B**.

Linear discrimination analysis (LDA) is an easy and effective classifier that was used because it gives high performance despite low computational cost and robustness for long term usage (Chen et al., 2013). The experiment uses a 10-fold cross-validation method to train the classifier multiple times to ensure that each data can be used as training data and test data. The accuracy of each test is the ratio of the number of correctly recognized patterns to the total number of tested patterns. The classification rate of each subject is taken as the average of 10 times of cross-validation training results. In this experiment, the final accuracies were calculated as the mean of 10 healthy subjects and five post stroke patients. Python programming language was used to implement the classification algorithm. One-Way ANOVA was used to compare the classification results of EMG, kinematics, feature fusion and decision fusion of EMG and kinematics, respectively. The significance threshold was set to 0.05 in the experiment.

Noise Contamination

Performance of a classification model in the presence of noise is an important area to address. Hence, we must assess the model's tolerance of input perturbations. For that reason, three different levels of Gaussian noise are added to the raw EMG data to contaminate it and then the contaminated data was fed to the classifiers. Since raw inputs' given order of magnitude is 10^{-5}, we set the noise levels accordingly to 1×10^{-5}, 2×10^{-5}, and 1×10^{-4} (Jia, 2020).

RESULT

Comparison of Classification Performance for Different Fusion Strategy

Representative EMG and kinematic data of different patterns from one normal subject are shown in **Figure 5**. In the figure, EMG signals are recorded from muscles of abductor pollicis brevis, flexor carpi radialis, extensor carpi ulnaris, and extensor digitorum. It can be observed that muscle activity patterns are different within four muscles during different tasks. In addition, Leap motion sensor were used to track the participant's hand to monitor the kinematics. Then the provided features of joint angles and finger-to-palm distances were estimated from the

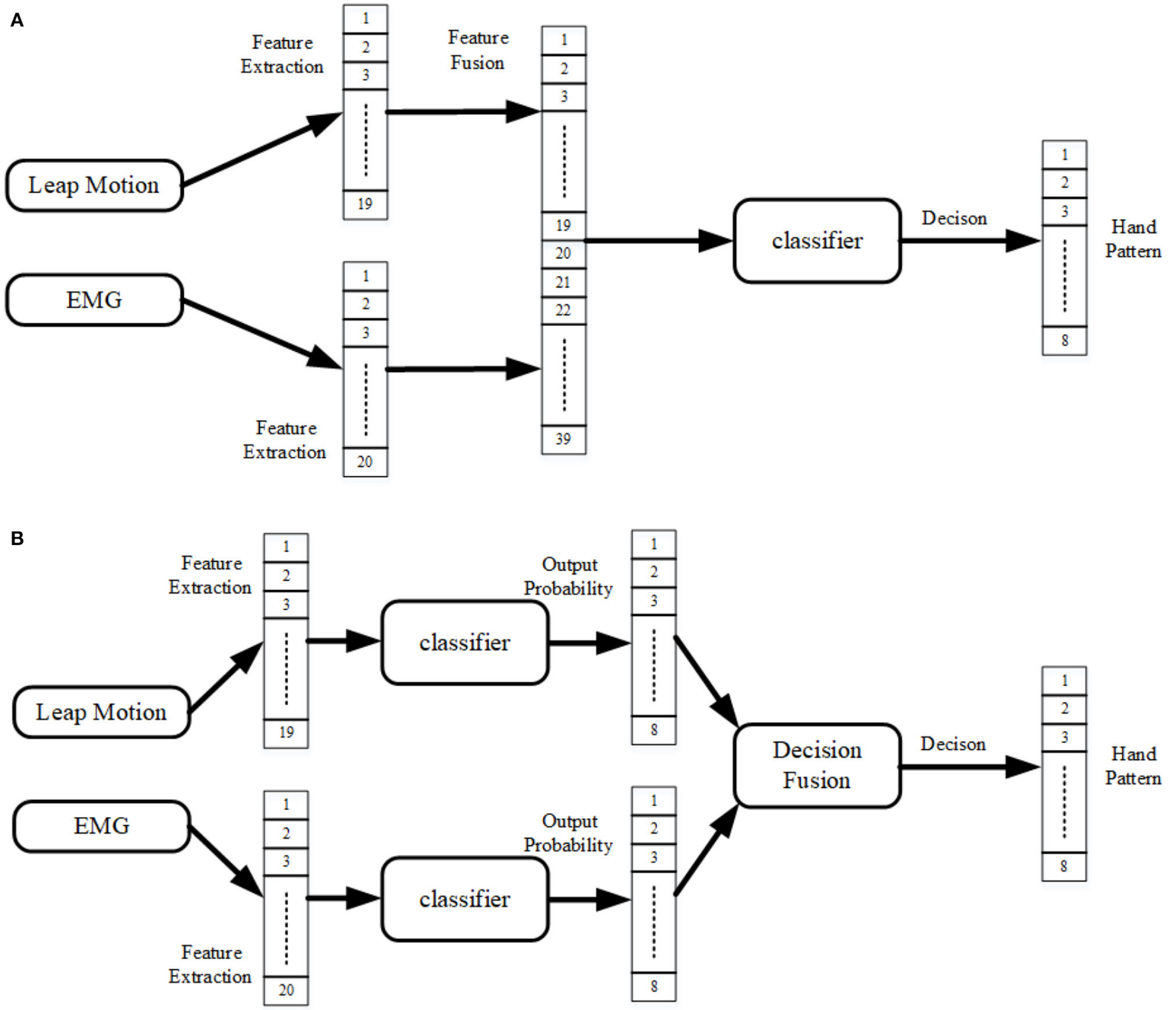


FIGURE 4 | Proposed classification framework of the system. **(A)** Feature fusion of EMG and kinematics (leap motion); **(B)** Decision fusion of EMG and kinematics (leap motion).

kinematics data. It can be seen from the figure that the kinematics are quite different between the eight hand patterns.

For 10 normal subjects, average hand pattern recognition accuracy was 88.71 ± 2.79% for EMG features only, 84.49 ± 6.77% for kinematic features only, 96.90 ± 1.81% for feature fusion of EMG and kinematics, 93.91 ± 2.57 % for decision fusion of EMG and kinematics (**Figure 6**). It can be observed after feature fusion that classification accuracy have increased to 96.90% and standard deviation have decreased to 1.81%. From the **Figure 6**, it can also be seen that the accuracy of pattern recognition using feature fusion of EMG and kinematic is significantly larger than that of using kinematic features alone ($p = 0.001$). Besides, decision fusion method comparison with kinematics only was statistically highly significant as well ($p = 0.01$). Moreover, there was significant difference between EMG and decision fusion of EMG and kinematic ($p = 0.003$,). In addition, there was no significant difference between kinematic features only and EMG features only in classification rate. Difference between feature fusion and decision fusion of kinematics and EMG was significant ($p = 0.05$).

For the post stroke patients, hand pattern recognition accuracy was 83 ± 8.21% for EMG features only, 84.71 ± 4.54% for kinematic features only, 96.43 ± 3.83% for feature fusion of EMG and kinematics, 91.18 ± 5.50% for decision fusion of EMG and kinematics (**Figure 7**). It is observable that both fusion methods are able to achieve average accuracy higher than 90%. From **Figure 7**, it is noticeable that pattern recognition accuracy using fusion features of kinematics and

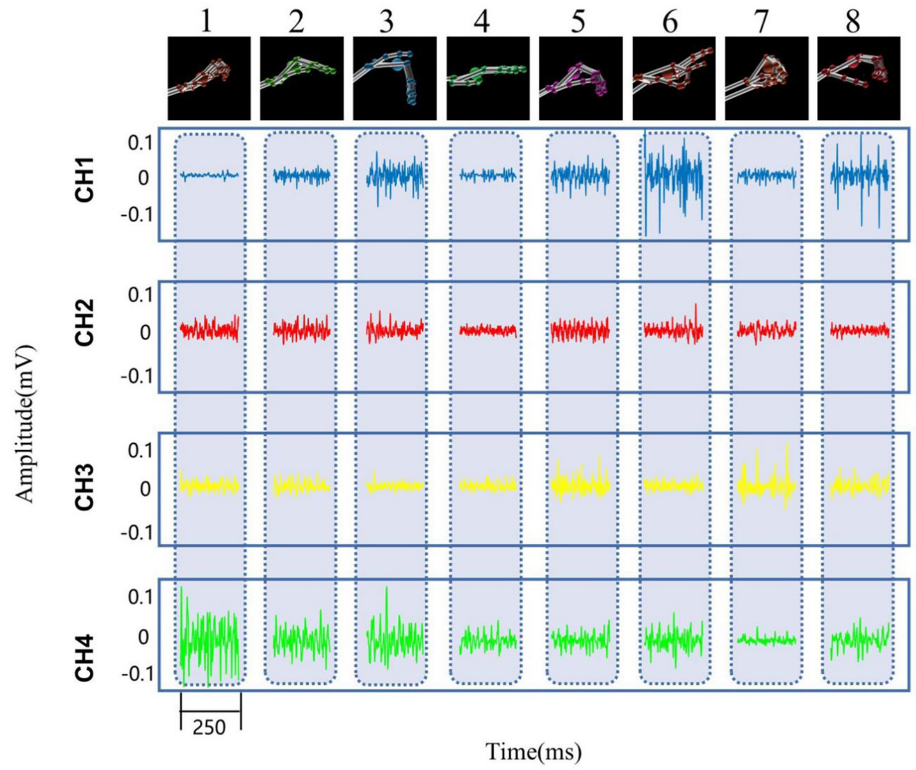


FIGURE 5 | Representative EMG and kinematic data of eight hand patterns recorded from one subject. Patterns 1–8 are: finger bend, finger close, finger flexion, hand open, lateral pinch, fingertip to fingertip, fist, and cylindrical grip, respectively. The top row kinematic data are recorded from Leap motion. The EMG data of CH1–CH4 are recorded from muscles of abductor pollicis brevis, flexor carpi radialis, extensor digitorum, and extensor carpi ulnaris, respectively.

EMG is significantly larger than solely using kinematics features ($p = 0.019$). Additionally, there was no significant difference in classification accuracy between decision fusion of kinematics and EMG method and kinematics features only method ($p = 0.151$).

Classification Performance With Noise on EMG Recordings

Comprehensive comparison between three levels of noisy contaminated data was made for EMG only, decision fusion and feature fusion of EMG and kinematics. **Table 2** displays robustness result for normal subjects and post stroke subjects. It can be observed from **Table 2** that classifiers realize near exact test accuracy when provided with 1×10^{-5} level of additive noise. But 1×10^{-4} level of additive noise decreases test accuracy greatly for EMG only classifier for both normal subjects and post stroke subjects. For 1×10^{-4} level of noisy data EMG decreases test accuracy from 85.91 to 58.88% for normal subjects and from 81.07 to 53.50% for post stroke patients. But test accuracy of decision fusion and feature fusion did not suffer such sharp decrease. For decision fusion decrease was from 93.54 to 88.13% for normal subjects and from 91.11 to 86.04% for post stroke patients. Similarly, feature fusion decrease was from 96.61 to 90.13% for normal subjects and from 95.61 to 91.07% for post stroke patients. It is distinguished that feature fusion method provides best performance which is robust to noises.

Classification Performance With Different Channel Combinations

Comparison of hand pattern recognition accuracy between different classification models are made with 15 different channels combinations. Combinations are as following: four single channels, six dual channels, four three channel combinations, and one with all four channels combined. Investigation is carried out for both normal subjects and post stroke patients.

Figure 8 displays recognition accuracy with different channels combination for normal subjects. The line showing 0.8449 represents pattern recognition accuracy using kinematics only. By increasing EMG channels from one to four, the overall classification accuracies are increased for EMG only classifier. When a single EMG channel is used for recognition, the highest recognition accuracy is produced by C1 for EMG only classifier. For dual channel combinations, C23 produced lowest accuracy, and C14 produced highest accuracy for EMG only classifier. Furthermore, between three channel combinations, C124 and C234 produced highest and lowest recognition accuracy for EMG only classifiers. And from the **Figure 8** we can see combining all four channels into one yields the best overall accuracy for all classifiers. Even with cases of single EMG channel, it is evident the test accuracy is still high after fusion with kinematics. This fusion based method makes the classifier robust despite potential failure

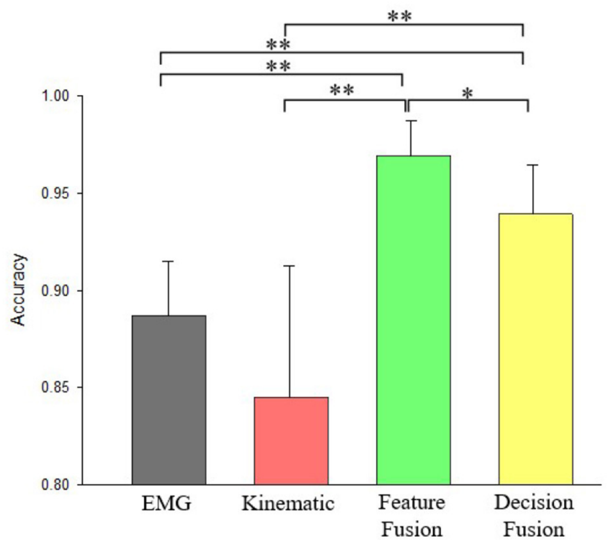


FIGURE 6 | Accuracy comparisons of pattern recognition using EMG only, kinematics only, feature fusion of kinematics+EMG, and decision fusion of kinematics+EMG for normal subjects. * Denotes $p < 0.05$, ** Denotes $p < 0.01$.

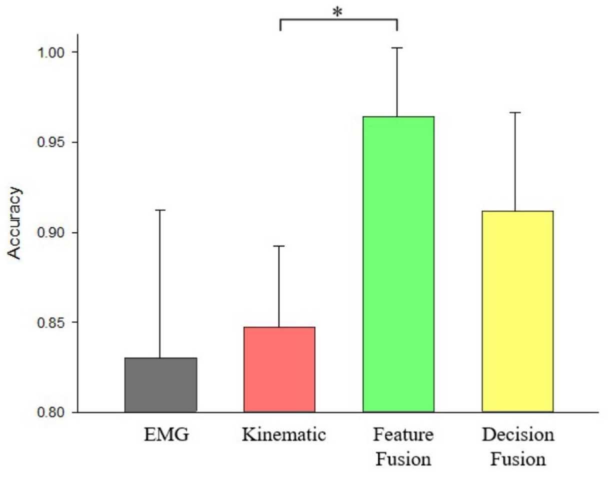


FIGURE 7 | Accuracy comparisons of pattern recognition using EMG only, kinematics only, feature fusion of kinematics+EMG, and decision fusion of kinematics+EMG for post stroke patients. * Denotes $p < 0.05$.

of specific EMG channels. Furthermore, feature fusion method produces slightly better classification performance than decision fusion method in this study.

On the other hand, **Figure 9** shows the recognition results with different channels for post stroke patients. From **Figure 9** it is visible that EMG only classifier produces lower test accuracy for post stroke patients than normal subjects. Classification accuracies are increased by rising the number of EMG channels, just as observed for normal subjects. Besides, it was apparent for both normal subjects and post stroke patients that channel combinations that included the channel C1 gave better performance than the combinations that didn't include this specific channel. This suggests C1 would be the optimal recording site for hand pattern recognition with EMG based classifier. Moreover, the lower classification accuracy were improved by fusion with kinematics. After fusion of EMG data with kinematics the classification performances among post stroke patient group are very close to results obtained for normal subjects.

DISCUSSION

In this study, a pattern recognition method was proposed that combined EMG with kinematics data for classification toward application in upper limb assistive and rehabilitation robotics. The feasibility of the proposed method was demonstrated by conducting experiments on 10 normal subjects and five post stroke patients. Classification accuracy was improved with the fusion method as only EMG and kinematics classification separately were unsatisfactory. Robustness of the model was validated with noise and channel combination comparison.

Fusion of EMG With Kinematics for Hand Pattern Recognition

EMG has been widely used for pattern recognition for many years, especially for applications in prosthesis (Lee and Saridis, 1984; Ajiboye and Weir, 2005; Kuiken et al., 2009). The use of EMG in driving assistive and rehabilitation robotics post stroke is increasing recently (Hu et al., 2013; Leonardis et al., 2015; Park et al., 2020). However, the EMG based pattern recognition performance in assistive and rehabilitation robotics post stroke remains unsatisfactory (Lu et al., 2019). In recent years, kinematic sensors have been increasingly used for hand pattern recognition. Various types of kinematics sensors were used for pattern recognition, such as motion capture devices (Pun et al., 2011), inertial measurement unit (IMU) (Kim et al., 2019), and strain sensors (Ferrone et al., 2016) etc. However, most commercially available kinematics devices are expensive and complex to set up such as wearable data gloves, Vicon motion sensor system. While on the other hand, leap motion device is a cheap, convenient and markerless hand kinematic recording device. Hence, combining kinematics obtained from leap motion device with EMG was interesting to make the proposed model cost-effective.

In this study, we propose a method of combining EMG and kinematics together for hand pattern recognition toward assistive and rehabilitation robots. The proposed fusion methods of EMG and kinematics can improve classification accuracy of hand patterns in daily activities. Experiments using proposed model of feature fusion and decision fusion was conducted on 10 normal subjects and five post stroke patients showed increased classification accuracy. This classification accuracy is comparable to other sensors fusion method such as EMG and force myograph (FMG) (Jiang et al., 2020), EMG and IMU (Georgi et al., 2015), and EMG and strain sensors (Landgraf et al., 2018). However, Leap motion is easy to use and off-the-shelf compared with these sensors.

TABLE 2 | Test accuracy of classifiers when given different levels of noise.

Noise level	Normal subjects			Patient after stroke		
	EMG	**Feature fusion**	**Decision fusion**	**EMG**	**Feature fusion**	**Decision fusion**
1e−5	85.91%	93.54%	96.61%	81.07%	91.11%	95.61%
2e−5	81.87%	92.37%	95.57%	77.21%	89.89%	95.39%
1e−4	58.88%	88.13%	90.13%	53.50%	86.04%	91.07%

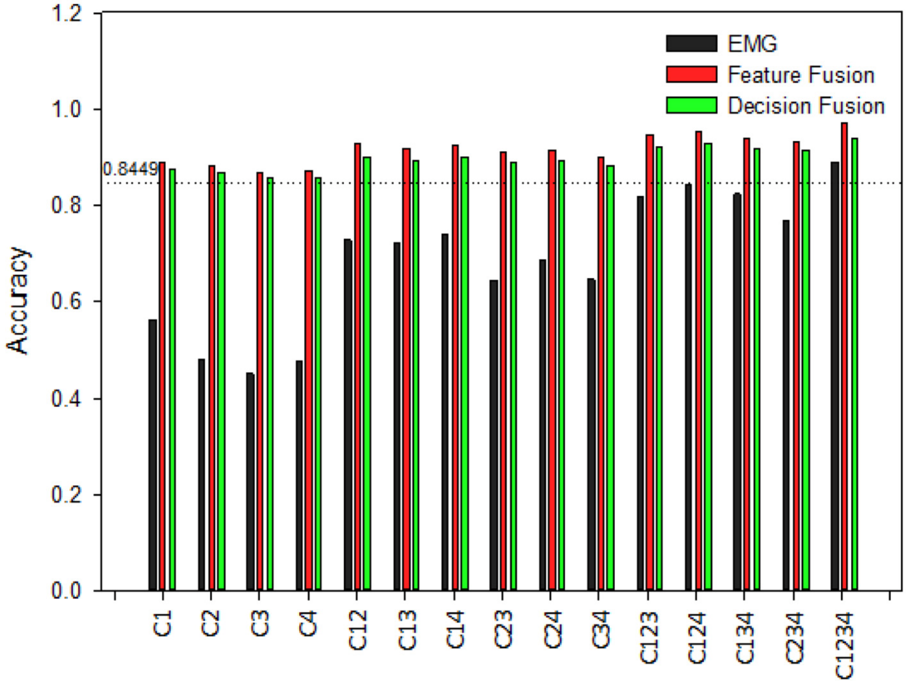


FIGURE 8 | Classification accuracy of different channel combinations for hand pattern recognition for normal subjects. C1–C4 are recorded from muscles of abductor pollicis brevis, flexor carpi radialis, extensor digitorum and extensor carpi ulnaris, respectively.

To our great knowledge, very few works has been reported about the fusion of EMG and Leap motion based kinematics for hand pattern recognition of daily activities. High accuracy was successfully achieved for post stroke patients by implementing the proposed method of feature fusion and decision fusion from EMG channels and Leap motion. The model obtained satisfactory test accuracy of 96.43% for feature fusion and 91.18% for decision fusion of EMG and kinematics. Ricardez et al. (2018) conducted similar experiment on only three normal subjects using 8-channel EMG and leap motion but overall classification accuracy was <90%. No publication yet have provided experimental results on stroke patients with fusion of EMG and Leap motion for hand pattern recognition. Highly accurate, cost-effective, and limited EMG electrode nature of the fusion method would be meaningful in a practical sense.

Robustness Analysis of the Proposed Methods

Checking robustness of the proposed method is imperative to investigate its stability against different levels of noise. **Table 2** displayed comparison of classification models after fed with three different level of noise. The chosen levels of noise were 1×10^{-5}, 2×10^{-5}, and 1×10^{-4}. **Table 2** showed great decrease in test accuracy for EMG only classifier whereas, the proposed feature fusion and decision fusion of EMG and kinematics classifiers saw small decrease meaning they are robust to EMG noise.

To check robustness of the classifier on the condition that one or multiple EMG channels fail, we compare the classification performance of 15 different EMG channel combinations. If channels fails, the overall classification accuracies would decrease with the number of channels. However, after fusion with kinematics the classifier would be robust from channels failure. It was observed that after fusion of single channel EMG with kinematics, the classification accuracy could reach close to 90% for both normal subjects and post stroke patients. Hence, the proposed fusion method would use minimal number of EMG electrodes without the deterioration of the classification accuracy. To our great knowledge, very few publications have verified robustness of fusion of EMG and Leap motion.

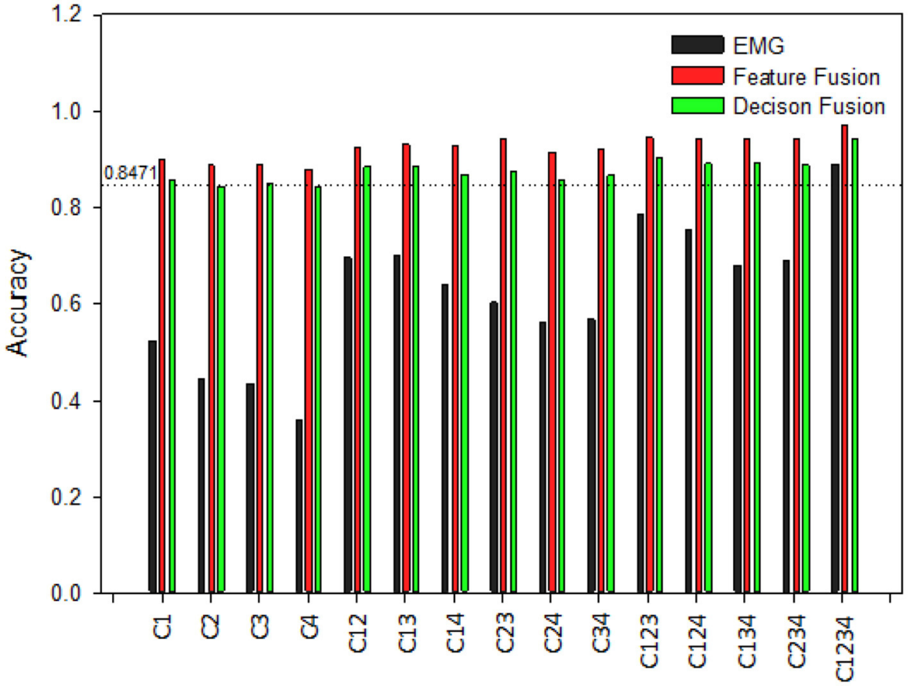


FIGURE 9 | Classification accuracy of different channel combinations for hand pattern recognition for post stroke patients. C1–C4 are recorded from muscles of abductor pollicis brevis, flexor carpi radialis, extensor digitorum, and extensor carpi ulnaris, respectively.

Limitation of the Study

In this study, FCA and FCA-C score shows that the five recruited patients had mild to no restrictions on activities of daily life. However, most post stroke patients have low FCA score and require assistance of robotics for daily activities. Lu et al. (2019) reported the feasibility of hand pattern recognition under the assistance of robotic hand on post stroke patients with FCA-C score ranging from 0 to 7. In this study, rehabilitation and assistive robotics haven't yet been implemented for stroke patients. In future, we would test our classification accuracy on stroke patients with lower FCA score with the help of assistive and rehabilitation robotics.

The proposed method uses simple feature fusion and decision fusion because the main purpose of this study was to validate the feasibility of fusion with EMG and kinematics for hand pattern recognition in normal subjects and stroke patients. The results produced an average accuracy of above 90% for both feature fusion and decision fusion method. In the future, we would use more complex algorithms to optimize fusion methods and further improve the classification accuracy. Another limitation of the proposed method is that the dataset were trained offline and no online experiment was done. Future work will focus on real-time pattern recognition using the method presented here.

CONCLUSION

This paper proposes a method that applies feature fusion and decision fusion using EMG features and kinematic features for hand pattern recognition toward application in upper limb assistive and rehabilitation robotics. The results showed that the infused kinematic features could improve the classification accuracy compared with EMG features only for hand pattern recognition of ADLs. Robustness of the proposed method was demonstrated by adding noise to EMG data and comparison between different channel combinations. Classification accuracy of feature fusion and decision fusion was above 90% for both normal subjects and post stroke patients which means that the system has significant potential in the field of assistive and rehabilitation robotics. Future work will be conducted with real-time pattern classification on stroke survivors.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Nanjing Brain Hospital Affiliated to Nanjing Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HZ and XZ initiated and supervised the research project. QZ, MZ, JR, SW, and LZ carried out the research. HZ, XZ, QZ, and SS wrote part of the manuscript. QZ, MZ, XZ, SW, LZ, JR, and HZ analyzed the results and prepared the figures and tables. All the authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (Grant U1613228), Siyuan Foundation (Grant HTKJ2020KL012002) and the open foundation of Key Laboratory of Biorheological Science and Technology (Chongqing University), Ministry of Education (Grant CQKLBST-2020-005).

ACKNOWLEDGMENTS

The authors would like to thank bachelor students Zhe Ma and Chengyuan Fang for their research contribution. We would also like to thank the reviewers for their valuable remarks and comments.

REFERENCES

Ajiboye, A. B., and Weir, R. F. (2005). A heuristic fuzzy logic approach to EMG pattern recognition for multifunctional prosthesis control. *IEEE Trans. Neural Syst. Rehabil. Eng.* 13, 280–291. doi: 10.1109/TNSRE.2005.847357

Arteaga, M. V., Castiblanco, J. C., Mondragon, I. F., Colorado, J. D., and Alvarado-Rojas, C. (2020). "EMG-based adaptive trajectory generation for an exoskeleton model during hand rehabilitation exercises," in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob). Presented at the 2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)* (New York, NY: IEEE), 416–421.

Bachmann, D., Weichert, F., and Rinkenauer, G. (2018). Review of three-dimensional human-computer interaction with focus on the leap motion controller. *Sensors* 18:2194. doi: 10.3390/s18072194

Bernhardt, J., and Mehrholz, J. (2019). Robotic-assisted training after stroke: RATULS advances science. *Lancet* 394, 6–8. doi: 10.1016/S0140-6736(19)31156-0

Castiblanco, J. C., Ortmann, S., Mondragon, I. F., Alvarado-Rojas, C., Jöbges, M., and Colorado, J. D. (2020). Myoelectric pattern recognition of hand motions for stroke rehabilitation. *Biomed. Signal Process. Control* 57:101737. doi: 10.1016/j.bspc.2019.101737

Chen, M., Cheng, L., Huang, F., Yan, Y., and Hou, Z.-G. (2017). "Towards robot-assisted post-stroke hand rehabilitation: fugl-meyer gesture recognition using sEMG," in *2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER). Presented at the 2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)* (Honolulu, HI: IEEE), 1472–1477.

Chen, X., Zhang, D., and Zhu, X. (2013). Application of a self-enhancing classification method to electromyography pattern recognition for multifunctional prosthesis control. *J. NeuroEng. Rehabil.* 10:44. doi: 10.1186/1743-0003-10-44

Cohen, M. W. (2020). Hand rehabilitation assessment system using leap motion controller. *AI Soc.* 35, 581–594. doi: 10.1007/s00146-019-00925-8

Dipietro, L., Ferraro, M., Palazzolo, J. J., Krebs, H. I., Volpe, B. T., and Hogan, N. (2005). Customized interactive robotic treatment for stroke: EMG-triggered therapy. *IEEE Trans. Neural Syst. Rehabil. Eng.* 13, 325–334. doi: 10.1109/TNSRE.2005.850423

Englehart, K., and Hudgins, B. (2003). A robust, real-time control scheme for multifunction myoelectric control. *IEEE Trans. Biomed. Eng.* 50, 848–854. doi: 10.1109/TBME.2003.813539

Feigin, V. L., Forouzanfar, M. H., Krishnamurthi, R., Mensah, G. A., Connor, M., Bennett, D. A., et al. (2014). Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. *Lancet* 383, 245–255. doi: 10.1016/S0140-6736(13)61953-4

Ferrone, A., Maita, F., Maiolo, L., Arquilla, M., Castiello, A., Pecora, A., et al. (2016). "Wearable band for hand gesture recognition based on strain sensors," in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob). Presented at the 2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)* (Singapore: IEEE), 1319–1322.

Geng, Y., Zhang, L., Tang, D., Zhang, X., and Li, G. (2013). "Pattern recognition based forearm motion classification for patients with chronic hemiparesis," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Presented at the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Osaka: IEEE), 5918–5921.

Georgi, M., Amma, C., and Schultz, T. (2015). "Recognizing hand and finger gestures with IMU based motion and EMG based muscle activity sensing," in *Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing. Presented at the International Conference on Bio-Inspired Systems and Signal Processing* (Lisbon: SCITEPRESS–Science and and Technology Publications), 99–108.

He, J., Zhang, D., Jiang, N., Sheng, X., Farina, D., and Zhu, X. (2015). User adaptation in long-term, open-loop myoelectric training: implications for EMG pattern recognition in prosthesis control. *J. Neural Eng.* 12:046005. doi: 10.1088/1741-2560/12/4/046005

Hu, X. L., Tong, K. Y., Wei, X. J., Rong, W., Susanto, E. A., and Ho, S. K. (2013). T he effects of post-stroke upper-limb training with an electromyography (EMG)-driven hand robot. *J. Electromyogr. Kinesiol.* 23:1065–1074. doi: 10.1016/j.jelekin.2013.07.007

Jia, G. (2020). Classification of electromyographic hand gesture signals using modified fuzzy C-means clustering and two-step machine learning approach. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28:8. doi: 10.1109/TNSRE.2020.2986884

Jiang, S., Gao, Q., Liu, H., and Shull, P. B. (2020). A novel, co-located EMG-FMG-sensing wearable armband for hand gesture recognition. *Sens. Actuators Phys.* 301:111738. doi: 10.1016/j.sna.2019.111738

Kim, M., Cho, J., Lee, S., and Jung, Y. (2019). IMU sensor-based hand gesture recognition for human-machine interfaces. *Sensors* 19:3827. doi: 10.3390/s19183827

Kuiken, T. A., Li, G., Lock, B. A., Lipschutz, R. D., Miller, L. A., Stubblefield, K. A., et al. (2009). Targeted muscle reinnervation for real-time myoelectric control of multifunction artificial arms. *JAMA*. 301, 619–628. doi: 10.1001/jama.2009.116

Landgraf, M., Yoo, I. S., Sessner, J., Mooser, M., Kaufmann, D., Mattejat, D., et al. (2018). "Gesture recognition with sensor data fusion of two complementary sensing methods*", in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob). Presented at the 2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)* (Enschede: IEEE), 795–800.

Lee, S., and Saridis, G. (1984). The control of a prosthetic arm by EMG pattern recognition. *IEEE Trans. Automat. Control* 29, 290–302. doi: 10.1109/TAC.1984.1103521

Lee, S. W., Wilson, K. M., Lock, B. A., and Kamper, D. G. (2011). Subject-specific myoelectric pattern classification of functional hand movements for stroke survivors. *IEEE Trans. Neural Syst. Rehabil. Eng.* 19:9. doi: 10.1109/TNSRE.2010.2079334

Lenzi, T., De Rossi, S. M. M., Vitiello, N., and Carrozza, M. C. (2011). "Proportional EMG control for upper-limb powered exoskeletons," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Presented at the 2011 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Boston, MA: IEEE), 628–631.

Leonardis, D., Barsotti, M., Loconsole, C., Solazzi, M., Troncossi, M., Mazzotti, C., et al. (2015) An EMG-controlled robotic hand exoskeleton for bilateral rehabilitation. *IEEE Trans. Haptics*. 8, 140–151. doi: 10.1109/TOH.2015.2417570

Li, H., Wu, L., Wang, H., Han, C., Quan, W., and Zhao, J. (2020). Hand gesture recognition enhancement based on spatial fuzzy matching in leap motion. *IEEE Trans. Ind. Inform*. 16, 1885–1894. doi: 10.1109/TII.2019.2931140

Lu, W., Tong, Z., and Chu, J. (2016). Dynamic hand gesture recognition with leap motion controller. *IEEE Signal Process. Lett*. 23, 1188–1192. doi: 10.1109/LSP.2016.2590470

Lu, Z., Tong, K., Zhang, X., Li, S., and Zhou, P. (2019). Myoelectric pattern recognition for controlling a robotic hand: a feasibility study in stroke. *IEEE Trans. Biomed. Eng*. 66:8. doi: 10.1109/TBME.2018.2840848

Mantecón, T., del-Blanco, C. R., Jaureguizar, F., and García, N. (2019). A real-time gesture recognition system using near-infrared imagery. *PLoS ONE* 14:e0223320. doi: 10.1371/journal.pone.0223320

Mehrholz, J., Pohl, M., Platz, T., Kugler, J., and Elsner, B. (2018). Electromechanical and robot-assisted arm training for improving activities of daily living, arm function, and arm muscle strength after stroke. *Cochrane Database Syst. Rev.* 2015:CD006876. doi: 10.1002/14651858.CD006876.pub5

Nizamis, K., Rijken, N., Mendes, A., Janssen, M., Bergsma, A., and Koopman, B. (2018). A novel setup and protocol to measure the range of motion of the wrist and the hand. *Sensors* 18:3230. doi: 10.3390/s18103230

Nogales, R., and Benalcazar, M. (2019). "Real-time hand gesture recognition using the leap motion controller and machine learning," in *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI). Presented at the 2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)* (Guayaquil: IEEE), 1–7.

Park, S., Fraser, M., Weber, L.M., Meeker, C., Bishop, L., Geller, D., et al. (2020). User-driven functional movement training with a wearable hand robot after stroke. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2265–2275. doi: 10.1109/TNSRE.2020.3021691

Pun, C.-M., Zhu, H.-M., and Feng, W. (2011). Real-time hand gesture recognition using motion tracking. *Int. J. Comput. Intell. Syst*. 4, 277–286. doi: 10.1080/18756891.2011.9727783

Ricardez, G. A. G., Ito, A., Ding, M., Yoshikawa, M., Takamatsu, J., Matsumoto, Y., et al. (2018). "Wearable device to record hand motions based on EMG and visual information," in *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA). Presented at the 2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)* (Oulu: IEEE), 1–6.

Rose, C. G., Pezent, E., Kann, C. K., Deshpande, A. D., and O'Malley, M. K. (2018). Assessing wrist movement with robotic devices. *IEEE Trans. Neural Syst. Rehabil. Eng*. 26, 1585–1595. doi: 10.1109/TNSRE.2018.2853143

Song, R., Tong, K., Hu, X., and Zhou, W. (2013). Myoelectrically controlled wrist robot for stroke rehabilitation. *J. NeuroEng. Rehabil.* 10:52. doi: 10.1186/1743-0003-10-52

Vermillion, B. C., Dromerick, A. W., and Lee, S. W. (2019). Toward restoration of normal mechanics of functional hand tasks post-stroke: subject-specific approach to reinforce impaired muscle function. *IEEE Trans. Neural Syst. Rehabil. Eng*. 27, 1606–1616. doi: 10.1109/TNSRE.2019.2924208

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Neurorobotics

ORIGINAL RESEARCH
published: 16 July 2021
doi: 10.3389/fnbot.2021.692562

Face-Computer Interface (FCI): Intent Recognition Based on Facial Electromyography (fEMG) and Online Human-Computer Interface With Audiovisual Feedback

Bo Zhu[1,2,3], *Daohui Zhang*[1,2]*, *Yaqi Chu*[1,2,3], *Xingang Zhao*[1,2]*, *Lixin Zhang*[4] *and Lina Zhao*[4]

[1] *State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China,* [2] *Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, China,* [3] *University of Chinese Academy of Sciences, Beijing, China,* [4] *Rehabilitation Center, Shengjing Hospital of China Medical University, Shenyang, China*

Edited by:
Dingguo Zhang,
University of Bath, United Kingdom

Reviewed by:
Hong Zeng,
Southeast University, China
Zhiyuan Lu,
University of Texas Health Science Center at Houston, United States
Yan Chen,
South China University of Technology, China

***Correspondence:**
Daohui Zhang
zhangdaohui@sia.cn
Xingang Zhao
zhaoxingang@sia.cn

Received: *08 April 2021*
Accepted: *21 June 2021*
Published: *16 July 2021*

Citation:
Zhu B, Zhang D, Chu Y, Zhao X, Zhang L and Zhao L (2021) Face-Computer Interface (FCI): Intent Recognition Based on Facial Electromyography (fEMG) and Online Human-Computer Interface With Audiovisual Feedback.
Front. Neurorobot. 15:692562.
doi: 10.3389/fnbot.2021.692562

Patients who have lost limb control ability, such as upper limb amputation and high paraplegia, are usually unable to take care of themselves. Establishing a natural, stable, and comfortable human-computer interface (HCI) for controlling rehabilitation assistance robots and other controllable equipments will solve a lot of their troubles. In this study, a complete limbs-free face-computer interface (FCI) framework based on facial electromyography (fEMG) including offline analysis and online control of mechanical equipments was proposed. Six facial movements related to eyebrows, eyes, and mouth were used in this FCI. In the offline stage, 12 models, eight types of features, and three different feature combination methods for model inputing were studied and compared in detail. In the online stage, four well-designed sessions were introduced to control a robotic arm to complete drinking water task in three ways (by touch screen, by fEMG with and without audio feedback) for verification and performance comparison of proposed FCI framework. Three features and one model with an average offline recognition accuracy of 95.3%, a maximum of 98.8%, and a minimum of 91.4% were selected for use in online scenarios. In contrast, the way with audio feedback performed better than that without audio feedback. All subjects completed the drinking task in a few minutes with FCI. The average and smallest time difference between touch screen and fEMG under audio feedback were only 1.24 and 0.37 min, respectively.

Keywords: face-computer interface, facial electromyography, facial movements, robotic arm control online, rehabilitation assistance robot

1. INTRODUCTION

Patients with paralysis and amputation are usually accompanied by loss of limb motor function. Particularly, for the patients with upper limb amputation, high paraplegia, or muscle weakness, it is hard to take care of themselves due to the loss of partial or total motor functions of hands or feet. In order to restore the patient's lost limb function or assist them for daily activities such as eating and drinking, artificial hands, exoskeletons, robotic arms, smart wheelchairs and other assistive robots have emerged (Wu et al., 2018; Kaur, 2021). How to establish a natural, efficient and stable human-computer interface (HCI) has become a difficult and hot point in the research of interactive control of rehabilitation aids (Mussa-Ivaldi et al., 2013; Venkatakrishnan et al., 2014; Gordleeva et al., 2020; Xiong et al., 2021).

Traditional HCI methods such as those based on buttons, joysticks, or touch screen are usually no longer applicable due to lack of limb function in the above-mentioned situations. In order to solve these problems and optimize the HCI of rehabilitation and assistive machines, many researchers have begun to study HCI based on human physiological signals such as electroencephalogram (EEG), surface electromyography (sEMG), electrooculography (EOG), and so on (Shin et al., 2017; Ding et al., 2019; Zhang et al., 2019; Gordleeva et al., 2020; Li et al., 2020). Compared with lower recognition accuracy or need additional stimulation for EEG-based HCI (such as motor imagery and steady state visual evoked potential) (Lin et al., 2016; Chu et al., 2018) and relative fewer recognizable intentions for EOG-based HCI (Bastos-Filho et al., 2014; He and Li, 2017) or hybrid gaze-brain machine interface (Li et al., 2017; Krausz et al., 2020; Zeng et al., 2020), EMG-based HCI has been widely used in the field of neurorehabilitation with the advantage of higher accuracy and stability, especially for decoding motor intentions of the limb with EMG (Ding et al., 2015; Hussain et al., 2016; Zhang et al., 2019). However, intent recognition based on limb EMG is still facing a huge challenge due to the abnormal signal in the absence of limb function (Jaramillo-Yánez et al., 2020; Xiong et al., 2021). Hence, instead of limb EMG, a novel intention recognition method based on facial electromyography (fEMG) and the HCI based on fEMG have been paid attention and partly researched (Hamedi et al., 2011; Tamura et al., 2012; Bastos-Filho et al., 2014; Nam et al., 2014; Inzelberg et al., 2018; Kapur et al., 2018, 2020).

There are many muscles on the human face, which can control different parts of the face to produce many different movements or expressions, such as eyebrows, eyes, lips, teeth, and so on. Thus, rich information can be decoded from fEMG signals (Hamedi et al., 2013; Inzelberg et al., 2018). Hamedi et al. (2011) recognize movement intentions from fEMG, and a total of 11 facial movements were recognized through electrodes attached to the forehead, with an accuracy rate of over 90%. In their work, a multipurpose interface was suggested that can support 2–11 control commands that could be applied to various HMI systems. Kapur et al. (2020, 2018) developed a portable and wearable device to collect EMG signals around the mouth and neck. More than 10 speech or silent voice commands were recognized from the collected EMG signal. Lu and Zhou (2019) used three electrodes to collect fEMG around the mouth to recognize five movements and used them to control the cursor on the computer to complete functions such as drawing and typing. Cler and Stepp (2015) developed a system using fEMG typing, and the system's typing ITR reached 105.1 bits/min. Nam et al. (2014) integrated multiple signals such as EOG and fEMG to control a humanoid robot. Zhang et al. (2020) controlled a two-degree-of-freedom (2-DOF) prosthesis based on fEMG. Bastos-Filho et al. (2014) and Tamura et al. (2012) used fEMG to control the movements of a wheelchair.

Although there have been some studies on the use of fEMG to recognize intentions to realize HCI, there are still many unsolved problems in this field. First of all, most of the current researches are only based on an organ of the human face, such as the actions of the mouth only or the movements of the eyes only (Hamedi et al., 2011; Lu and Zhou, 2019). Therefore, the performance of fusion of forehead, eyes, mouth, and other parts needs further research. What's more, the facial muscles are neither intertwined like the muscles that control the fingers of the forearm, nor are they independent of each other like the muscles of the upper arm and thigh. There are few research on which features and models are suitable for fEMG classification and recognition. In addition, most of the existing researches only control cursors or mobile robots (Tamura et al., 2012; Bastos-Filho et al., 2014; Nam et al., 2014; Cler and Stepp, 2015), and there are few researches using fEMG to control interactive device with human such as robotic arms. In particular, there is a lack of experiments to control the robotic arm to assist users in completing daily tasks such as eating or drinking based on fEMG. What's more, most of the existing studies only have visual feedback, and lack other feedback methods such as auditory feedback.

In order to solve these problems, we conducted detailed research and experiments. In this study, a complete face-computer interface (FCI) framework based on fEMG including offline analysis and online control of mechanical equipments was proposed. This is a limbs-free method, thus patients can use it to control prostheses, exoskeletons, robotic arms, and computers to take care of themselves and communicate with the world. Healthy people can also use it as a third way of interaction, for example, when controlling an intelligent robotic arm, they can use it as a third hand. In our research, six facial movements related to eyebrows, eyes, and mouth were used in FCI. In order to select better models and features for online FCI, 12 models, eight ways of calculating features, and three different feature combination input methods for the model were compared in detail in the offline stage. In the online stage, four well-designed sessions were designed to control a robotic arm to complete drinking water task in three ways (by touch screen, by fEMG with, and without audio feedback) for verification and performance comparison of proposed FCI framework. To our best of knowledge, this is the first study of using fEMG to control a robotic arm to complete a drinking experiment with audiovisual feedback.

Abbreviations: Base: FCI, face-computer interface; fEMG, facial electromyography; HCI, human-computer interface. Facial movements: LEb, lift eyebrows; LEBO, left eye blink once; REBO, right eye blink once; Bk, bick; TML, tilt mouth to left; TMR, tilt mouth to right. Models: LR, logistic regression; NB, Naive Bayes; DT, decision tree; SVM, support vector machines with Linear Kernel; MLP, multilayer perceptron; Ridge, ridge classifier; RF, random forest; QDA, quadratic discriminant analysis; Ada, Ada boost; GBC, gradient boosting classifier; LDA, linear discriminant analysis; LGBM, light gradient boosting machine Features: MAV, mean absolute value; RMS, root mean square; MC, mean changes; MAC, mean absolute changes; MAX, maximum value; ZC, zero crossings; VAR, variance; ARC, auto regression coefficient. Feature combinations: SF, single-feature; AF, all-features. Online control: VB, control by virtual buttons; E-AF, control by fEMG with audio feedback; E-NAF, Control by fEMG without audio feedback.

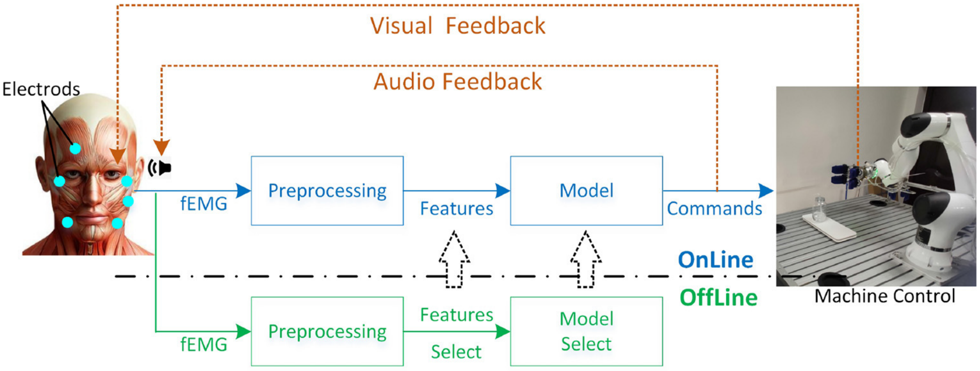


FIGURE 1 | The frame of the face-computer interface (FCI).

In summary, the main contributions and highlights of this research are as follows: (1) A complete Face-Computer Interface (FCI) framework based on fEMG has been proposed. At the same time, the effectiveness of FCI has been proven through well-designed experiments. (2) The performance of frames with and without audio feedback was compared. (3) Multiple models and features were carefully compared and analyzed to select the best model and features suitable for FCI.

The rest of this article would introduces the concept of the FCI framework first, and then six facial movements, eight features, 12 models and three different feature combination methods for model inputing were explained immediately after. The offline data acquisition with its analysis and four online experiment sessions to complete drinking water task based on different ways were written in detail after that. Detailed experimental results and discussion were introduced at the end.

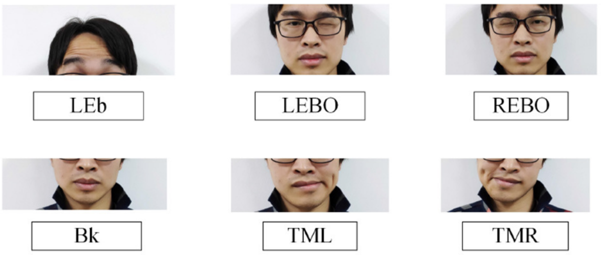


FIGURE 2 | Illustration of facial movements. From left to right and from top to bottom are Lift Eyebrows (LEb), Left Eye Blink Once (LEBO) slowly, Right Eye Blink Once (REBO) slowly, Bick (Bk), Tilt Mouth to Left (TML), and Tilt Mouth to Right (TMR).

2. METHODS

2.1. Frame of FCI

As shown in **Figure 1**, the purpose of the FCI framework is to use fEMG for online control of specific machines such as exoskeleton, prosthetic hand, robotic arm, computer application, and so on. The fEMG is acquired by some electrodes attached on human's face. Then fEMG signals will go through offline preprocessing, feature extraction, and model selection to determine the appropriate features and model for online control stage. After a suitable online model is selected, fEMG will undergo the same filtering and other preprocessing as when offline. The commands or signals output by the model are used to control the machine which is a robotic arm with a soft grip in our research. As usual, subjects can watch the movement of the controlled device in real time to provide visual feedback. In addition, we have also added a voice broadcast of the output commands of the model to provide more feedback information to the users.

2.2. Facial Movements

For the purpose to study FCI based on fEMG in this paper, almost the entire facial areas are involved as the research object, including the forehead area related to eyebrows movements, the area around the eyes related to eyelid actions, and the area related to mouth movements. In order to make the selected actions can be performed by most participants, six movements selected elaborately are used for research. As shown in **Figure 2**, those movements include Lift Eyebrows (LEb), Left Eye Blink Once (LEBO) slowly, Right Eye Blink Once (REBO) slowly, Bick (Bk), Tilt Mouth to Left (TML), Tilt Mouth to Right (TMR), and REST of course. The LEb requires participants to raise the left and right eyebrows at the same time. The Bk is like simulating stationary chewing a hard food. Subjects need to consciously blink the corresponding eye slowly (close corresponding eye for more than hundreds of milliseconds) when doing the LEBO or the REBO movements. Participants can choose to shift only the corner of the mouth to the left or both the corner of the mouth and the mandible to the left according to their personal habits when performing the TML. Actually, the requirement for performing the TMR is same as the TML. And the REST requires participants relaxation and doing nothing.

2.3. Features and Models

In order to determine a more suitable model and features for on-line control machines in FCI frame, 12 models and eight

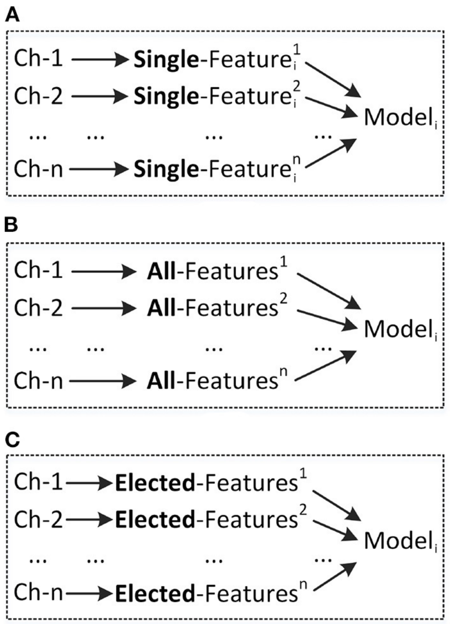


FIGURE 3 | Different combinations of channels, features, and models. **(A)** Using single feature (SF) as model input. **(B)** Using all features (AF) as model input. **(C)** Using elected features (EF) as model input.

types of feature were compared in detail, respectively. The 12 models are {Logistic Regression (LR), Naive Bayes (NB), Decision Tree (DT), Support Vector Machines (SVM) with Linear Kernel, Multilayer Perceptron (MLP), Ridge Classifier (Ridge), Random Forest (RF), Quadratic Discriminant Analysis (QDA), Ada Boost (Ada), Gradient Boosting Classifier (GBC), Linear Discriminant Analysis (LDA), Light Gradient Boosting Machine (LGBM)} (VanderPlas, 2016; Jaramillo-Yánez et al., 2020). Furthermore, as shown in **Figure 3**, three different feature combination methods for model inputing were analyzed in detail, which are Single-Feature (SF) per channel, All-Features (AF), Elected-Features (EF). Notably, the features in EF were selected according to the order of performance of SF.

Assume that x_i is the i-th point in an EMG signal with N points. The calculation method of each feature is as follows (Roberto and Dario, 2016; Jaramillo-Yánez et al., 2020).

(1) Mean absolute value, MAV

$$MAV = \frac{1}{N}\sum_{i=1}^{N}|x_i|$$

(2) Root mean square, RMS

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}$$

(3) Mean changes, MC

$$MC = \frac{1}{N}\sum_{i=1}^{N-1}(x_{i+1} - x_i)$$

(4) Mean absolute changes, MAC

$$MAC = \frac{1}{N}\sum_{i=1}^{N-1}|x_{i+1} - x_i|$$

(5) Maximum value, MAX

$$MAX = max(x_i), i = 1, 2, ..., N$$

(6) Zero crossings, ZC

$$ZC = \sum_{i=1}^{N-1} sgn(-x_i x_{i+1})$$

Where the $sgn(\zeta)$ represents the sign of ζ

$$sgn(\zeta) = \begin{cases} 1 & \zeta > 0 \\ 0 & other \end{cases}$$

(7) Variance, VAR

$$VAR = \frac{1}{N-1}\sum_{i=1}^{N-1}(x_i - \bar{x})^2$$

(8) Auto regression coefficient, ARC

$$x_i = \sum_{k=1}^{p} a_k x_{i-k} + e_i$$

Where the coefficients a_k are the features. The p is the order of ARC which is three in this study.

3. EXPERIMENT

3.1. Subjects and Devices

A total of seven healthy subjects (2 females, 26.2 ± 5.2 years old) participated in all experiments in this study. And it was reviewed and approved by the Ethical Committee of the Shenyang Institute of Automation. Before starting all processes such as data collection and robotic arm control, all subjects were informed in detail of all experimental procedures and possible dangers, and signed an informed consent form.

Four main types of hardware devices [a fEMG collector, a robotic arm, a personal computer (PC), and a high-speed router] were used in this study. The NeusenW64 (Neuracle Co., Ltd, China) with a maximum of 64 unipolar channels (or 32 channels of bipolar electrodes) and maximum of 2,000 Hz sampling rate was used for the acquisition of fEMG. The fEMG collector used

TABLE 1 | Correspondence of electrode numbers, positions, muscles, movements, and commands.

Numbers	Positions	Muscles	Movements	Commands
1	Forehead	Frontal muscle	LEb	Positive direction
2	Left eye corner	Left orbicularis oculi	LEBO	Y axis (forward or back)
3	Right eye corner	Right orbicularis oculi	REBO	X axis (left or right)
4	Masseter skin	Masseter	Bk	Negative direction
5	Left corner of mouth	Left risorius	TML	Gripper (open or close)
6	Right corner of mouth	Right risorius	TMR	Z axis (up or down)

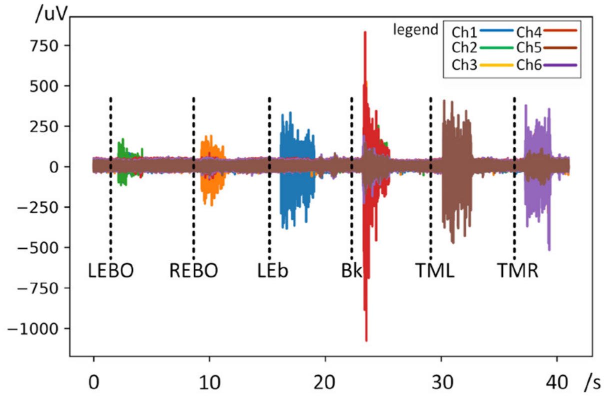


FIGURE 4 | Waveform diagram of a round in data acquisition. The black dotted lines indicate the action prompt.

wifi to transmit data to the PC in real time through a high-speed wireless router. The PC containing with 64-bit Windows-10 system and Python3.7 programming environment was used for all data acquisition, data analysis, intent recognition model training, and control of the robotic arm (Elfin, Han's Robot Co., Ltd, China) with a soft gripper.

All fEMG acquisition processes used 1,000 Hz sampling rate to acquire signals. After the original signal was acquired, the signals used for further analysis were preprocessed by IIR notch filtering to remove power frequency interference and 10–450 Hz second-order bandpass Butterworth filtering.

3.2. fEMG Acquisition for Offline Analysis

3.2.1. Electrodes Configuration

In order to collect the fEMG signals corresponding to different actions, six monopolar electrodes were attached to different parts of the human face. The labels, positions, and corresponding muscles of all electrodes are shown in the **Table 1**. The No. 1 electrode was placed on forehead for acquainting fEMG signals when LEb being performed. Electrodes No. 2 and No. 3 were placed on the extended corners of the left and right eyes, respectively to collect the fEMG signals generated by the actions of LEBO and REBO. When the subjects clenched their teeth, the position of the masseter muscle would be obviously raised, and the No. 4 electrode was placed on the corresponding raised part of the left face. Electrodes No. 5 and No. 6 were placed on the left and right corners of the mouth to collect the fEMG signals generated by the actions of TML and TMR, respectively. In order to balance the electrical field of the left face and the right face, two short-circuited reference electrodes were placed on the mastoid behind the left and right ears. The ground electrode was placed behind the right ear and next to the reference electrode. All electrodes are ordinary disposable electrodes with conductive paste. Before applying the electrodes, the subject's face was wiped with alcohol and waited for it to dry. Each numbered electrode was placed according to the structural features of the human face, such as the corners of the eyes, the edge of the mandible, the corners of the mouth, etc., to prevent large positional deviations from the collection of different sessions.

3.2.2. Acquisition Paradigm

For the six actions {LEb, LEBO, REBO, Bk, TML, TMR} defined in **Figure 2**, each subject participated in 25 rounds of data collection, and each action was collected once in a round. In order to familiarize the subjects with the collection process and maintain the consistency of their actions, the first 5 rounds were used as adaptive training. The next 20 rounds were for the formal collection of fEMG data. In the entire collection process, each action was executed 25 times in total, of which the fEMG signals generated by the last 20 actions were regarded as valid signals. After the first five rounds of adaptive training were completed, the subjects had 3 min of rest and then started the formal collection. During each round of collection, the order of appearance of the six actions was random. There was a 5 s rest between the two actions, and each action lasted about 3 s. Before a action was executed, the name prompt and voice prompt of the action were given by screen and a speaker at the same time. Participants rested for 1 min after each round. After the formal collection of 10 rounds, the subjects rested for 5 min and then performed the next 10 rounds of collection. The entire collection time lasted ∼1 h.

3.2.3. Offline Data Analysis and Processing

Figure 4 shows a demo of a round of waveform in data acquisition and **Figure 5** shows a demo of waveform comparison between different actions and channels. It can be seen from the waveform that the fEMG signal corresponding to each action has an obvious difference. The signal amplitudes from LEBO and REBO are obviously smaller than those of the other actions, while the amplitude from Bk is the largest, followed by LEb. Thus, we cannot simply identify the action category from the

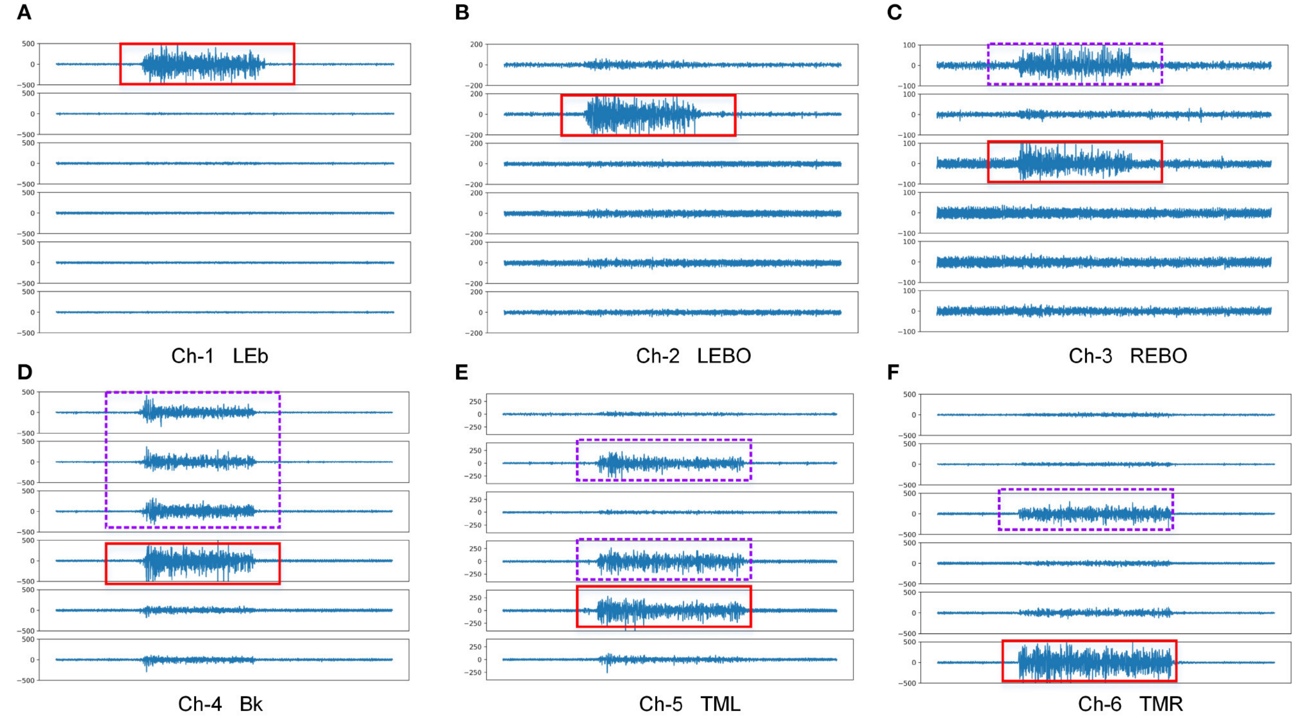


FIGURE 5 | Waveform comparison between different actions and channels. In the figure, the waveform in the solid red frame is the main channel signal corresponding to the action, and the waveform in the purple dashed frame is redundant signal collected by other channels. **(A)** Ch-1 LEb, **(B)** Ch-2 LEBO, **(C)** Ch-3 REBO, **(D)** Ch-4 Bk, **(E)** Ch-5 TML, **(F)** Ch-6 TMR.

amplitude of each signal because of the mutual influence between the channels as shown in **Figure 5**. In the **Figure 5**, the waveform in the solid red frame is the main channel signal corresponding to each action, and the waveform in the purple dashed frame is redundant signal collected by other channels. The action Bk has an effect on almost all channels, while the LEb has almost no effect on other channels. Almost all channels have an influence from actions that are not their counterparts. Ch-1 is affected by REBO because it is attached to the above of the right eye, and of course it is also affected by Bk. Ch-2 is affected by Bk and TML and similarly Ch-3 is affected by Bk and TMR. Ch-4 is affected by TML for Ch-4 is attached at left face. Ch-5 and Ch-6 are the least affected, but are also partially affected by Bk.

Since the intent could not be easily identified from fEMG signal, the features and models mentioned in the third subsection of the second section were used to identify the action category and were analyzed in detail at offline. A 200 ms sliding window with a 50 ms sliding interval containing fEMG signals was used to calculate each feature. For each channel of each action, the signal between 1,500 and 2,650 ms after the prompt was divided into 20 samples. Similarly, the 350 ms signal from 350 ms before each prompt to the prompt moment was divided into four samples. In this way, for a round of collecting 6 actions, there were 20 samples for each movement and 24 samples for REST. To sum up, there were actually 2,880 samples for each participant (2,880 = 20 rounds $*$ 20 samples of each round $*$ 6 actions + 20 rounds $*$ 24 REST samples of each round). Five-fold cross-validation was used to analyze the performance of each model for each subject. For each cross-validation, 80% of the data (2,304 samples = 6 actions * 320 samples for each action + 384 samples for REST) was used for training, and the remaining 20%(576 samples) was used for testing. In other words, 320 samples of each action were used for training in each cross-validation, and the remaining 80 samples were used to test the performance of trained model.

3.3. Control Robotic Arm Online

In order to simulate the scenario where FCI is used in the environment of lack of limb function, a water drinking task was carefully designed. Participants were required to complete the task of using fEMG to control the robotic arm to drink water for themselves during the online control period. A six-degree-of-freedom robotic arm was controlled in this phase and **Figure 6** shows the details of this scene. A soft gripper mounted on the end of the robotic arm was used to grab the object as shown in **Figures 6A,B**. The soft gripper was driven by a dynamic driving manner so as not to damage the object while grasping the object. A drinking glass with a diameter of 7.5 cm and a height of 13 cm with a straw was used for the drinking experiment.

In the experiment, the robotic arm can be controlled in two ways. One of them is to control through the teach pendant which has a touch screen. On the touch screen of the teach pendant, when the robot arm is controlled in manual control mode, a virtual button interface that can control the motion of the robot arm is presented as shown in **Figure 6C**. One can control the

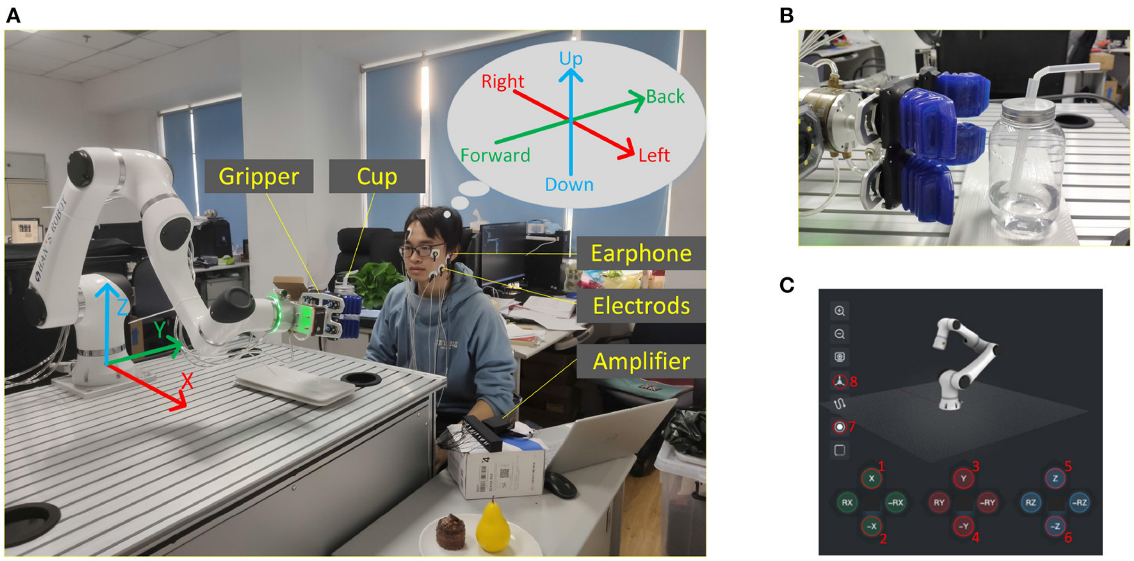


FIGURE 6 | Online control of robotic arm through fEMG. **(A)** A snapshot of the online control of the robotic arm to drink water. The robotic arm body and its coordinate system, the soft gripper, fEMG electrodes, earphone for audio feedback, amplifier, PC, etc. are marked. **(B)** A photo of the soft gripper before grabbing the cup. **(C)** Virtual button interface for manipulating the robotic arm.

movements of the robotic arm corresponding to the function of virtual button by pressing and holding the virtual button on the touch screen. Another way to control the robotic arm is through the FCI method proposed in this paper.

During the whole experiment, the robotic arm works under the position control mode of Cartesian space. The three-dimensional orthogonal Cartesian coordinate system is shown in **Figure 6**. In this research, the external application sends eight commands to the robotic arm through the API of the robotic arm system. These eight commands are {moving to the positive of X-axis, moving to the negative of X-axis, moving to the positive of Y-axis, moving to the negative of Y-axis, moving to the positive of Z-axis, moving to the negative of Z-axis, closing gripper, and opening gripper}. These eight commands correspond one-to-one with the virtual buttons labeled 1–8 in **Figure 6C**. When the robotic arm is working, the operator only needs to pay attention to the direction of movement of the claw at the end of the manipulator, and does not need to pay attention to the joint space of the manipulator. The mapping from Cartesian space to joint space is done by the manipulator API. For the safety of operation, the end of the robot arm runs at a lower speed of 3 cm/s during the movement. In this experiment, the working space of the gripper is limited to a cuboid space with a range of [20, 120] cm of X-axis, [−70, 100] cm of Y-axis, and [5, 120] cm of Z-axis. The projection of the working space at the end of the robotic arm on the X-Y plane is restricted to not exceed the desktop as shown in **Figure 6A**, except for the side parallel to X-axis at positive direction of Y-axis, which can exceed 30 cm so that the water cup can be sent to the subject's mouth.

As the subjects were naive for controlling robotic arm, the experiment was divided into four sessions. Subjects participated in different sessions for 4 consecutive days, and each session took from 1 to 2 h. The first session was for participants to familiarize themselves with the robotic arm and its control process. In the second session, the participants used virtual buttons to operate the robotic arm to complete the task of drinking water. In the third and fourth sessions, subjects used FCI to complete the task of drinking water with and without audio feedback, respectively.

3.3.1. Session #1-Familiar With the Robotic Arm

In this session, participants' goal was to understand the movement of the end of the manipulator in the robotic arm workspace and then use virtual buttons to control the gripper of the robotic arm for single-axis movement. A professional robotic arm engineer explained to each participant the working space of the robotic arm and the uniaxial movement of the end of the robotic arm in the Cartesian coordinate system. In this process, participants did not need to understand the working principle of the robotic arm such as joint angle, joint space, kinematics, etc., but only need to know that the gripper can move along the three axes in Cartesian space. They even did not need to know the concept of Cartesian space. The instructor explained to them as follows: "The gripper will move forward when this button (−Y) is pressed. And the gripper will move to the left when this button (+X) is pressed. After that button (Close) is pressed, the gripper will close..." After understanding the operating mode of the end of the robotic arm, the subjects used eight virtual buttons on the touch screen to control the direction, opening or closing of the gripper. Each participant was asked to run every control command 10–15 times in order to become familiar with the movement of the gripper. For each subject, this session lasts about 1 h.

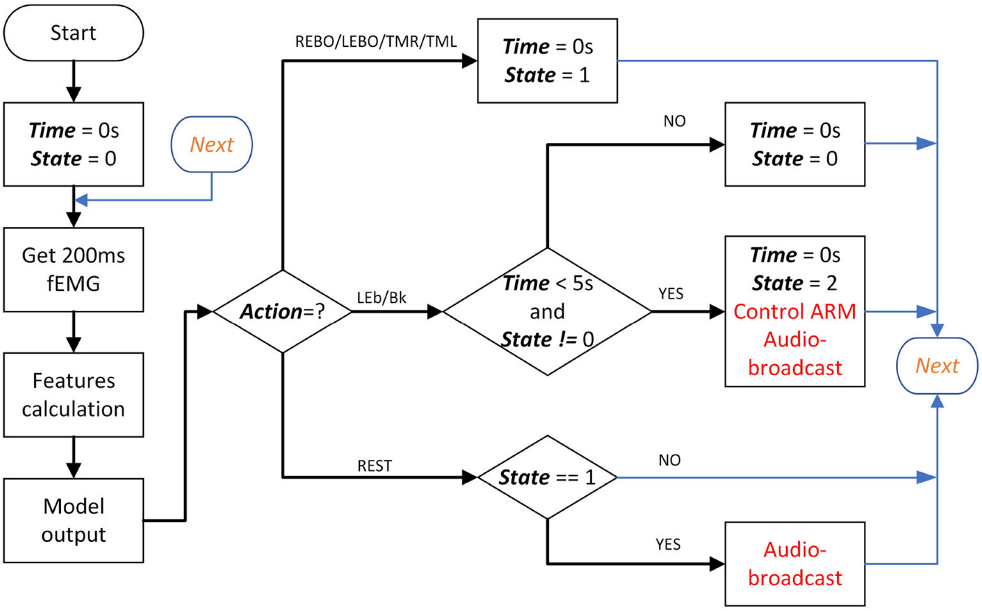


FIGURE 7 | Online control flow chart base on FCI. In the figure, the black, bold, and italicized words are the main variables, and the red part is the robotic arm control and voice broadcast.

3.3.2. Session #2-Complete Drinking Task Based on Virtual Buttons

In this session, subjects were asked to use virtual buttons to control the movement of gripper to complete drinking task 10 times. In each drinking task, as shown in **Figure 6**, the cup was placed on a 30∗15 cm rectangular saucer. The center of the saucer was fixed at the position (x: 90, y: 45, z: 0) cm of the robotic arm coordinate system. In order to ensure that the subjects were controlled based on their actual location rather than remembering repeated paths, in each task, the position of the cup on the saucer and the initial position of the gripper were random. At the same time, in order to eliminate the difference in experimental performance caused by randomness, the random range of the position of the cup and the gripper was restricted rather than completely random, and each participant conducts at least 10 experiments in each session. The initial position of the gripper was (x: x_i, y: -45, z: z_i)cm, $i = 1, 2, 3, ..., 10$, in each drinking task. The x_i was randomly generated from [20, 120] cm and z_i was randomly generated from [5, 120] cm. The subject sat in the direction of the positive Y-axis and was asked to use virtual buttons to control the gripper to complete the task of drinking water. A water drinking task can be divided into three stages: (1) the gripper was moved from a random initial position to the place where the water cup was placed; (2) the water cup was picked up and moved to the subject's mouth; (3) the subject was asked to drink a small amount of water and then controlled the robotic arm to place the water cup back to the saucer. The time spent on each task was recorded when the gripper started to move, and stopped when the water cup was put back on the saucer. The subjects needed to complete 10 repeated drinking tasks during this session and there was a 2 min rest period between each task, during which the experiment assistant would reset the water cup and the gripper to the initial position.

3.3.3. Session #3 and #4-Complete Drinking Task Based on FCI

In this two sessions, participants used fEMG instead of virtual buttons to control gripper to complete the task. Except for the change of the command input interface, the other settings remained the same as in the second session. The models and features selected during the offline phase were used here. Before this two sessions started, in order to eliminate the difference between the different sessions, the subjects were asked to collect five additional rounds of fEMG data to update the model with the initial parameters trained in the offline phase. When fEMG was used as a control method, six facial movements were mapped into eight commands corresponding to eight gripper movements. As shown in **Table 1** and **Figure 7**, each manipulator motion control goes through two stages. The first stage is to use four actions to select one of the four axes (3 coordinate axes and hand grip to open or close). Then, in the second stage, the remaining two actions will be used to select the direction of movement. For the direction of movement of the claw, the first action selected which axis to move along, and the next second action selected to move in the positive or negative direction of that axis. The first actions {REBO, LEBO, TMR}, respectively indicate the selected X, Y, or Z axis. LEb represents the positive direction of the corresponding axis and Bk represents the negative direction. The first action only needs a short duration of about 1 s to ensure that the model can be recognized, but LEb or Bk needs to hold on when the gripper is running. When the action stops, the gripper also stops moving accordingly. For the opening or closing of the gripper, the first

TABLE 2 | Broadcast contents in audio feedback after the first action or the second action is recognized.

First movement	Broadcast content	Second movement	Broadcast content
REBO	Left or right	LEb	Left
		Bk	Right
LEBO	Forward or back	LEb	Back
		Bk	Forward
TMR	Up or down	LEb	Up
		Bk	Down
TML	Gripper	LEb	Close
		Bk	Open

facial movement must be TML. When the second action is Bk, it means closing the gripper and the opposite LEb means opening the gripper. When controlling the opening and closing of the gripper, both the first action and the second action only need to last for a short time. After completing the command, the open and closed state of the gripper will be self-locking. In order to prevent the water cup from being erroneously released at high altitude, the gripper opening and closing control is only valid when the gripper is directly above the desktop and the coordinate value of the z-axis is <10 cm. Throughout whole process, the second action {LEb or Bk} is valid only if it is started within 5 s after the first action is successfully recognized or the last non-resting action is itself, otherwise it is invalid and ignored. In **Figure 7**, when the variable ***State*** is 0, it means that it is currently in the first stage; when it is 1, it means that it has entered the second stage from the first stage; and when it is 2, it indicates the state where the second stage continues.

In the third and fourth sessions, the subjects were required to complete drinking task 10 times as in the second part. In these two sessions, the subjects had three attempts to familiarize themselves with the commands of fEMG before starting the task. The difference between the third session and the fourth session is that the third session has a earphone to provide audio feedback to the subjects, while the fourth session has no audio feedback. The audio feedback in the third session mainly announces the intention of the participants after the action recognition. Participants can confirm whether the recognition results are consistent with their own intentions according to the broadcast content in order to adjust the control strategy in time. The content of the broadcast is the upcoming or ongoing movement of the gripper, as shown in **Table 2**, and he corresponding intent direction is shown in **Figure 6A**.

4. RESULTS

4.1. Offline Performance

Figure 8 shows the average accuracy of all subjects in different models with single-feature (SF). It can be seen from the result that the maximum accuracy of classifiers {LR, NB, SVM, MLP, RF, QDA, Ada, GBC, LGBM} exceeds 90% when using SF as input. Features {MAV, RMS, MAC, MAX, VAR} performed well on

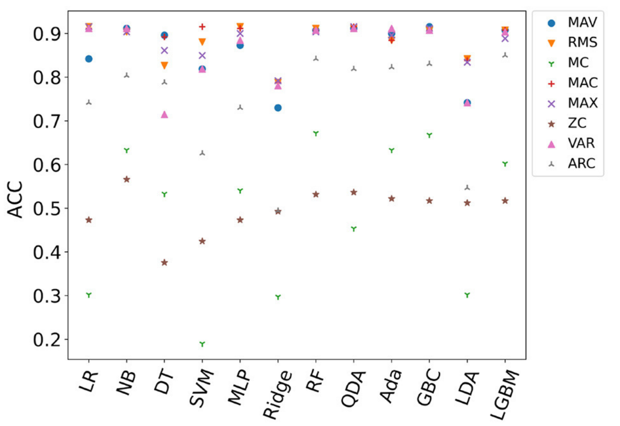


FIGURE 8 | Comparison of the average accuracy of all subjects in different models with single-feature (SF).

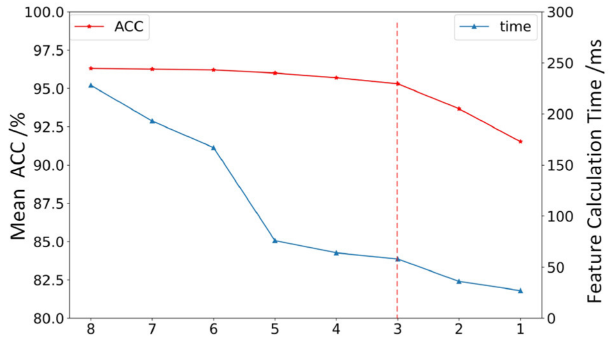


FIGURE 9 | The features calculation time and the average accuracy of the GBC model under feature reduction process. The abscissa represents the performance when the eight features are removed one by one in the order of {MC, ZC, ARC, WAV, MAX, VAR, RMS, MAC}.

multiple classifiers. The features used in the online phase require less time for feature calculation and high recognition accuracy. In order to select the feature set that meets this condition, the average accuracy of all models when a single feature is used as the model input is sorted as {MC, ZC, ARC, WAV, MAX, VAR, RMS, MAC} from low to high. **Figure 9** shows the calculation time and the average accuracy of the GBC model under feature reduction process when the eight features are removed one by one in the order of {MC, ZC, ARC, WAV, MAX, VAR, RMS, MAC}. It can be seen from the figure that as the features decrease, the accuracy and calculation time are also decreasing. However, the degree of accuracy decrease in the early stage is small, and the decrease rate increases in the later stage, and its inflection point appears when there are only three features left and the calculation time is already small enough at this time. Therefore, features {VAR, RMS, MAC} are used as EF here.

Comparison of the average accuracy of all subjects between SF, AF, and EF is shown in **Figure 10**. For SF, the maximum accuracy rate of each model are selected first across different single-feature

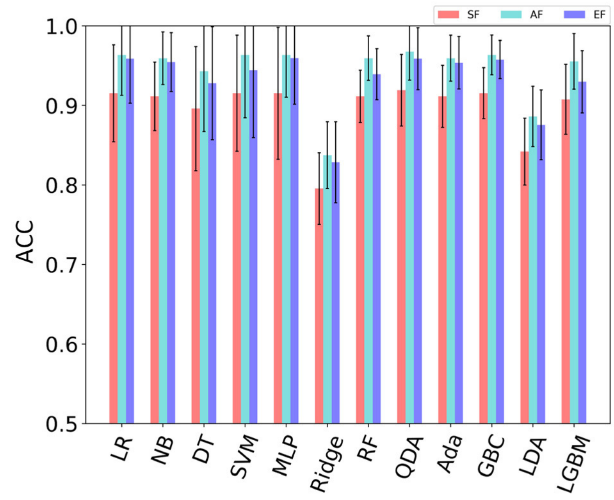


FIGURE 10 | Comparison of the average accuracy of all subjects in different combinations of features. The result is the average of all subjects, and the standard deviation is indicated in the figure. SF: The maximum accuracy rate of each model are selected first across different single-feature (SF) for per single subject, and then the average value of the selected maximum is calculated across all subjects. AF: Performance when using all features. EF: Elected features, here are RMS, MAC, and VAR.

(SF) for per single subject, and then the average value of the selected maximum is calculated across all subjects. The *t*-test was used to analyze the difference between SF and AF, and all *p*-values were <0.01 for all models, which indicates that the performance of SF and AF is significantly different. But there is no statistically significant difference between EF and AF. In other words, the difference between EF and AF is not obvious. When using AF as the model input, there are nine models {LR, NB, SVM, MLP, RF, QDA, Ada, GBC, LGBM } with an average accuracy of more than 95% on all subjects. Among them, the accuracy of 6 models {LR, NB, MLP, QDA, Ada, GBC, LGBM} still exceeds 95% when using EF {RMS, MAC, VAR}. When using AF, the accuracy of QDA is the largest among all models, which is 96.7%. The next one comes from GBC, which is 96.3%. When using EF, the accuracy of MLP is the largest among all models, which is 95.9%. The next one is also from GBC, which is 95.3%. It can be seen that the difference between the best performance when using AF and EF is small, not more than one percentage point. At the same time, the smallest standard deviation of the accuracy comes from GBC in all models. This indicates that GBC has the most stable performance. As a result, GBC with EF input was selected for online testing. When using the EF features{RMS, MAC, VAR} as the input of the GBC model, the average value of the five cross-validation for each participant is shown in **Figure 11D**. Among them, the accuracy of subject S6 was the highest with 98.8%.

4.2. Online Control Robotic Arm

The performance of different subjects using different methods to control the robotic arm to complete the task of drinking water online is shown in **Figure 11**. The time spent by seven subjects using virtual buttons to control the robotic arm to complete the task ranges from 1.79 to 6.62 min. Where the mean is 3.26 min, and the median is 3.02 min. The ranges of the time of using FCI with audio and without feedback are [1.93, 7.9] and [2.18, 9.55] min, respectively. Where the means are 4.5 and 5.72 min, and the medians are 4.13 and 5.60 min. In terms of mean and variance, the performance using virtual buttons as the control method is better than fEMG control with audio feedback, and the performance without audio feedback is the worst. For these three cases, the pairwise permutation test was used to analyze their statistical significance. Ten thousand permutation tests were carried out for each condition of each subject, and the *p*-value is shown in **Table 3**. As can be seen from the table, for the control by virtual buttons or by fEMG with audio feedback, the significance of the difference between each subject is quite different. S6 has no statistical difference, S1, S2, and S3 have a certain difference, while the difference of S4, S5, and S7 is very significant. Therefore, statistically speaking, there are some differences between virtual buttons and fEMG with audio feedback, but the overall difference is not very significant. For the control by fEMG with and without audio feedback, the *p*-values of all subjects are <0.001, indicating that the difference between this two cases is very statistically significant. The average time of each participant in the 10 tasks is marked in **Figure 11D**. It can be seen that there is no order of magnitude difference in the time consumed by virtual buttons and fEMG with audio feedback. The average gap between the two is 1.24 min. The biggest gap comes from the subject S7, which is 2.1 min, and the smallest difference is only 0.37 min which is from S6.

5. DISCUSSION

This study performed offline analysis and online experiment respectively based on the proposed FCI. In the offline stage, 12 models and eight ways of calculating features were compared in detail. A total of seven participants performed 25 rounds of fEMG signal acquisition for six facial movements, and each generated 20 rounds of valid signals. It can be seen from the offline analysis results that the selected action has good recognizability under the studied model and features. As shown in **Figures 10**, **11D**, when the EF features {RMS, MAC, VAR} were selected as the model input, the maximum recognition accuracy among the seven participants reached 98.8%, and the minimum reached 91.4%. In the feature calculation methods studied, the time domain features were mainly compared without the frequency domain features. Because the main purpose of comparing features is to select some good features for online use, and these features selected usually require fast and stable calculations. When calculating features related to the frequency domain, the speed is usually slower than that in the time domain. Moreover, it can be seen from the offline and online results that the recognition accuracy and efficiency are sufficient based on EF features {RMS, MAC, VAR} or even a single feature only.

In order to verify the effectiveness of the proposed FCI, an experiment to control a robotic arm was designed. During the online control of the robotic arm, four progressive sessions were

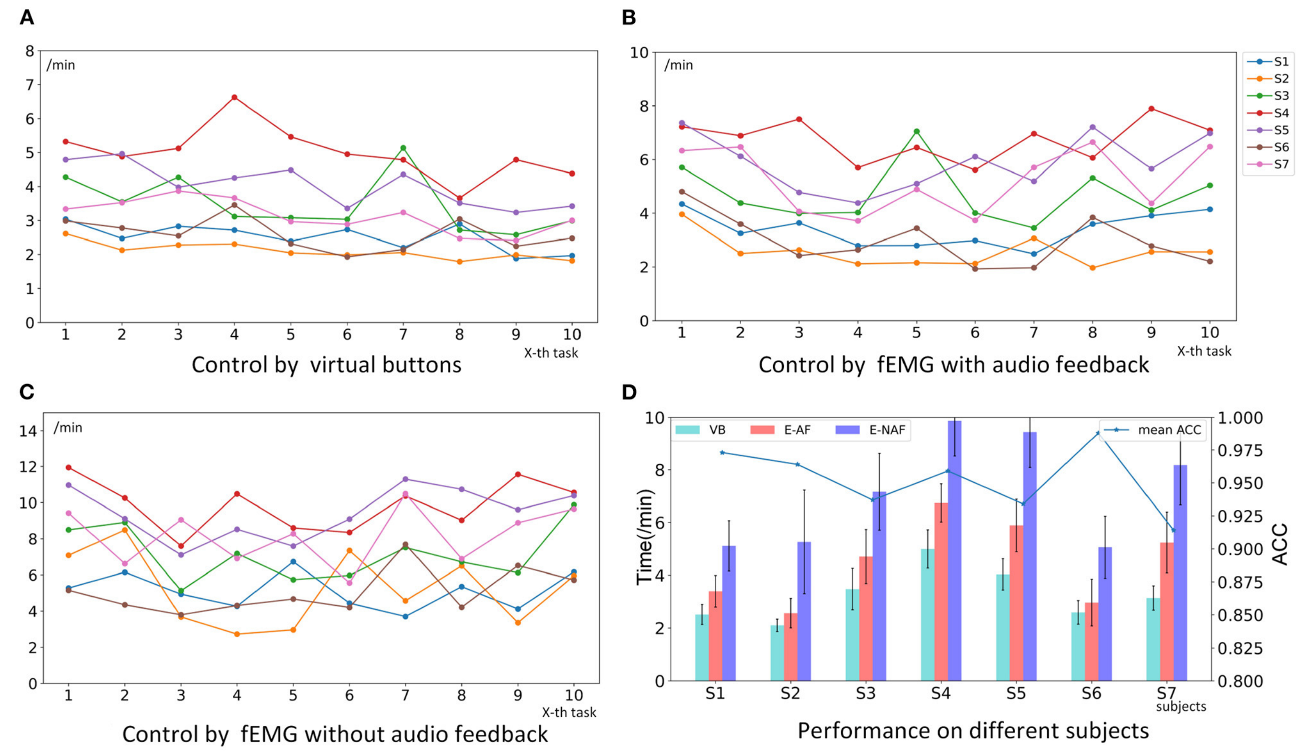


FIGURE 11 | The performance of different subjects using different methods to control the robotic arm to complete the task of drinking water online. **(A)** Control by virtual buttons (VB). **(B)** Control by fEMG with audio feedback (E-AF). **(C)** Control by fEMG without audio feedback (E-NAF). **(D)** Performance on different subjects. The bars represent the average time-consuming of 10 online tasks, and the broken line represents the average recognition accuracy (ACC) of 5 cross-validation using Gradient Boosting Classifier (GBC) with elected features (EF) {RMS, MAC, VAR} as input when offline.

TABLE 3 | In the online phase, the *p*-value of the three-way pairwise permutation test statistical analysis.

Pair	S1	S2	S3	S4	S5	S6	S7
VB & E-AF	0.0015	0.019	0.008	0.0002	0.0004	0.28	<0.0001
VB & E-NAF	All <0.0001						
E-AF & E-NAF	0.0003	0.0005	0.0007	<0.0001	<0.0001	0.0003	0.0004

Among them, VB, control by virtual buttons; E-AF, control by fEMG with audio feedback; E-NAF, control by fEMG without audio feedback.

carefully designed. In the experiment, in order to reduce the burden on the participants and keep FCI close to the daily habits of most people, participants only learned the motion control of the gripper without having to understand the working principle of the robotic arm too much. This way is as natural as people controlling arbitrary objects to move in three-dimensional space in normal life. Subjects first used the virtual buttons on the touch screen to control the gripper to perform the task of drinking water, and then used FCI to control the robotic arm. A total of six facial actions from almost entire face in the online phase were used to map to the eight movements commands of the grippers of end at robotic arm. When the virtual buttons, fEMG with audio feedback and fEMG without audio feedback were used to control the hand grips, the performance on the virtual buttons was the best, which was reasonable and expected. Using the button control method is just a benchmark, or even an upper limit that can be reached at present. This method is only suitable for scenarios where hand functions are available. In the case of lack of limb functions, this method cannot continue to be used. In contrast, the FCI method with relatively poor performance but not very large gap is more suitable for disabled people.

It can be seen from the results of the online stage that the movements recognition accuracy is not the only factor that affects the performance of the experiment. There are many factors that affect the performance of the online control stage. These factors include: (1) the accuracy of intention recognition; (2) whether the feedback information is sufficient, such as whether there is voice feedback; (3) whether the actions of the participants are consistent with their intentions; (4) the participant's reaction speed, such as the time it takes to correct the movement or recognition error; (5) the participant's path planning and control strategy for the movement of the gripper (Nam et al., 2014; Zhang et al., 2020). As shown in **Figure 11**, the accuracy rate from subjects S3, S5, or s2 was lower than that from S4, but in the online process, the former performed better than the latter. S6 had the highest accuracy rate reached 98.8%, but his (/her) time-consuming performance was not the best, slightly inferior to S2. Similarly, the accuracy of S7 was the lowest, but the performance of online control was not the worst. There is a trend that the more time the subjects spend based on virtual buttons, the more time they spend based on FCI. The performance of subjects S1, S2, and S6 under using FCI was even higher than that of subjects S3, S4, and S5 from using virtual buttons. Therefore, if the time subjects take to use the buttons to control the robotic arm to complete the task of drinking water is acceptable, then the time spent using FCI should also be within an acceptable range.

For fEMG-based FCI, the method with audio feedback performed better than that without audio feedback. When there is audio feedback, not only did the participants spend less time to complete the task, but their performance was also more stable. It can be seen from **Figures 11B,C** that, when there is no audio feedback, the time spent on 10 tasks fluctuates more severely. **Figure 11D** also shows the standard deviation of the time it takes on different task trials. It can also be seen that the standard deviation when there is no voice feedback is larger than when there is voice feedback.

FCI based on fEMG has some inherent disadvantages, such as the need to attached electrodes on the subject's face, this may not be accepted by some people. However, this requirement is more suitable than EEG-based HCI that requires EEG caps and conductive paste. In addition, like other HCI based on physiological signals, FCI based on fEMG also requires additional signal acquisition for model training. And fatigue caused by long-term continuous use will reduce performance without a lot of corresponding exercise. Compared with a button-based HCI, FCI based on fEMG also has some advantages. One of the advantages is that the user can keep his eyes on the movement of the object to be controlled without leaving it when using FCI. However, when using HCI that is operated by hand based on buttons or joysticks, the subjects often need to shift their attention to their hands first when switching commands, and the sight of the subjects even jumped back and forth between the hand and the object to be controlled sometimes.

6. CONCLUSION

A complete FCI framework based on fEMG including offline analysis and online control of mechanical equipments was proposed. In the offline stage, 12 models, eight ways of calculating features, and three ways of feature input were studied and compared in detail. The three EF features {RMS, MAC, VAR} and the GBC model with an average offline recognition rate of 95.3%, a maximum of 98.8%, and a minimum of 91.4% were selected for use in online scenarios. Four well-designed sessions were designed for online verification and performance comparison of FCI. In the online phase, seven subjects were required to use virtual buttons, fEMG with and without audio feedback to control the gripper at the end of the robotic arm to complete the drinking experiment. In contrast, the way with audio feedback performed better than the way without audio feedback. There is no order of magnitude difference in the time consumed by virtual buttons and fEMG with audio feedback. The average gap between the two is only 1.24 min, and the smallest difference is only 0.37 min. The effectiveness and applicability of the proposed FCI framework has been proven.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Committee of the Shenyang Institute of Automation. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

BZ and YC carried out experimental design, data collection, analysis, and wrote the first draft of this article. LZhan and LZhao provided anatomical analysis support, corrected the electrode placement position, and designed the main types of movements used. DZ provided technical guidance in online experiments. XZ provided the experimental platform and its management and completed the final proofread. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (U1813214, 61773369, 61903360, 92048302, U20A20197), the Self-planned Project of the State Key Laboratory of Robotics (2020-Z12), and China Postdoctoral Science Foundation funded project (2019M661155).

REFERENCES

Bastos-Filho, T. F., Cheein, F. A., Müller, S. M. T., Celeste, W. C., de la Cruz, C., Cavalieri, D. C., et al. (2014). Towards a new modality-independent interface for a robotic wheelchair. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 567–584. doi: 10.1109/TNSRE.2013.2265237

Chu, Y., Zhao, X., Zou, Y., Xu, W., Han, J., and Zhao, Y. (2018). A decoding scheme for incomplete motor imagery EEG with deep belief network. *Front. Neurosci.* 12:680. doi: 10.3389/fnins.2018.00680

Cler, M. J., and Stepp, C. E. (2015). Discrete versus continuous mapping of facial electromyography for human-machine interface control: performance and training effects. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 572–580. doi: 10.1109/TNSRE.2015.2391054

Ding, Q., Han, J., Zhao, X., and Chen, Y. (2015). Missing-data classification with the extended full-dimensional Gaussian mixture model: applications to EMG-based motion recognition. *IEEE Trans. Indus. Electron*. 62, 4994–5005. doi: 10.1109/TIE.2015.2403797

Ding, Q., Zhao, X., Han, J., Bu, C., and Wu, C. (2019). Adaptive hybrid classifier for myoelectric pattern recognition against the interferences of outlier motion, muscle fatigue, and electrode doffing. *IEEE Trans. Neural Syst. Rehabil. Eng*. 27, 1071–1080. doi: 10.1109/TNSRE.2019.2911316

Gordleeva, S. Y., Lobov, S. A., Grigorev, N. A., Savosenkov, A. O., and Kazantsev, V. B. (2020). Real-time EEG–EMG human-machine interface-based control system for a lower-limb exoskeleton. *IEEE Access* 8, 84070–84081. doi: 10.1109/ACCESS.2020.2991812

Hamedi, M., Salleh, S. H., Astaraki, M., and Noor, A. M. (2013). Emg-based facial gesture recognition through versatile elliptic basis function neural network. *BioMed. Eng. OnLine* 12:73. doi: 10.1186/1475-925X-12-73

Hamedi, M., Salleh, S. H., Tan, T. S., Ismail, K., Ali, J., Dee-Uam, C., et al. (2011). Human facial neural activities and gesture recognition for machine-interfacing applications. *Int. J. Nanomed*. 2011, 3461–3472. doi: 10.2147/IJN.S26619

He, S., and Li, Y. (2017). A single-channel EOG-based speller. *IEEE Trans. Neural Syst. Rehabil. Eng*. 25, 1978–1987. doi: 10.1109/TNSRE.2017.2716109

Hussain, I., Spagnoletti, G., Salvietti, G., and Prattichizzo, D. (2016). An EMG interface for the control of motion and compliance of a supernumerary robotic finger. *Front. Neurorobot*. 10:18. doi: 10.3389/fnbot.2016.00018

Inzelberg, L., Rand, D., Steinberg, S., David-Pur, M., and Hanein, Y. (2018). A wearable high-resolution facial electromyography for long term recordings in freely behaving humans. *Sci. Rep*. 8:2058. doi: 10.1038/s41598-018-20567-y

Jaramillo-Yánez, A., Benalcázar, M., and Mena-Maldonado, E. (2020). Real-time hand gesture recognition using surface electromyography and machine learning: a systematic literature review. *Sensors* 20:2467. doi: 10.3390/s20092467

Kapur, A., Kapur, S., and Maes, P. (2018). "Alterego: a personalized wearable silent speech interface," in *23rd International Conference on Intelligent User Interfaces, IUI '18* (New York, NY: Association for Computing Machinery), 43–53. doi: 10.1145/3172944.3172977

Kapur, A., Sarawgi, U., Wadkins, E., Wu, M., Hollenstein, N., and Maes, P. (2020). "Non-invasive silent speech recognition in multiple sclerosis with dysphonia," in *Proceedings of the Machine Learning for Health NeurIPS Workshop*, eds A. V. Dalca, M. B. McDermott, E. Alsentzer, S. G. Finlayson, M. Oberst, F. Falck, and B. Beaulieu-Jones (PMLR), 25–38. Available online at: http://proceedings.mlr.press/v116/kapur20a.html

Kaur, A. (2021). Wheelchair control for disabled patients using EMG/EOG based human machine interface: a review. *J. Med. Eng. Technol*. 45, 61–74. doi: 10.1080/03091902.2020.1853838

Krausz, N. E., Lamotte, D., Batzianoulis, I., Hargrove, L. J., Micera, S., and Billard, A. (2020). Intent prediction based on biomechanical coordination of EMG and vision-filtered gaze for end-point control of an arm prosthesis. *IEEE Trans. Neural Syst. Rehabil. Eng*. 28, 1471–1480. doi: 10.1109/TNSRE.2020.2992885

Li, S., Zhang, X., and Webb, J. D. (2017). 3-d-gaze-based robotic grasping through mimicking human visuomotor function for people with motion impairments. *IEEE Trans. Biomed. Eng*. 64, 2824–2835. doi: 10.1109/TBME.2017.2677902

Li, Z., Zhang, D., Zhao, X., Wang, F., and Han, J. (2020). A temporally smoothed MLP regression scheme for continuous knee/ankle angles estimation by using multi-channel sEMG. *IEEE Access* 8, 47433–47444. doi: 10.1109/ACCESS.2020.2979008

Lin, K., Cinetto, A., Wang, Y., Chen, X., Gao, S., and Gao, X. (2016). An online hybrid bci system based on SSVEP and EMG. *J. Neural Eng*. 13:026020. doi: 10.1088/1741-2560/13/2/026020

Lu, Z., and Zhou, P. (2019). Hands-free human-computer interface based on facial myoelectric pattern recognition. *Front. Neurol*. 10:444. doi: 10.3389/fneur.2019.00444

Mussa-Ivaldi, F. A., Casadio, M., and Ranganathan, R. (2013). The body-machine interface: a pathway for rehabilitation and assistance in people with movement disorders. *Expert Review of Med. Devices* 10, 145–147. doi: 10.1586/erd.13.3

Nam, Y., Koo, B., Cichocki, A., and Choi, S. (2014). GOM-Face: GKP, EOG, and EMG-Based multimodal interface with application to humanoid robot control. *IEEE Trans. Biomed. Eng*. 61, 453–462. doi: 10.1109/TBME.2013.2280900

Roberto, M., and Dario, F. (2016). "Surface electromyography for man-machine interfacing in rehabilitation technologies," in *Surface Electromyography: Physiology, Engineering, and Applications*, ed T. Samad (Hoboken, NJ: John Wiley & Sons, Inc.), 540–560.

Shin, S., Tafreshi, R., and Langari, R. (2017). "Real-time EMG-based human machine interface using dynamic hand gestures," in *2017 American Control Conference (ACC)*, 5456–5461. doi: 10.23919/ACC.2017.7963803

Tamura, H., Murata, T., Yamashita, Y., Tanno, K., and Fuse, Y. (2012). Development of the electric wheelchair hands-free semi-automatic control system using the surface-electromyogram of facial muscles. *Artif. Life Robot*. 17, 300–305. doi: 10.1007/s10015-012-0060-2

VanderPlas, J. (2016). *Python Data Science Handbook*. Sebastopol, CA: O'Reilly Media.

Venkatakrishnan, A., Francisco, G. E., and Contreras-Vidal, J. L. (2014). Applications of brain-machine interface systems in stroke recovery and rehabilitation. *Curr. Phys. Med. Rehabil. Rep*. 2, 93–105. doi: 10.1007/s40141-014-0051-4

Wu, Q., Wang, X., Chen, B., and Wu, H. (2018). Patient-active control of a powered exoskeleton targeting upper limb rehabilitation training. *Front. Neurol*. 9:817. doi: 10.3389/fneur.2018.00817

Xiong, D., Zhang, D., Zhao, X., and Zhao, Y. (2021). Deep learning for EMG-based human-machine interaction: a review. *IEEE/CAA J. Autom. Sin*. 8, 512–533. doi: 10.1109/JAS.2021.1003865

Zeng, H., Shen, Y., Hu, X., Song, A., and Wen, P. (2020). Semi-autonomous robotic arm reaching with hybrid gaze-brain machine interface. *Front. Neurorobot*. 13:111. doi: 10.3389/fnbot.2019.00111

Zhang, J., Wang, B., Zhang, C., Xiao, Y., and Wang, M. Y. (2019). An EEG/EMG/EOG-based multimodal human-machine interface to real-time control of a soft robot hand. *Front. Neurorobot*. 13:7. doi: 10.3389/fnbot.2019.00007

Zhang, X., Li, R., Li, H., Lu, Z., Hu, Y., and Alhassan, A. B. (2020). Novel approach for electromyography-controlled prostheses based on facial action. *Med. Biol. Eng. Comput*. 58, 2685–2698. doi: 10.1007/s11517-020-02236-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 02 August 2021
doi: 10.3389/fnbot.2021.685961

Multi–Joint Angles Estimation of Forearm Motion Using a Regression Model

Zixuan Qin[1]*, *Sorawit Stapornchaisit*[1], *Zixun He*[1], *Natsue Yoshimura*[2,3] *and Yasuharu Koike*[2]

[1] *Department of Information and Communications Engineering, Tokyo Institute of Technology, Yokohama, Japan,* [2] *Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan,* [3] *Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency (JST), Saitama, Japan*

To improve the life quality of forearm amputees, prosthetic hands with high accuracy, and robustness are necessary. The application of surface electromyography (sEMG) signals to control a prosthetic hand is challenging. In this study, we proposed a time-domain CNN model for the regression prediction of joint angles in three degrees of freedom (3-DOFs, include two wrist joint motion and one finger joint motion), and five-fold cross validation was used to evaluate the correlation coefficient (CC). The CC value results of wrist flexion/extension motion obtained from 10 participants was 0.87–0.92, pronation/supination motion was 0.72–0.95, and hand grip/open motion was 0.75–0.94. We backtracked the fully connected layer weights to create a geometry plot for analyzing the motion pattern to investigate the learning of the proposed model. In order to discuss the daily updateability of the model by transfer learning, we performed a second experiment on five of the participants in another day and conducted transfer learning based on smaller amount of dataset. The CC results improved (wrist flexion/extension was 0.90–0.97, pronation/supination was 0.84–0.96, hand grip/open was 0.85–0.92), suggesting the effectiveness of the transfer learning by incorporating the small amounts of sEMG data acquired in different days. We compared our CNN-based model with four conventional regression models, the result illustrates that proposed model significantly outperforms the four conventional models with and without transfer learning. The offline result suggests the reliability of the proposed model in real-time control in different days, it can be applied for real-time prosthetic control in the future.

Keywords: convolutional neural networks, geometry plot, regression model, surface electromyography, transfer learning

Edited by:
Thilina Dulantha Lalitharatne,
Imperial College London,
United Kingdom

Reviewed by:
Siti Anom Ahmad,
Putra Malaysia University, Malaysia
Yinfeng Fang,
Hangzhou Dianzi University, China

****Correspondence:***
Zixuan Qin
qin.z.aa@m.titech.ac.jp

Received: *26 March 2021*
Accepted: *07 July 2021*
Published: *02 August 2021*

Citation:
Qin Z, Stapornchaisit S, He Z, Yoshimura N and Koike Y (2021) Multi–Joint Angles Estimation of Forearm Motion Using a Regression Model. Front. Neurorobot. 15:685961. doi: 10.3389/fnbot.2021.685961

INTRODUCTION

The human hand plays a crucial role in many activities of daily living (ADL), therefore, loss of the upper extremity could significantly impact functional independence (National Academies of Sciences, 2017). To support the ADL of amputees, prosthetic hands were developed, which can mimic human-hand motions to complete activities. Several studies have investigated the possibility of controlling the prosthetic hand using surface electromyography (sEMG) signals because the use of sEMG signals can directly utilize human neural pathways to restore ADL function (Scheme and Englehart, 2011). For abled people, hand movements occur due to the contraction and relaxation of the forearm muscles controlled by the brain. When the forearm is lost, sEMG signals of the remaining muscle or electroencephalography (EEG) signal can be considered as a control signal for prosthetic hands. When compared to EEG signals, sEMG signals have higher accuracy and reliability (Scheme and Englehart, 2011; Farina et al., 2014; Xia et al., 2018). However, the method of using bio-signals (such as EEG or sEMG signals) to control a robotic hand is still a challenge.

Many studies considered using pattern recognition via conventional machine learning methods, such as support vector machine (SVM), to perform motion classification (Yoshikawa et al., 2006; Angkoon et al., 2012). Some researches trained regression predictors using deep learning methods, such as recurrent neural networks (RNNs), convolutional neural networks (CNN), even a combination of RNNs and CNNs (recurrent convolutional neural networks, RCNN) (Atzori et al., 2016; Wang et al., 2016; Hu et al., 2018; Xia et al., 2018).

Some researchers have used CNN-based models to perform activity recognition and motion classification. Jiang and Yin (2015) proposed a deep convolutional neural network (DCNN) to learn the optimal features from activity images for motion recognition, and obtained high recognition accuracy. Rehman et al. (2018) found that CNN can be used to recognize sEMG patterns for long-term classification even though the sEMG signal is not stable. They extracted four time-domain features and compared them with classical machine learning methods, such as linear discriminant analysis (LDA), stacked sparse autoencoders with features (SSAE-f), and SSAE with raw samples (SSAE-r). CNN was found to be better. Huang and Chen (2019) proposed a hybrid structure combining CNN with LSTM to form a CNN–LSTM model and used the time-frequency domain feature extracted from the EMG signal to classify hand movements. The model was compared with the physical features of conventional methods such as SVM with spectrogram. Ameri et al. (2019) developed a regression-based CNN network for online EMG estimation of wrist motions and compared them with SVM. Bai et al. (2021) proposed a hybrid CNN-LSTM model that effectively combines feature extraction and time series regression for deep learning using sEMG to recognize hand gesture, the experiment results shown that the EMG signal processed by Fast Fourier Transform (FFT) as the characteristic value has better performance, and complex gesture signals can be accurately predicted. In addition to classifiers, CNN is also widely used in regression analysis of hand motions. Koch et al. (2020) chose RNNs to do hand movements regression from sEMG signal, the results proved that even with the relatively simple networks the hand gestures can be regressed quite accurately. Bao et al. (2021a) proposed a CNN-LSTM framework named Deep Kalman Filter Network (DKFN) to estimate wrist and finger kinematics using sEMG, the results shown 0.6–0.8 R^2 result of fingers kinematics and 0.7–0.9 R^2 of wrist kinematics. However, all the studies that used CNN to do classification or regression analysis without discussing the learning methodology of the model from sEMG signals. Moreover, in sEMG pattern recognition, the model accuracy decreases in another day due to the electrode shift or skin impedance, therefore, updating the control system for different day is significant.

Nowadays, modern researches start to use transfer learning to deal with the possible changes in sEMG signals in different days, because transfer learning requires only a short training session to recalibrate the system for new sEMG signals. The new source of information can be leveraged to build a more precise and reliable classifier or regression model for a new dataset, transfer learning allows the capture of more general and robust features, so that the model is then able to use these general features to build a more effective performance of a new sEMG activity. Côté-Allard et al. (2017) combined CNN with transfer learning to classify hand gestures from sEMG, and the model was robust and achieved average classification accuracy of 97.81% on seven gestures. Ameri et al. (2020) proposed an approach based on CNNs with transfer learning to overcome the lack of robustness to confounding factors in EMG pattern recognition-based control, including CNN classification and regression model. Thus, transfer learning not only has effect on classification model, but also can be applied to regression prediction. (Bao et al., 2021b) proposed a state-of-the-art transfer learning method with regression supervised domain adaptation (SDA) for wrist kinematics estimation using sEMG signal, effectively reduced the burden of regular model re-training/recalibration under domain shift effects. In our work, we considered to use transfer learning for checking the proposed model robustness in different days.

Although it is difficult to discuss the weight of the CNN layers for analyzing muscle anatomy and motion activation, Stapornchaisit et al. (2019) used a topology graph to plot the weight of the independence component (IC) to investigate the relationship between the weight of each channel. Thus, we can attempt an analysis of the movement area on both sides of the human forearm by backtracking from the weight parameters in the final layer of the proposed model.

In this research, we proposed a high-accuracy CNN-based regression model to predict forearm joint angles include two wrist motions and one finger motion, and discussed the learnability and high stability of the proposed model for amputees to update the parameters every day. In our experiment, sEMG data and joint angles were collected from participants at the same time as data. As for the sEMG signal, the muscle activities which related to wrist motion and finger motion are mixed, it is difficult to separate these motion. Thus, we designed three degrees-of-freedom (3-DOFs, include wrist flexion/extension and pronation/supination as wrist motions and hand grip/open as finger motion) joint angles as the output of the proposed CNN model based on multi-array sEMG input, the output of the three joint angles will be sent to prosthetic hand for controlling in the future. Then, we used the backtracking method to create a geometry plot for muscle area analysis. We designed a second experiment for smaller amount of dataset and checked that whether transfer learning can improve the regression prediction accuracy. We compared our model with four conventional regression model with and without transfer learning respectively, the conventional regression models are: linear regression (LR), support vector regression (SVR), k-nearest neighbors (KNN) and decision tree regression (DT), and the model comparison results proved that proposed model significantly outperforms conventional regression models. This

TABLE 1 | Participant information.

Participant ID	Age(years)	Handedness	Gender
S1	24	Right	Male
S2	23	Right	Male
S3	21	Right	Male
S4	23	Right	Female
S5	23	Right	Male
S6	26	Right	Male
S7	43	Right	Male
S8	25	Right	Male
S9	24	Right	Male
S10	25	Right	Male

paper shows the stability and superiority of the proposed model, and we consider it can be applied for future real-time prosthetic hand control.

MATERIALS AND METHODS

Participants

Ten right-handed participants (S1–S10) with intact limbs participated in this study; they were nine males and one female, aged 21–43. Data from the participants were acquired at the Tokyo Institute of Technology, Japan. The information of the participants is presented in **Table 1**. The study protocol was approved by the ethics committee of the Tokyo Institute of Technology and was conducted in accordance with the Declaration of Helsinki. All participants were asked to read the participant information sheet and provide written informed consent to participate in the study.

sEMG Data Acquisition

We used a bipolar multi-array electrode with 32 channels (Yasuharu et al., 2020) to acquire EMG signals (SMK Corp., SEIREN Co., Ltd.). **Figure 1** shows the multi-array electrode sleeve, the left and right figures show the wrist flexor and wrist extensor sides, respectively. There are 40 electrodes (5 × 8), each five electrodes measure four channels (two electrodes are shared by one channel), and one reference. The flexor side consists of 16 channels, including ch5–ch8, ch13–ch16, ch21–ch24, and ch29–ch32. The extensor side consists of 16 channels, including ch1–ch4, ch9–ch12, ch17–ch20, and ch25–ch28. The two sides are arranged as a 4 × 4 matrix, and the channel numbers help us to understand the position of each channel. The ADC resolution of the multi-array electrode is 16 bit. The sEMG signal was sent to PC by a Bluetooth module. We acquired raw sEMG signal from this system that consists of a dataset in our experiment, and lab streaming layer (LSL) was used to synchronize data. Before the experiment, we use water to make the sensors and participants' skin perfectly fit, and check the sEMG signal quality. We chose 500 Hz as the sampling frequency (Yasuharu et al., 2020) due to the limited signal transmission speed of Bluetooth low energy, and we do not need to adjust the electrode location for each participant.

Joint Angle Acquisition

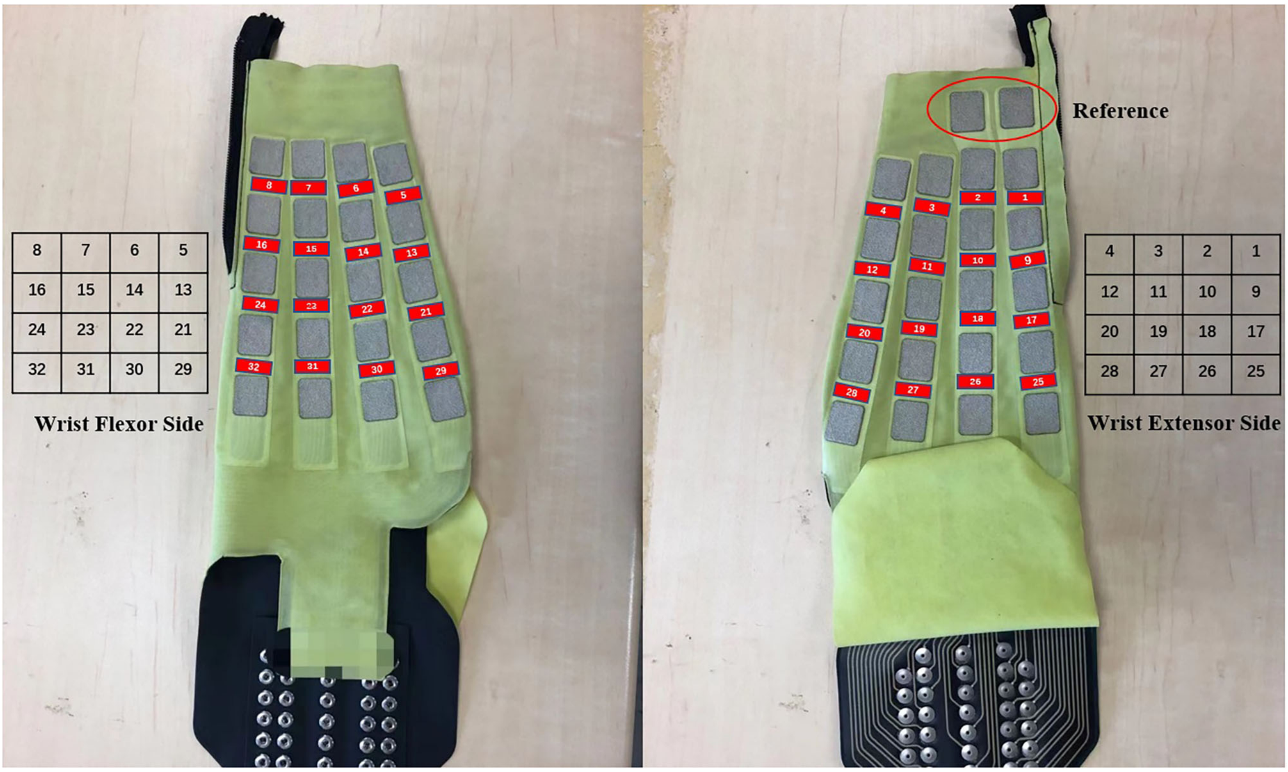
We used the Perception Neuron Motion Capture system (Noitom Ltd., China) to collect joint angles when participants performed specified motions. This system is a type of motion capture device that is applied to movement analysis for film makers, game developers, sports analysis, and biomechanics research related to this work (Kim et al., 2019). In our work, we used only the right hand and right forearm parts. Each participant was required to wear the glove with motion capture markers (**Figure 2B**) and complete calibration to build a skeleton model on the Axis Neuron software provided by Noitom Corp. After successful calibration, we can check the rotation angles by choosing the specific name such as "RightHand," etc. (Yuanhui, 2014) and axis of rotation (X, Y, or Z axis).

While using the Axis Neuron, we chose the BVH data type as the data format, which includes joint hierarchy and movement data. The sensory data of each joint were stored in the corresponding joint local coordination system, body joints of Perception Neuron were arranged as tree structures, and children joints were connected corresponding to their parent joint (Chen et al., 2017). Each joint angle data in the BVH data is calculated from the children joint coordination system to the parent joint coordinate system by internally multiplying the transformational matrix. Finally, the data are shown in global coordinates.

Figure 2A shows the 3-DOFs movements in our study (right hand): black dash line shows the DOF of wrist flexion (WF) and wrist extension (WE); red dash line shows the DOF of pronation (P) and supination (S); blue dash line shows the DOF of hand grip (HG) and hand open (HO). WF/WE and P/S are wrist motion, HG/HO is finger motion. **Figure 2B** shows the required markers (the red points,: RightForeArm, RightHand, and RightMiddle1) to acquire the joint angles, "RightHand" marker was to acquire the joint angles of WF and WE motion, "RightForeArm" marker was to acquire the joint angles of P and S motion, and "RightHandMiddle1" marker was to acquire the joint angles of HG and HO motion.

According to the Neuron Coordinate document provided by Noitom Ltd. and Aiuto Co. Ltd., Japan (Aiuto Co., Ltd., 2020), the global coordinates of the BVH data are shown in **Figure 2B**, where the green, red, and yellow axes are the Z-axis, X-axis, and Y-axis, respectively. Thus, WF/WE joint angles are obtained based on "RightHand" marker rotating around the Z axis, P/S joint angles are obtained based on "RightForeArm" marker rotating around the X axis, and HG/HO joint angles are obtained based on "RightMiddle1" marker rotating around the Z axis. The value of the angle is positive when the rotation is counterclockwise and negative when the rotation is clockwise. Therefore, the joint angles of WF, P, and HG are positive angles, and that of WE, S, and HO are negative angles.

The sampling rate of data collection from the Perception Neuron Motion Capture system was 120 Hz. During the experiment, 3-DOF joint angles and sEMG signal were collected at the same time and synchronized by LSL system. **Figure 3** shows an example of the synchronization between sEMG signal and joint angles.



8	7	6	5
16	15	14	13
24	23	22	21
32	31	30	29

4	3	2	1
12	11	10	9
20	19	18	17
28	27	26	25

FIGURE 1 | Multi-array electrode forearm sleeve (32 channels, right-handed). The red parts show the positions of 32 channels. Left: Wrist flexor side after wearing the sleeve, includes ch5–ch8, ch13–ch16, ch21–ch24, and ch29–ch32. Right: Wrist extensor side after wearing the sleeve, includes ch1–ch4, ch9–ch12, ch17–ch20, and ch25–ch28. The matrix represents the channel distribution on multi-array electrode sleeve and the number represents the channel number. Geometry plot for motion analysis is based on the two matrices.

Experiment Protocol and Data Processing

Participants (**Table 1**) were asked to sit on a chair in front of a screen. After calibrating the joint angles on Perception Neuron Motion Capture system, we checked both the quality of joint angles and sEMG signal. Before the experiment, each participant underwent trial sessions to familiarize themselves with the experimental process (**Figure 4**). After completing each trial, the participants were asked to relax for approximately 2 min before the next trial to prevent fatigue.

For ensuring long-term stability that can be used in the real-time control of a prosthetic hand, we must ensure high accuracy and robustness in another day. We invited the 10 participants to conduct experiment, and five of them were invited to do the experiment again with smaller amount of dataset. Transfer learning was applied to solve this problem, this process is called *the second experiment* in this paper. In order to distinguish the second experiment from the previous experiment (with 10 participants and larger dataset), we call the previous experiment *initial experiment*.

During the initial experiment, we acquired sEMG signal and joint angles at the same time. Each participant performed 10 trials, the following order of each trail shown as follows: WF/WE twice, P/S twice, and HG/HO twice. At the beginning of the experiment, participants performed their hand as a central position (CP). The screen then displayed the movement to be performed, and the participants had to rotate their forearm or finger joint from the CP as per the displayed movements. For each motion, participants rotated their hand to the maximum angle that can be reached, and rotated it back to the CP. The next motion is performed as per the following animation. For example, for WF/WE, the order of motion should be as follows: CP-WF-CP-WE-CP. **Figure 2B** shows a participant performing the experiment and motions according to the animation on the screen. The experimental paradigm was created using MATLAB (The MathWorks, Inc., USA). The second experiment also used the same experimental paradigm, totally 5 trials to reduce the dataset amount.

After the data acquisition step, we used the filter proposed in (Koike and Kawato, 1995) to process the absolute value of the raw sEMG signal to the integrated EMG (IEMG) signal, and normalized the IEMG signal from 0 to 1. **Figure 3** shows the pre-processed result of the raw sEMG signal and the synchronized joint angle data (from participant S1). In **Figure 3A**, blue line is the raw sEMG (ch10), and orange line is the filtered EMG signal. Combining **Figures 3A,B**, we can see clear data synchronization between EMG and angle data *via* LSL time stamps. Owing to the difference in sampling rate between the filtered sEMG signal and joint angle data, we resampled the sEMG signal

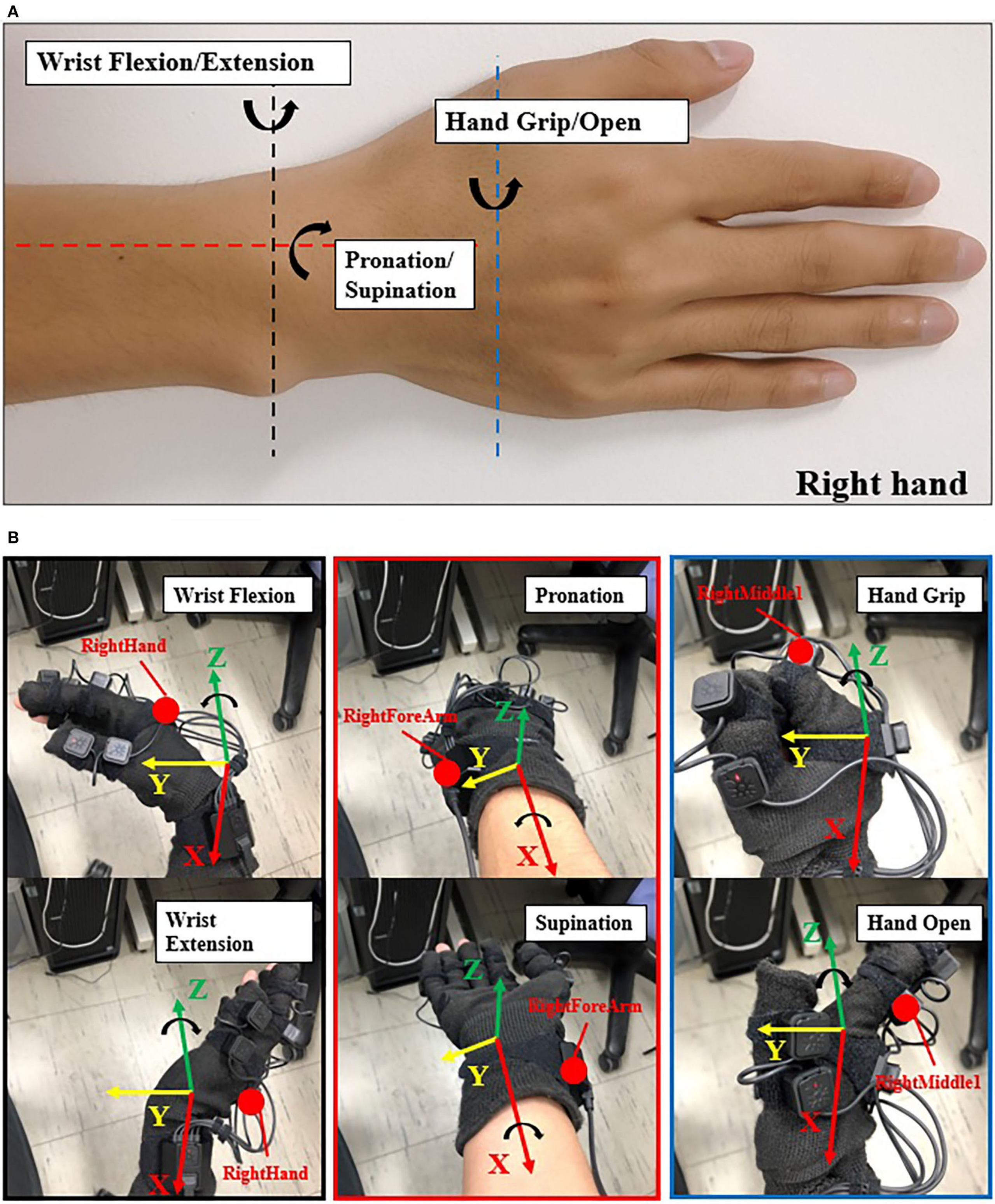


FIGURE 2 | Illustration of joint angle data collection. **(A)** The 3-DOFs movements (right hand) in this study: black dash line shows the DOF of wrist flexion (WF) and wrist extension (WE); red dash line shows the DOF of pronation (P) and supination (S); blue dash line shows the DOF of hand grip (HG) and hand open (HO). WF/WE and P/S are wrist motion, HG/HO is finger motion. **(B)** Data collection with Perception Neuron Motion Capture Glove with 9 markers (right hand). The red point here shows the focus markers we chose. While performing the experiment, we chose “RightForeArm” marker to acquire joint angles of pronation and supination (P/S), “RightHand” marker to acquire joint angles of wrist flexion and wrist extension (WF/WE), and “RightHandMiddle1” to acquire joint angles of hand grip and hand open (HG/HO) motion. The coordinate here shows the BVH global coordinates, red axis is X axis, yellow axis is Y axis, and green axis is Z axis. WF/WE angles are obtained

(Continued)

FIGURE 2 | based on "RightHand" marker rotating around the Z axis; P/S angles are obtained based on "RightForeArm" marker rotating around the X axis; and HG/HO angels are obtained based on "RightMiddle1" marker rotating around the Z axis. Counterclockwise rotation shows positive angle and clockwise rotation shows negative angle, i.e., WF, P, and HG angles are positive values, and WE, S and HO angles are negative values.

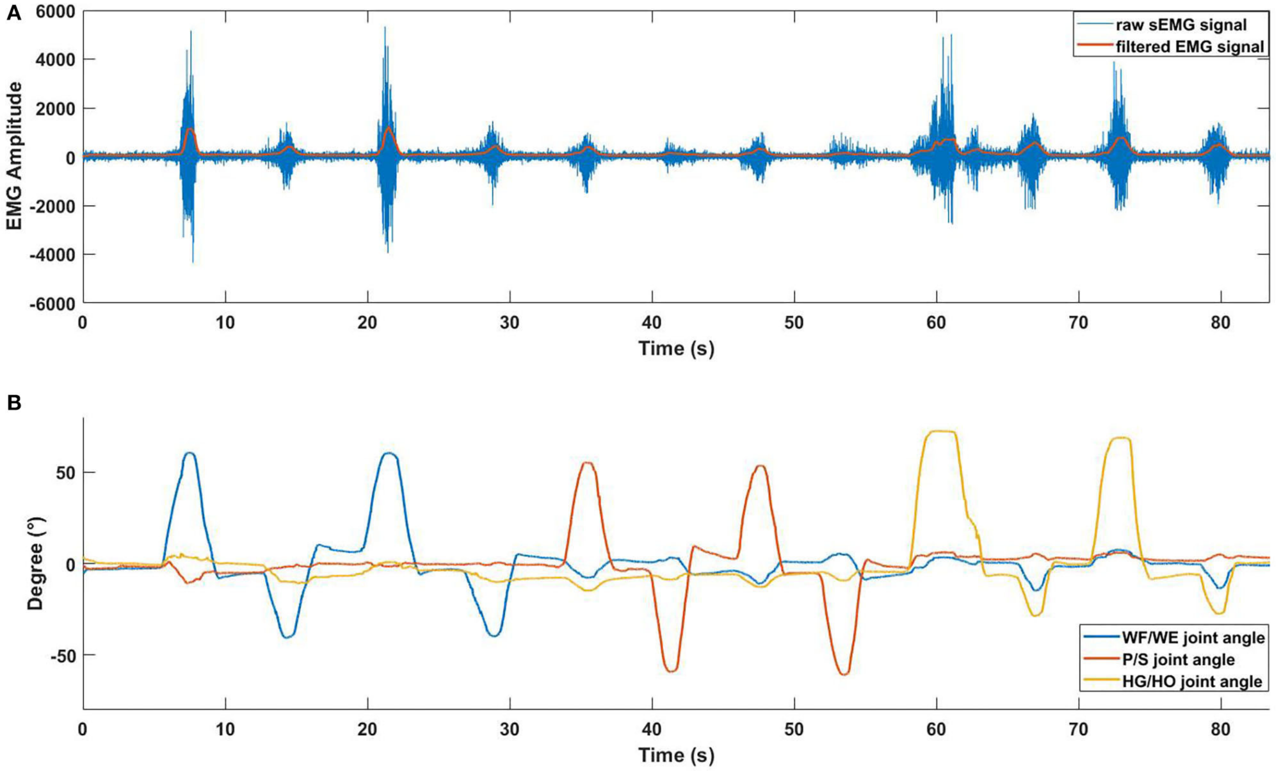


FIGURE 3 | EMG signal with joint angle data (participant S1). **(A)** Pre-processed sEMG signal with raw sEMG signal. The data is obtained from ch10. Blue line represents raw sEMG signal, orange line represents filtered EMG signal before the normalization. The horizontal axis represents the duration time of one trial, vertical axis represents the amplitude of sEMG signal. **(B)** Collected 3-DOF joint angle data using Perception Neuron Motion Capture system.

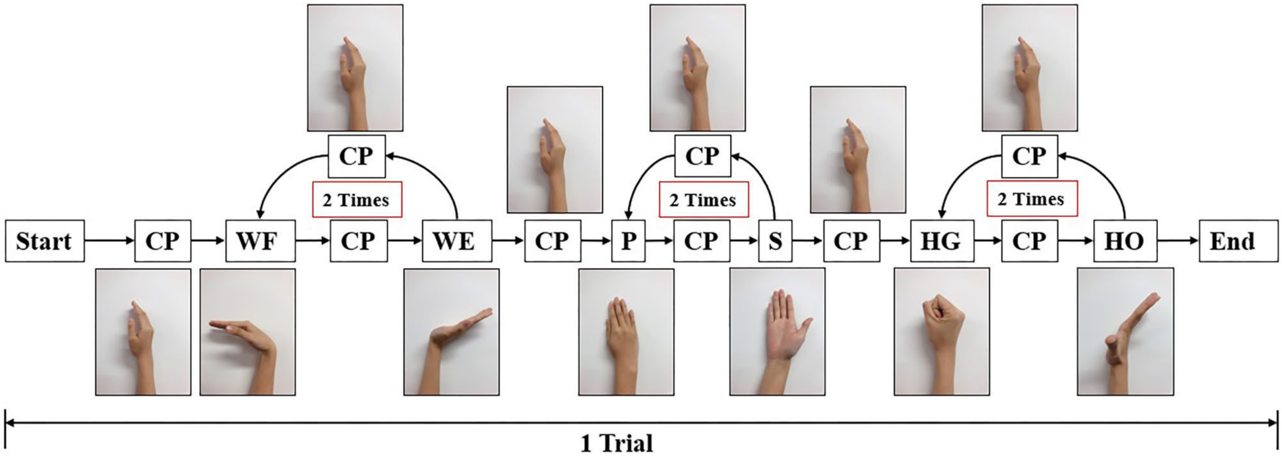


FIGURE 4 | Experiment paradigm for one trial. In each trial, participant should start from central position (CP), perform wrist flexion (WF) and wrist extension (WE) two times, pronation (P) and supination (S) two times, and hand grip (HG) and hand open (HO) two times. After each motion, they should move their forearm to CP and then perform the next motion. The screen displays the next motion to be performed.

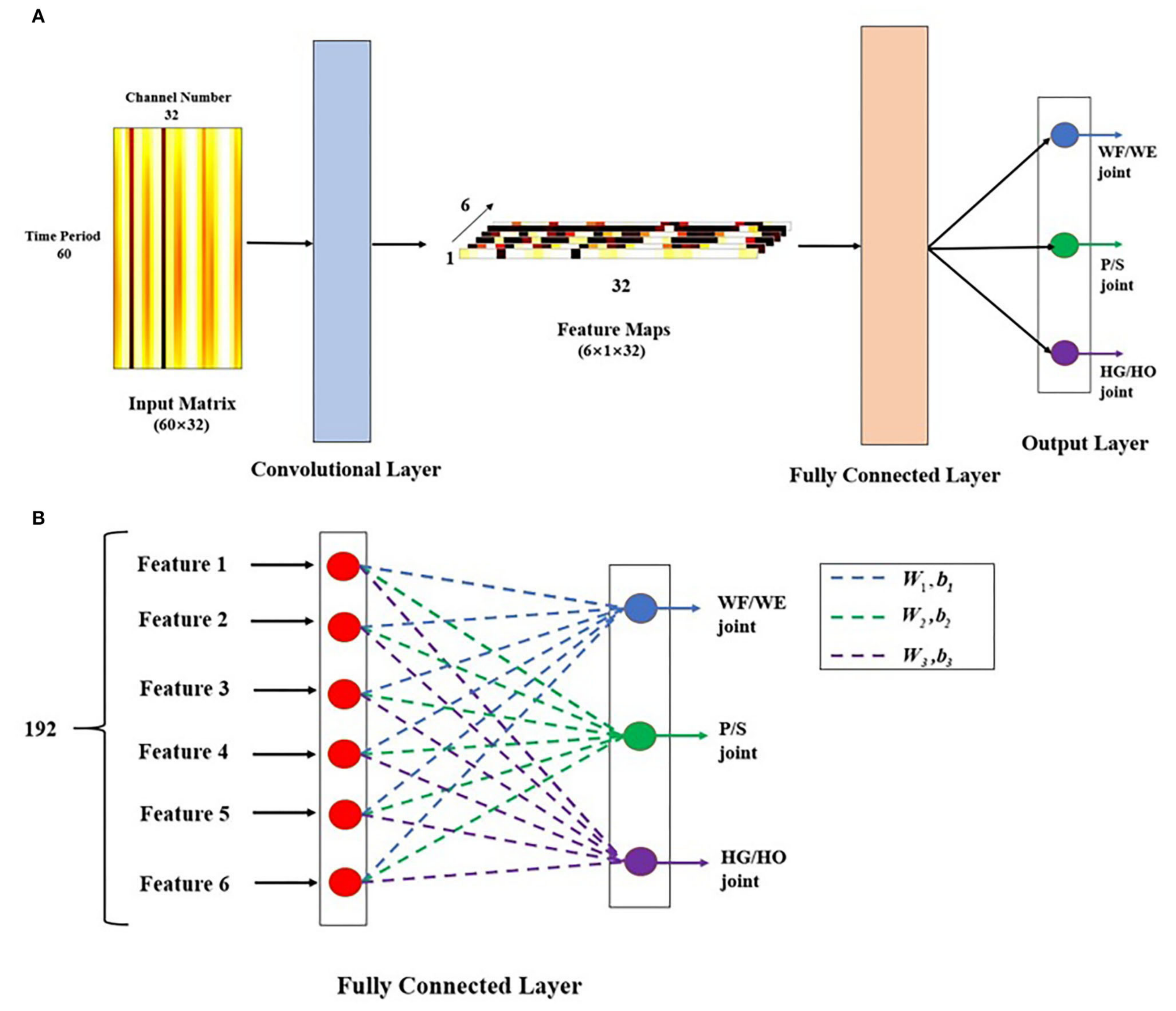


FIGURE 5 | Proposed CNN-based regression model. **(A)** Model structure. This model includes two layers: convolutional layer and fully connected (FC) layer. The input matrix is 500 ms time-domain EMG signal in 32 channels, as a 60 × 32 image. Here, the image is the filtered EMG signal from a participant performing the WF motion. In the convolutional layer, there are six 60 × 1 filters, stride = 1, padding = 0, and activation function is tanh. The output after convolutional layer is six 1 × 32 feature maps, then go through FC layer directly to get three joint angles as 3 × 1 vector as output of this model. **(B)** Details description of FC layer. After the six feature maps were sent to FC layer. They connect end to end to become a 192 × 1 sequence. Each red point shows one transposed feature map with 32 × 1 represents a neuron. The sequence performs dot multiplication on three different weight ($\boldsymbol{W}$) and sum the corresponding bias ($\boldsymbol{b}$) to obtain different joint angles. Blue dashed line represents $\boldsymbol{W_1}$ and $\boldsymbol{b_1}$, which is related to WF/WE joint angle, green dashed line represents $\boldsymbol{W_2}$ and $\boldsymbol{b_2}$, which is related to P/S joint angle, and purple dashed line represents $\boldsymbol{W_3}$ and $\boldsymbol{b_3}$, which is related to HG/HO joint angle.

data from 500 Hz to 120 Hz to match angle data from perception neuron.

To aggregate as dataset, we used a window of 500 ms to segment sEMG data and joint angles, and the ideal dataset should be a 60 × 32 matrix corresponding to a 1 × 3 vector with 3-DOFs joint angles. Hence, we calculated the mean value of joint angles as target angles in the dataset. The interval of the sliding window is 100 ms, which means a 400 ms overlap of the data. After segmenting all the 10 trials of EMG data and joint angle data, for the initial experiment, we randomly combined the 10 trials data into five groups of datasets to perform five-fold cross validation. Then, the 10 trials data were aggregated into five groups of time-domain datasets. For the second experiment, there are 5 trails, one trial data is considered as a group with a total of five groups for the five-fold CV.

CNN-Based Regression Model

CNN is the main architecture of deep learning for a multi-array of data, such as images, signals, and languages. CNN works on the local receptive field, shared weights, and pooling (Le, 2015).

In this work, we proposed a regression model for joint angle estimation based on CNN (**Figure 5A**). This model includes two layers: a convolutional layer and a fully connected (FC) layer. Similar to the conventional two-dimensional (2D) CNN, the input is a time-domain matrix data, which we regard as a 2D image. The input size is 60 × 32, where 60 represents the number of input sample that is 500 ms (the sampling rate of the processed EMG signals and joint angles are 120 Hz); 32 represents the number of channels.

In the convolutional layer, we designed the architecture as a channel-wise CNN (CW-CNN), which was mentioned in (Sakhavi et al., 2018), Discussion section will discuss the choice of the convolutional layer. There are six channel-wise filters, and each filter size is 60×1 in order to compress the time period dimension from 60 to 1. The outputs of the convolutional layer are six 1 × 32 vector feature maps such that we can use backtrack to easily analyze each channel. The activation function in this layer is a tanh function, without any padding and max pooling layer, and with a stride size of 1.

After the convolutional layer, the feature maps directly go to the FC layer (**Figure 5B**). There are six 1 × 32 feature maps, which are transposed and connected end-to-end to become a 192 × 1 vector. There are three groups of weight (W) and bias (b) in this layer, each of which can be used to calculate one joint angle. The 192 × 1 vector input was used to obtain different joint angle outputs using (1).

$$Angle_i = W_i^T \cdot fm + b_i \tag{1}$$

where i is the joint number and $i = 1,2,3$; $\mathbf{Angle_i}$ is the joint angle; $\mathbf{Angle_1}$ is the WF/WE joint angle; $\mathbf{Angle_2}$ is the P/S joint angle; $\mathbf{Angle_3}$ is the HG/HO joint angle; **fm** is the feature map, which is a fully connected feature map with a size of 192×1; and $\mathbf{W_i}$ is the FC layer weight. Among them, $\mathbf{W_1}$ is related to the WF/WE joint, $\mathbf{W_2}$ is related to the P/S joint, and $\mathbf{W_3}$ is related to the HG/HO joint. $\mathbf{W_i} = \{\mathbf{w_i^j}\}$, where $j = 1,2,\ldots,6$ corresponds to one feature map $\mathbf{fm^j}$ with 32 weight numbers, and $\mathbf{b_i}$ is a bias. After the FC layer, the output of the regression model is a 3 × 1 vector, which includes three joint angles: WF/WE, P/S, and HG/HO. The reason to choose the 3-DOFs angles as model output is that, in the future control of prosthetic hand, the three joint angles will be regarded as control command and sent to the three motors respectively in prosthetic hand directly, so that to rotate each motor to the corresponding angular position.

For the proposed CNN-based model, each trial was individually trained to use k-fold cross-validation (k-fold CV) (Stone, 1974; Rodríguez et al., 2010). Cross validation is used to evaluate the predictive performance of a model, particularly the performance of the trained model on new data, which can reduce overfitting to a certain extent. Further, more effective information can be obtained from limited data. The k-fold CV reduces the variance by averaging the results of k different group trainings; hence, the performance of the model is less sensitive to the division of data. In this study, $k = 5$, *i.e.*, we trained and tested the dataset using a five-fold CV.

To evaluate the model using five-fold CV, we use the correlation coefficient (CC), which is used to statistically measure the strength of the relationship between two variables (Taylor, 1990). For example, in machine learning, CC is used to measure the relationship strength between the estimated angle series and measured angle series in a dataset. CC values range between −1.0 and 1.0, a correlation of 1.0 shows a perfect positive correlation and −1.0 shows a negative correlation; CC = 0.0 shows no linear relationship between the two series. Therefore, if we want to train an ideal rotation angle predictor model, we should obtain the CC result as approximately 1.0, i.e., for the predictions and measurements to be positively correlated at a high level. The equation of CC is shown in (2).

$$CC = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}} \tag{2}$$

where X and Y are two series variables and n is the number of samples. In this study, the dataset was divided into five-folds. In every iteration, the proposed regression model was learned using four-folds. The remaining fold was tested to calculate performance indicators such as CC. Thus, we obtained five evaluation values after the five-fold CV of a motion pattern, and the average of the five results of each fold was calculated to obtain the final testing result (Diamantidis et al., 2000; Sakhavi et al., 2018), as shown in (3).

$$CC_5 = \frac{1}{5}\sum_{k=1}^{5} CC_k \tag{3}$$

Geometry Plot of FC Layer Weight

The geometry plot shows the weight of the feature map (Stapornchaisit et al., 2019), which has a significant contribution to the different forearm motions in the FC layer. Here, the weights of the FC layer were separated into two parts and distributed corresponding to channels (**Figure 1**) representing the wrist flexor side and wrist extensor side, respectively. As previously mentioned, in the FC layer, the size of weight $\mathbf{W_i}(\mathbf{i} = \mathbf{1}, \mathbf{2}, \mathbf{3})$ is a 192×1 vector and each $\mathbf{w_i^j}$ corresponds to one feature map $\mathbf{fm^j}$, which has 32 weight numbers corresponding to the channels. Thus, a high or low value of weight indicates the importance of the channels for specific forearm movements. All six feature maps contribute to the computation of joint angles with the FC layer weight. To check the motion pattern using weight, we can separate the FC layer weight into six 32 × 1 weight parts corresponding to feature maps and perform superposition calculation as a 32 × 1 vector. We then arrange the 32 weight numbers based on the 32 channels positions as the expected geometry plot. To analyze the WF/WE motion, we plotted a geometry plot based on $\mathbf{W_1}$; for P/S, we used $\mathbf{W_2}$; and for HG/HO, we used $\mathbf{W_3}$ (see Results Section and **Figure 6**).

RESULTS

Five-Fold Cross Validation Results

The model used for training was created using pytorch1.3.1, GeForce RTX 2080 GPU, and CUDA10.1. For the training, we

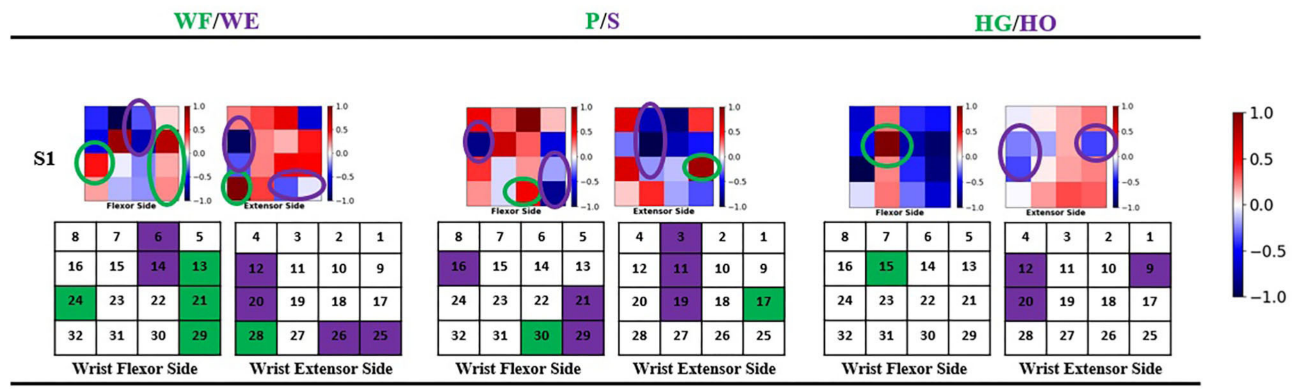


FIGURE 6 | Geometry plot of participant S1. Each group of geometry plot shows flexor side (left) and extensor side (right). The color bar shows the weight normalized from −1 to 1. White color shows the channel does not contribute to motions, red color shows 0–1 and contributes to positive joint angles (WF, P, HG), and blue color shows −1–0 and contributes to negative joint angles (WE, S, HO). The circles show the motion pattern position: green circle shows the motion pattern corresponding to positive joint angles, and purple circle shows the motion pattern corresponding to negative joint angels. Under the geometry plot, the matrix shows the motion pattern color with channel index.

TABLE 2 | CC results of five-fold cross validation (mean CC ± std).

	Initial Experiment			Second Experiment		
	WF/WE	**P/S**	**HG/HO**	**WF/WE**	**P/S**	**HG/HO**
S1	0.9298 ± 0.0121	0.9500 ± 0.0259	0.8849 ± 0.0410	0.9746 ± 0.0089	0.9189 ± 0.0273	0.9183 ± 0.0292
S2	0.8928 ± 0.0269	0.9266 ± 0.0405	0.9413 ± 0.0465	-	-	-
S3	0.8721 ± 0.0311	0.8628 ± 0.0164	0.8229 ± 0.0488	-	-	-
S4	0.8525 ± 0.0457	0.8556 ± 0.0311	0.8484 ± 0.0186	-	-	-
S5	0.8755 ± 0.0298	0.7235 ± 0.0459	0.7765 ± 0.0331	-	-	-
S6	0.9017 ± 0.0271	0.8563 ± 0.0406	0.8915 ± 0.0485	0.9210 ± 0.0243	0.8434 ± 0.0712	0.9298 ± 0.0443
S7	0.9087 ± 0.0355	0.9323 ± 0.0169	0.7589 ± 0.0105	0.9247 ± 0.0106	0.9647 ± 0.0080	0.8864 ± 0.0310
S8	0.8636 ± 0.0178	0.9392 ± 0.0269	0.8086 ± 0.0152	0.9050 ± 0.0372	0.9507 ± 0.0259	0.8577 ± 0.0404
S9	0.8617 ± 0.0210	0.8223 ± 0.0213	0.8519 ± 0.0164	-	-	-
S10	0.8799 ± 0.0231	0.8445 ± 0.0191	0.8540 ± 0.0186	0.9107 ± 0.0272	0.9464 ± 0.0281	0.8665 ± 0.0398
Mean	0.8838 ± 0.0270	0.8713 ± 0.0284	0.8439 ± 0.0297	0.9272 ± 0.0216	0.9248 ± 0.0321	0.8918 ± 0.0369

* *"-" means this participant did not participate in the second experiment.*

used the Adam optimizer to update the model parameters to minimize the loss function and mean-square error (MSELoss) as our loss function. The learning rate was 0.001. In the initial experiment, the training epoch was 15; For second experiment dataset, training epoch was 5 for transfer learning.

The result of the five-fold CV was evaluated using the CC. The mean CC values of all participants are listed in the left part (initial experiment) of **Table 2**, and **Figure 7** shows the box plot corresponding to **Table 1**. An example of the comparison between the estimated joint angles using the proposed regression model and measured joint angles via the Perception Neuron Motion Capture system are plotted in **Figure 8**, where the red solid line represents the measured angles and green dashed line indicates the estimated angles. This figure shows the prediction result of S1, CC = 0.9465 for WF/WE joint angles, CC = 0.9686 for P/S joint angles, and CC = 0.9074 for HG/HO joint angles.

We invited the participant S1, S6, S7, S8 and S10 to complete the second experiment. The second experiment was conducted two months after the initial experiment. We used the trained model to test the new dataset directly to check the mean test CC, we called this procedure as *Direct Model Testing*. As we expected, the mean CC values are lower than the CC value in initial experiment respectively (**Figure 9**, red bar), this is because the magnitude and quality of the sEMG signal often differ grately in another day. Then, we used the trained model and fixed the convolutional layer, using transfer learning to check the model via five-fold CV. The average CC results are listed in the right part (second experiment) of **Table 2**. A comparison of the mean CC results in initial experiment, direct model testing and second experiment is shown in **Figure 9**, blue bar shows the mean CC value in initial experiment, red bar shows the direct model testing result, and the yellow bar shows the mean CC value in second experiment. An example of the comparison between the

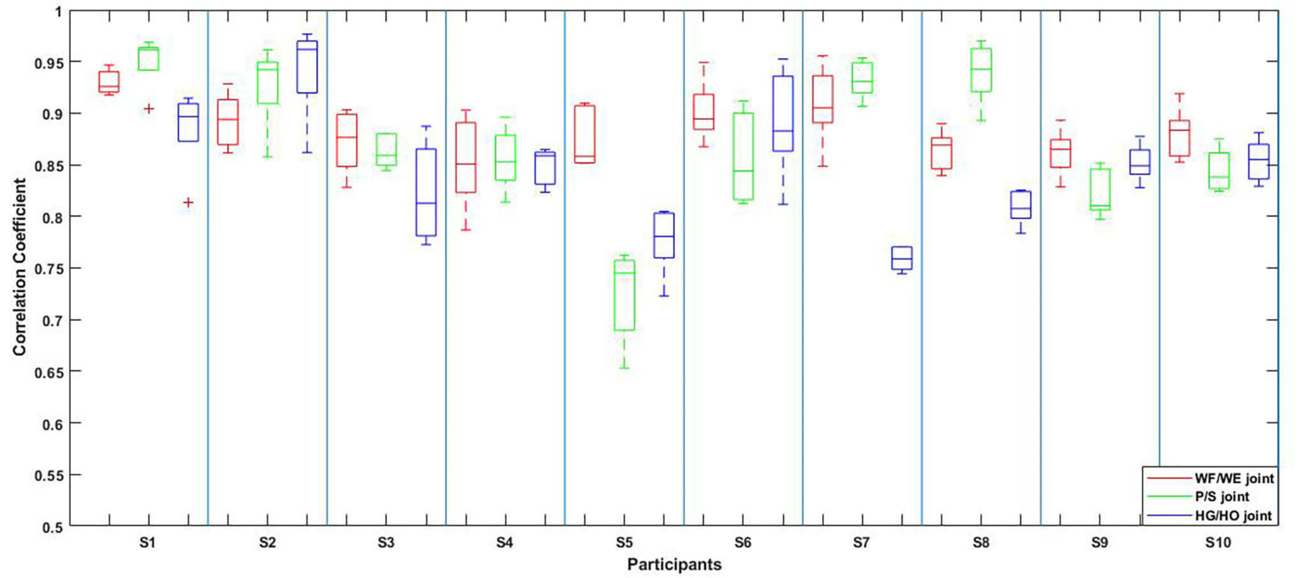


FIGURE 7 | Box plot of five-fold CV results of all the participants according to **Table 2**. The red box is the CC result of WF/WE joint angles, the green box is the CC result of P/S joint angles, and the blue box is the CC result of HG/HO joint angles.

estimated joint angles using the proposed regression model and measured joint angles via Perception Neuron Motion Capture are plotted in **Figure 10**, where the red solid and green dashed lines represent the measured and estimated angles, respectively. The **Figure 10** shows the prediction result of S1, CC = 0.9793 for WF/WE joint angles, CC = 0.9553 for P/S joint angles, and CC = 0.9573 for HG/HO joint angles.

Model Performance Comparison

In order to highlight the advantages of proposed CNN-based regression model in joint angles prediction based on EMG signal, this model was compared with four conventional regression model: linear regression (LR), support vector regression (SVR), k-nearest neighbors (KNN) and decision tree regression (DT). We compared these five regression models with and without transfer learning, and the results were shown as **Figures 11**, **12**. We will discuss the model performance comparison in the Section Model comparison.

Figure 11 shows the average CC results of the five regression models applied to each participant, the top figure is the comparison result of initial experiment (without using transfer learning; 10 participants), the bottom figure is the result of second experiment (with transfer learning; five participants). The results show that for any participant, whether transfer learning was applied or not, the proposed model outperformed the four conventional regression models.

Figure 12 is the comparison result between the five regression models, the vertical axis is the average CC of all participants, **Figure 12A** is the result of initial experiment (without using transfer learning; 10 participants) and **Figure 12B** is the result of second experiment (with transfer learning; five participants), average CC and standard deviation are shown as mean ± std format. We can find that for all participants, proposed CNN regression model outperformed than the other four conventional regression models, and the standard deviation of proposed CNN model is smaller and stable.

Geometry Plot Results

Table 2; **Figure 7** show that the proposed regression model accurately predicts the joint angles in the three DOFs of forearm motion. Therefore, we investigated the accurate model performance learnt directly from the filtered EMG signal and if the muscle activity can be analyzed from geometry plot. Different muscle areas perform different motions; therefore, we backtrack from the FC weight to create a geometry plot. In this section, we use the trained model trained obtained from all participants to find the correlation motion pattern. The corresponding result of participant S1 are shown in **Figure 6** (result of all 10 participants can be found in **Supplementary Material**), each geometry plot includes two parts: the left side denotes the wrist flexor side and right side denotes the wrist extensor side. The color bar ranges from −1 to 1, the blue part shows −1–0, i.e., the area contributes to negative joint angles (WE, S, HO); the red part shows 0–1, i.e., the area contributes to positive joint angles (WF, P, HG); and the white part is 0, i.e., the area has no contribution to the corresponding motion. The green circle shows positive joint angles such as WF, P, and HG, and the purple circle shows negative joint angles such as WE, S, and HO.

DISCUSSION

We proposed a CNN-based regression model for real-time prediction of joint angle of wrist and hand motion using sEMG signals. We used the CC value to evaluate the model training

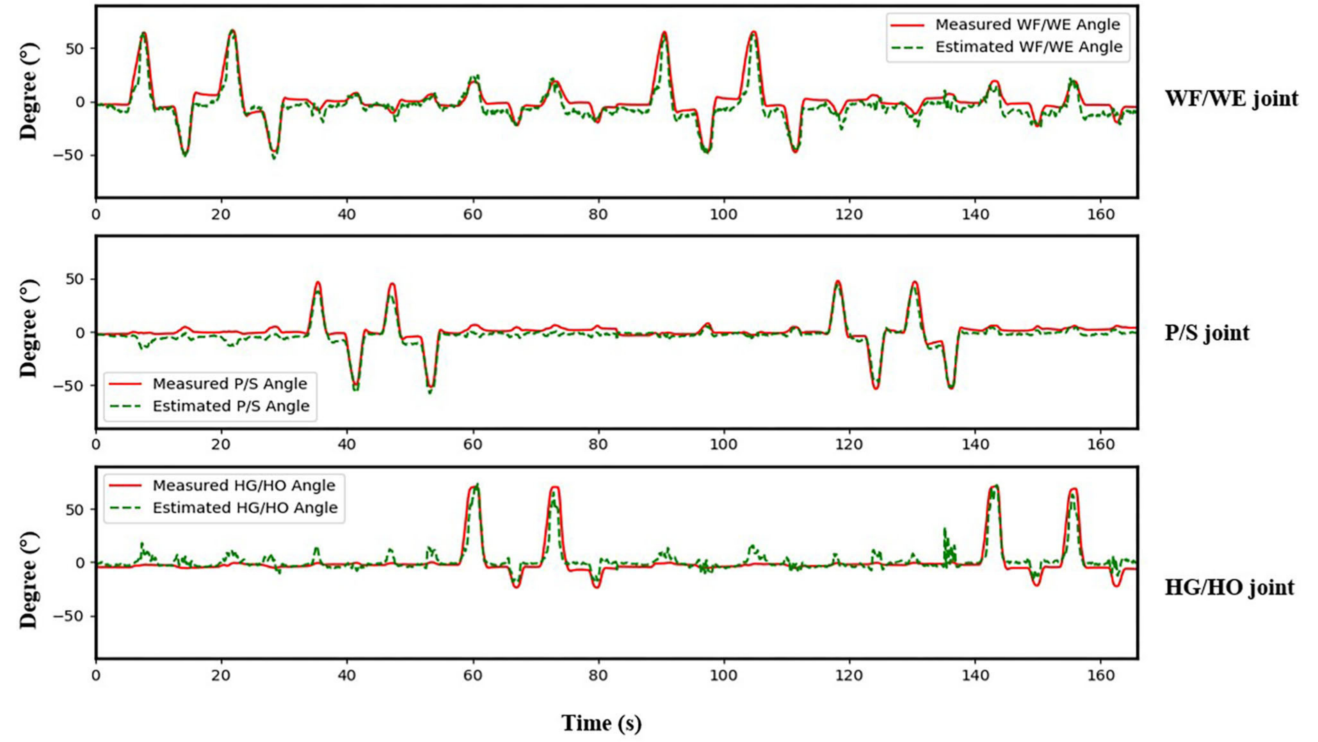


FIGURE 8 | Testing result of participant S1. Top: WF/WE joint; Middle: P/S joint; Bottom: HG/HO joint. Red solid line is measured joint angles using Perception Neuron Motion Capture system. Green dashed line is estimated joint angles using proposed regression model. This result shows CC = 0.9465 for WF/WE joint angles, CC = 0.9686 for P/S joint angles, and CC = 0.9074 for HG/HO joint angles.

effect for 10 participants in initial experiment and explore the learnability of this model by directly using EMG signals. Further, we backtracked the FC layer weight to create a geometry plot for the wrist flexor and wrist extensor sides, and checked the area of the muscles for the corresponding motions. We hope that the model can be robust and applied in different day, participant S1, S6, S7, S8 and S10 were invited to repeat our experiment for smaller number of datasets as second experiment. Firstly, we tested the new dataset directly using the existing trained model to prove the prediction accuracy reduced in another day. Then, we applied transfer learning, fixed the first layer, updated the FC layer parameters to train the new smaller dataset to check the testing CC result and the geometry plot, so that to verify an improvement in the testing CC result. We also compared proposed CNN model with four conventional regression model (LR, SVR, KNN and DT) to prove the superiority of the proposed model. The results showed the model can be used in different day with small number of sEMG data using transfer learning.

Convolutional Layer Design

In this study, we proposed a model with two layers: a convolutional layer and an FC layer. The function of the FC layer is to flatten the feature maps into a single vector; hence, the key point in designing the CNN-based regression model is the convolutional layer. Sakhavi et al. (2018) considered three types of convolution kernel for linear mixture of EEG signal input, including CW-CNN, channel mixing CNN (CM-CNN), and channel-wise convolution with channel mixing (C2CM). According to Sakhavi et al. (2018) research, the difference between CW-CNN, and CM-CNN and C2CM is that CW-CNN does not demonstrate channel mixing. Channel mixing leads to a widened network, and without channel mixing, the receptive field of the network is emphasized.

The process of skeletal muscle contraction and relaxation can be expressed as follows: the electrical signals from the brain reach the nerve endings, and the action potentials are transmitted to the cell membranes of the nerve endings. A series of chemical changes occur to change the conformation of tropomyosin and expose muscle movement. The binding site of the protein and myosin, and head of myosin is activated generating power to swing the head and slide the thin filaments. Since the arrival of electrical signals at the neurons to contract or relax a muscle, a time delay should be generated at times. Thus, the input data in our study is a linear mixture of 32 EMG channels with temporal dimension, and we hope that the time period input can provide sufficient information of the sEMG signal and muscle force pattern that occur during the time delay.

The channel-wise filter (CW-CNN) designed in the convolutional layer can reduce the temporal dimension to one-dimensional feature maps, including 32 channels information. Thus, the channels corresponding to each number from the

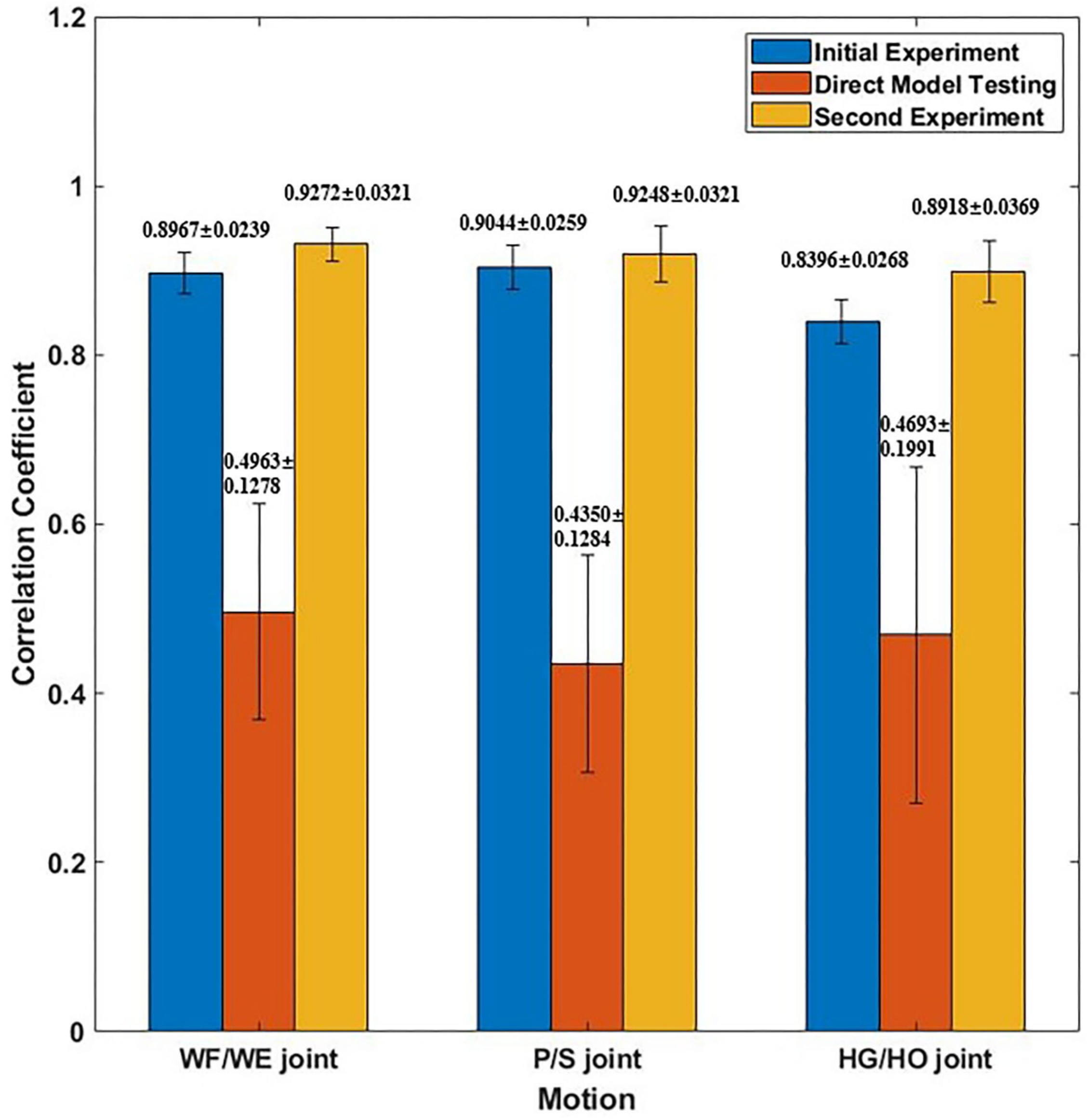


FIGURE 9 | Mean CC result of participant S1, S6, S7, S8 and S10. Blue bar shows the mean CC result in initial experiment (15 epochs); Red bar shows the mean CC result of direct model testing, that is to use the trained model in initial experiment to test on the new dataset directly; Yellow bar shows the mean CC result in second experiment, which fixed the first layer and only to train FC layer (five epochs).

feature maps are independent of each other. Further, because the temporal dimension is reduced, we obtained the muscle force pattern from each channel using such a channel-wise convolution kernel. Unlike the CW-CNN proposed in Sakhavi et al. (2018), which was used to process the EEG signal, we used a one-dimensional convolution kernel to obtain the muscle force pattern, which is called as a *force pattern filter* (FP filter).

Five-Fold Cross Validation

In the initial experiment, we used the five-fold CV to train and test the proposed model. For each participant, we created five groups of datasets, we trained the four groups and tested them on the remaining dataset. Hence, there were five CC results for each participant. Equation (3) was used to calculate the average values of the CC of different participants for joint WF/WE, joint P/S, and joint HG/HO. **Table 2** left part (initial experiment) shows the five-fold CV results of all participants in the form of mean ± std. This result is plotted in **Figure 7**. After checking the raw sEMG signal from each participant, the qualities of the sEMG signal of S1, S2 and S8 are the best, S3, S4, S6, S7, S9 and S10 are slightly noisy, and sEMG signal quality of S5 is the worst and noisy channels are more than that of the other four participants. Therefore, we inferred that several noisy channels confuse our model in predicting P/S joint angles and HG/HO joint angles, which contain more inner muscle sEMG signals, and hence are more difficult to predict than the WF/WE joint angles. We can

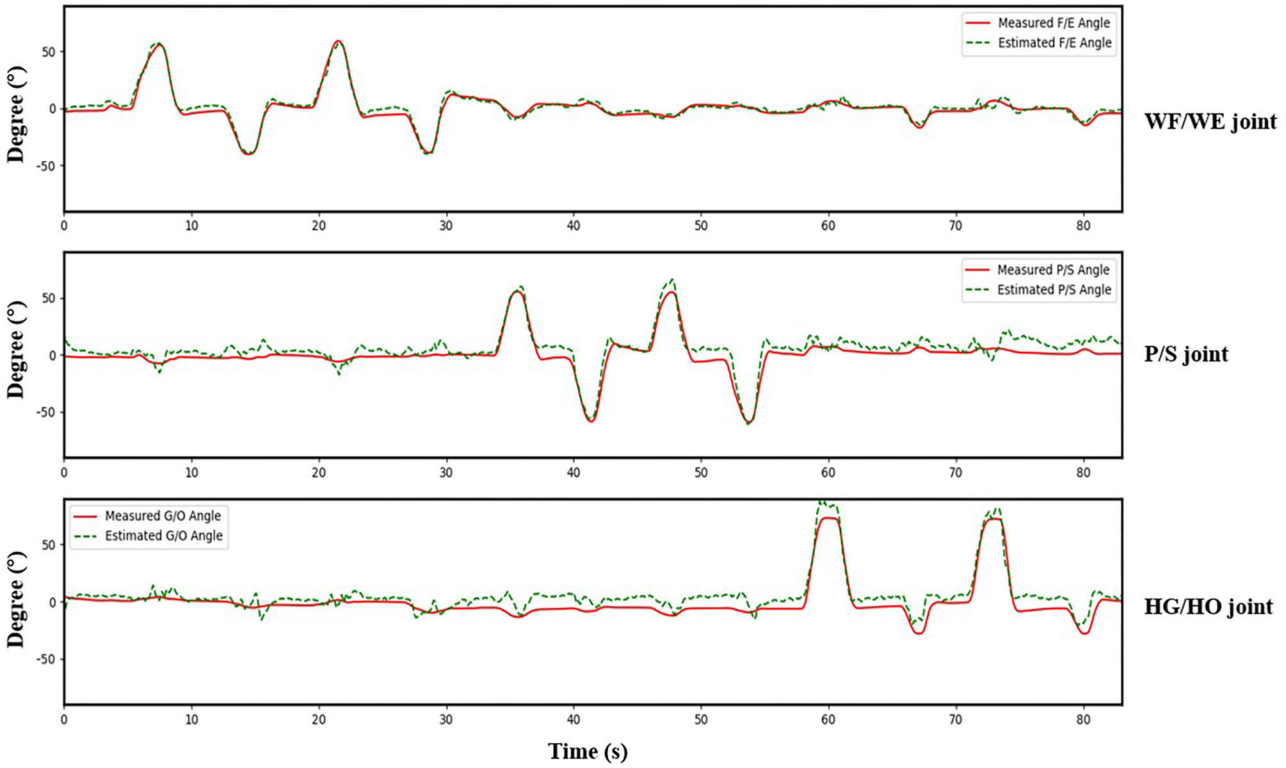


FIGURE 10 | A testing result of participant S1. Red solid line is measured joint angles using Perception Neuron Motion Capture system. Green dashed line is estimated joint angles using proposed regression model. This result shows CC = 0.9793 for WF/WE joint angles, CC = 0.9553 for P/S joint angles, and CC = 0.9573 for HG/HO joint angles.

further conclude that with a less noisy sEMG signal as input, the proposed regression model can perform very well in predicting the three important DOF joint angles. **Figure 8** shows one of the testing results obtained for the five-fold CV of S1. The top figure shows the WF/WE joint, middle figure shows the P/S joint, and bottom figure shows the HG/HO joint. We found that the predicted joint angles met our expectations. When the participant performed hand grip and open motion, the WF/WE joint showed small angles and vice versa. Further, the HG/HO joint showed a smaller angle when performing wrist flexion and extension motion, which is appropriate because WF/WE and HG/HO have common muscle areas. The participants S1, S6, S7, S8 and S10 with relatively high-quality sEMG signals (**Figure 3** shows the raw sEMG signal of S1 from one of the channels) participated in the second experiment.

While the second experiment, with the same experimental paradigm (**Figure 4**), we obtained five trials dataset, and each trial data was regarded as one group of datasets to continue with the five-fold CV training and testing. If we used a trained model for prosthetic hand control, the prediction accuracy decreases in another day. Furthermore, even for the same participant, the quality of the sEMG signal always changes, with the existing trained model, the testing result in another day should be worse; hence, amputees should train the model to calibrate the control system before using the prosthetic hand. We used the existing model to do transfer learning, the first layer (convolutional layer) was fixed, and the testing CC result of the five-fold CV of the five participants is shown in **Table 2**, where the mean CC of the five participants of the WF/WE, P/S, and HG/HO joints was 0.9272 ± 0.0216, 0.9248 ± 0.0321, and 0.8918 ± 0.0369, respectively.

Figure 9 shows the corresponding result, the blue bar is the mean testing CC result of S1, S6, S7, S8 and S10 in initial experiment from 10 trial dataset and trained using five-fold CV; the red bar is the testing CC result in Direct Model Testing, which means to test the existing trained model on new dataset directly; the yellow bar is the testing result of S1, S6, S7, S8 and S10 in second experiment, fixed the convolutional layer and trained FC layer in five epochs, as shown in the right part of **Table 2**. When compared to the blue bar, which indicates the initial experimental result, the red bars show that the mean CC value of the participants in Direct Model Testing lower than in Initial Experiment, WF/WE joint CC reduced to 0.4963 ± 0.1278 from 0.8967 ± 0.0239, P/S joint CC reduced to 0.4350 ± 0.1284 from 0.9044 ± 0.0259, HG/HO joint CC reduced to 0.4693 ± 0.1991 from 0.8396 ± 0.0268. This result is in line with our expectations. With small number dataset and only five epochs, the mean CC results are improved, the yellow bars show that, compared to the red bars which indicates Direct Model Testing, WF/WE joint CC reaches to 0.9272 ± 0.0321,

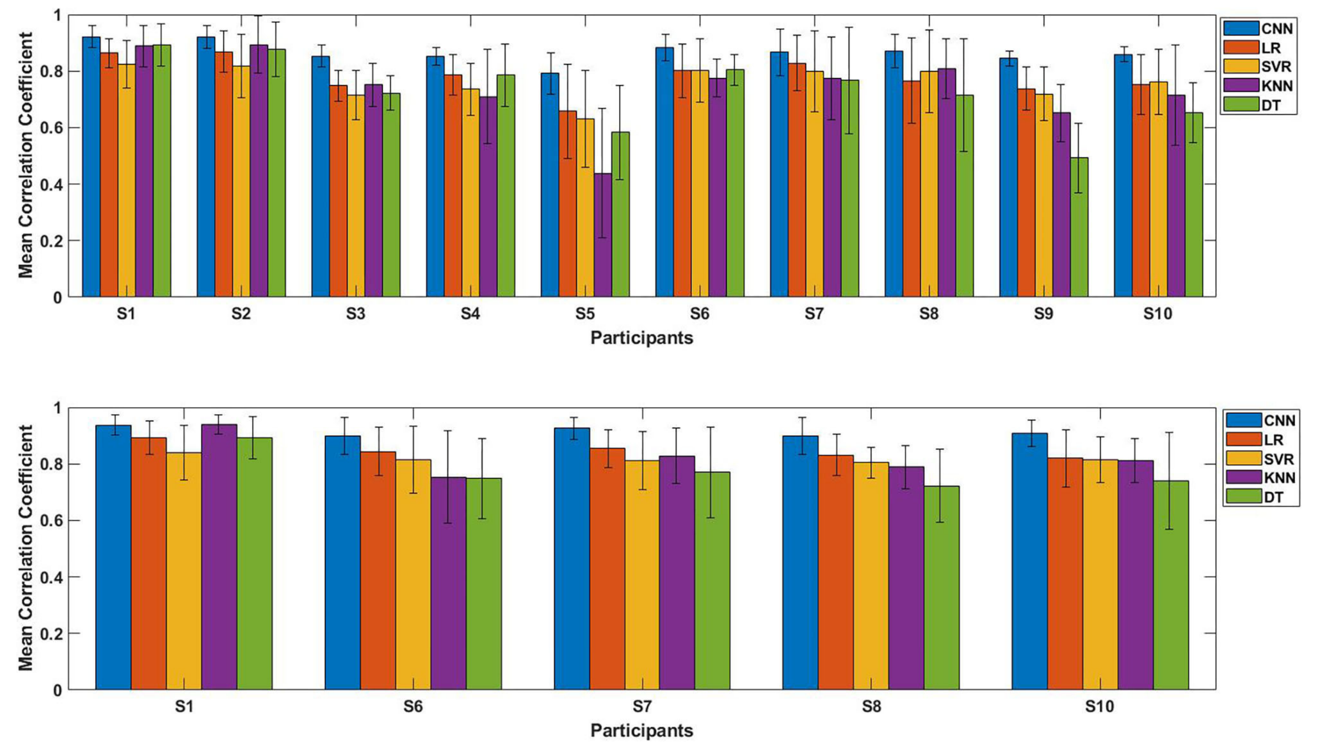


FIGURE 11 | The performance comparison of each participants using five regression models (Proposed CNN, LR, SVR, KNN, DT). Top: Comparison result of initial experiment (without transfer learning; 10 participants); Bottom: Comparison result of second experiment (with transfer learning; five participants).

P/S joint CC reaches to 0.9248 ± 0.0321, HG/HO joint CC reaches to 0.8918 ± 0.0369. We can find that with smaller dataset and only 5 trials, the model can keep the high CC value every day, and even higher than before. Let us compare yellow bar (second experiment) to blue bar (initial experiment), average CC of WF/WE joint, P/S joint and HG/HO joint were improved. The **Figure 10** shows a testing result of the five-fold CV of S1 in second experiment, top figure shows WF/WE joint, middle figure shows P/S joint, and bottom figure shows HG/HO joint.

Then we will discuss the reason why using transfer learning can improve the performance (**Figure 9**). In the initial experiment, parameters from both FP filter and FC layer were trained from the 10 trails datasets of each participant, the parameters of the model include the information of all of the dataset. As we discussed in Section convolutional layer design, reference to CW-CNN (Sakhavi et al., 2018), the proposed FP filter designed in the first convolutional layer can be used to obtain the muscle force pattern from each channel, namely, the feature map extracted from the EMG signal input should be force pattern, and different people have their own force pattern. In the second experiment, if we fix the convolutional layer (FP filter) and only update FC layer parameters, the model still contains the previous 10 trial dataset information, and the FC layer parameters can adapt the model to the new data set. From the perspective of the entire training process, it is equivalent to adding a new dataset to the original dataset, totally 15 trial datasets. Therefore, this approach can ensure that in the daily update training, although there is only less training dataset, the model parameters always contain the training information of all previous datasets. The overall training dataset is constantly superimposed. The prediction result will become better, and it will also be more conducive to amputees to update the training daily.

Model Comparison

To confirm the proposed CNN model performs well, we compared it with four conventional regression models (LR, SVR, KNN and DT). **Figure 11** shows the results of each participants using different regression model with error bars, top figure is the initial experiment result (without transfer learning), bottom figure is the second experiment result (with transfer learning); **Figure 12** shows the results of the model comparison on all participants with transfer learning (**Figure 12A**) and without transfer learning (**Figure 12B**). Both **Figures 11**, **12** show that the proposed model outperforms the four conventional models in the dataset of 10 participants. The comparison results also prove that the proposed model performs better in another day using transfer learning with even small amount of dataset and fewer training epoch. When measuring muscle activity, the magnitude and quality of the signal often differ greatly depending on the contact resistance between the electrode and the skin. Therefore, in the trained model, there is a problem that the estimation accuracy of the data on another day is lowered. In order to solve

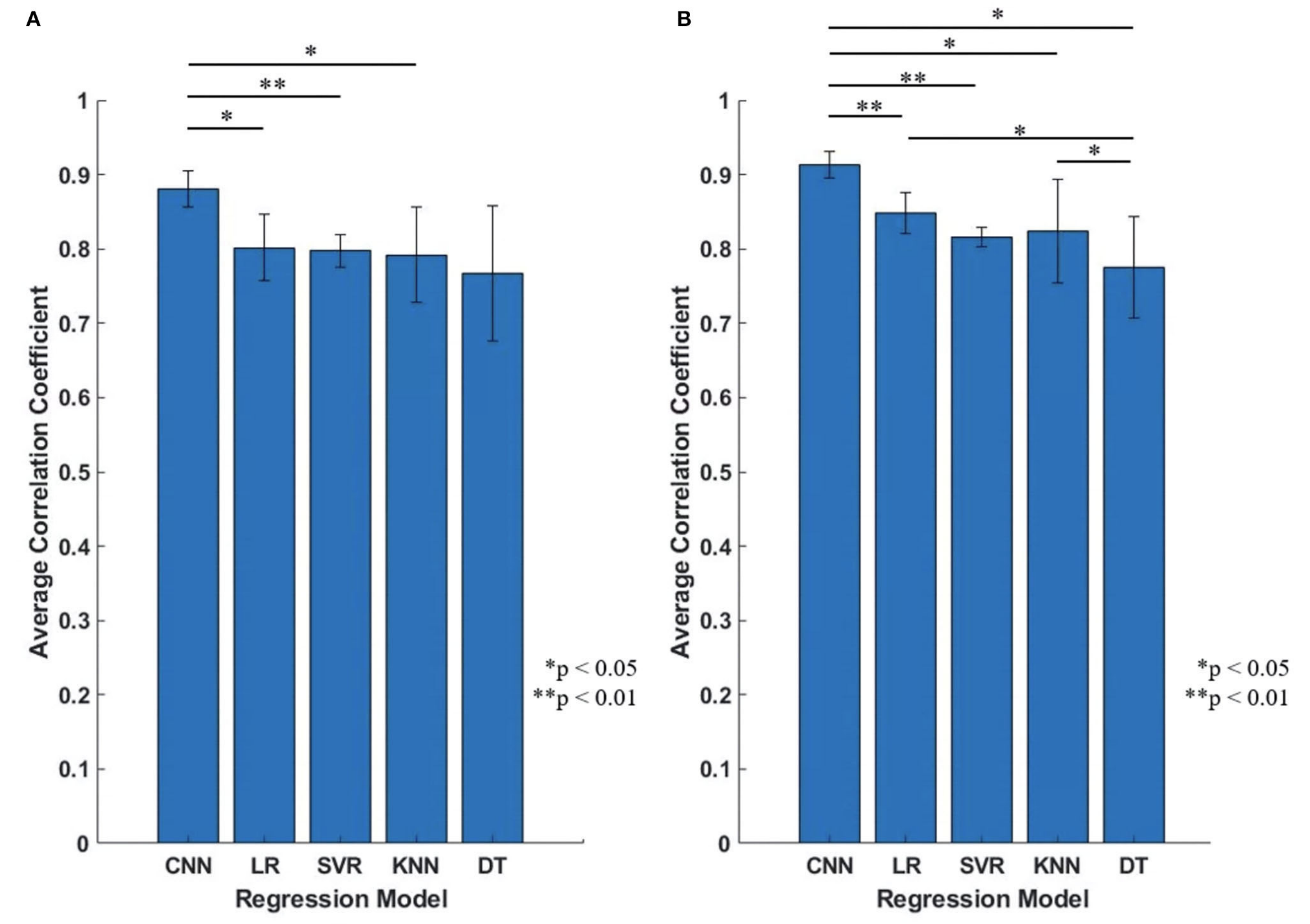


FIGURE 12 | The performance comparison of the five regression models (Proposed CNN, LR, SVR, KNN, DT). **(A)** Initial experiment (without transfer learning; 10 participants); **(B)** Second experiment (with transfer learning; five participants). Statistical differences were calculated using Student's *t* test with Benjamini and Hochberg false discovery rate (BHFDR) correction for multiple comparison.

this problem, we aimed to improve the estimation accuracy even with a small amount of data on new day. The average CC still shows the highest compared to the four traditional regression model with transfer learning (**Figure 12B**).

As presented in **Figure 12**, the statistical analysis was performed to show the proposed CNN shows significantly higher performance than the conventional methods. Statistical differences were calculated using Student's *t* test with Benjamini and Hochberg false discovery rate (BHFDR) (Benjamini and Hochberg, 1995) correction for multiple comparisons. From **Figure 12A**, the CC value for the proposed CNN is 0.8803 ± 0.0247, is significantly higher than LR ($p = 0.021 < 0.05$), SVR ($p = 0.0016 < 0.01$) and KNN ($p = 0.0337 < 0.05$); From **Figure 12B**, the CC value for CNN is 0.9133 ± 0.0175, is significantly higher than LR ($p = 0.0044 < 0.01$), SVR ($p = 0.0004 < 0.01$), KNN ($p = 0.031 < 0.05$), and DT ($p = 0.015 < 0.05$). In our CNN model, we only used one convolutional layer to make the model have better performance than conventional regression models, instead of using multiple convolutional layers or more complex deep learning models to achieve high-precision predictions, which makes regression prediction for joint angles with higher efficiency.

Motion Pattern

In this study, we assumed that the proposed regression model can obtain the corresponding motion pattern through the sEMG signal of different channels on the wrist muscle. After the FP filter in convolutional layer, the feature maps include the force pattern information of the participant, then the FC layer interacts directly with the feature maps to calculate the joint angle output, we thought that it should be motion pattern. Actually, the motion pattern showed as geometry plot in different muscle area conforms to anatomy. **Figure 6** shows the geometry plot of participant S1, and the participants' geometry plots can be checked from **Supplementary Material**. Each geometry plot includes the flexor side (left) and extensor side (right). In the actual multi-array electrode sleeve, the left border of the flexor side is connected to the right border of the extensor side, and the right border of the flexor side is connected to the left border of the extensor side. The WF/WE geometry plot was

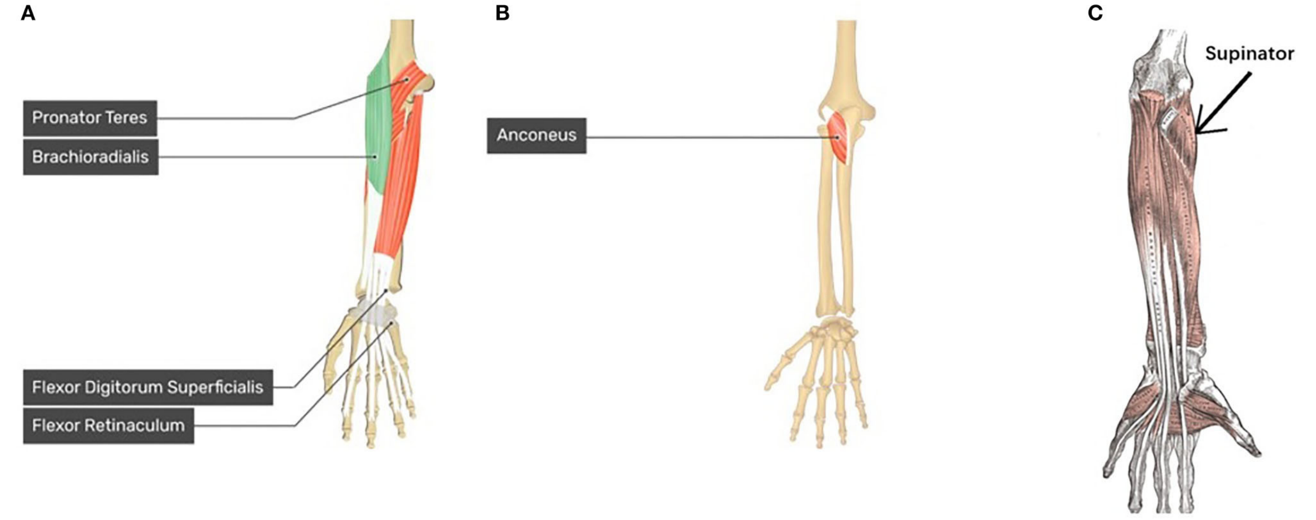


FIGURE 13 | Forearm and Hand anatomy for motion pattern analysis. **(A)** Anatomy of brachioradialis which related to WF motion, pronator teres which related to P motion, and flexor digitorum superficialis (FDS) which related to HG motion (left hand) (GetBodySmart2021, 2021); **(B)** Anatomy of anconeus which related to WE motion (left hand) (GetBodySmart2021, 2021); **(C)** anatomy of supinator related to S motion (left hand) (Kenhub, 2021).

constructed using $\mathbf{W_1}$ from FC layer and those of the P/S and HG/HO using $\mathbf{W_2}$, $\mathbf{W_3}$, respectively. The circles show the motion pattern position; the green circle shows the motion pattern corresponding to the positive joint angles and purple circle shows the motion pattern corresponding to negative joint angles. In the geometry plot, the matrix shows the motion pattern color with the channel index. The geometry plots of the participants (**Figure 6**; **Supplementary Material**) show that all participants show similar motion patterns, although the channels are not the same, and the adjacent area can display the muscle activity. This is because the different sizes of the participant's forearm and dislocation of the multi-array electrode sleeve may lead to this result. Using the anatomy of the forearm (**Figure 13**), the motion patterns can be discussed as follows.

- WF/WE joint: When WF/WE motion is performed, both sides show WF or WE motion patterns. This is an example of muscle contraction and muscle relaxation during muscle activity. S1 and S2 show almost the same motion pattern, and **Table 2** shows that the two participants demonstrate best CC results with the WF motion occurring on channels 13, 21, 29, and 28 near channel 24, while WE motion occurs on channels 6, 14, 12, 20, 25, and 26. S3–S10 show similar patterns but near the above area. For S5, channel 17 is the adjacent area of channel 24; hence, it is a similar motion pattern when compared with the WF pattern shown in S1 and S2 (in this pattern, S3 and S4 shown in channels 15 and 16 represent dislocation of the sleeve).
- P/S joint: When P/S motion is performed, both sides show wrist pronation or supination motion pattern, which is an example of muscle contraction and muscle relaxation during muscle activity. The P/S motion generated from deep layer muscle is compared to WF/WE. The motion pattern may not be representing the correct anatomy of the P/S motion activity in the forearm; however, they show a similar motion pattern.
- HG/HO joint: The difference between the previous motions (WE/WF and P/S) is that HG/HO motion is generated from forearm motion and deeper muscles that lead to finger motion (we can regard hand grip and open as finger motion). The geometry plot shows that the FC layer weights of each participant are quite different. However, we can find a similar motion pattern area from **Supplementary Material**.

Moreover, according to human anatomy (**Figure 13**) and comparing with **Figure 6**, we can discuss the motion pattern using the anatomy of the forearm muscles:

- The WF motion produced by the brachioradialis of the forearm (**Figure 13A**) and WE motion produced by the anconeus of the forearm (**Figure 13B**) correspond to the area of channels 13, 21, 29, and 28 (brachioradialis) and the area of channels 25 and 26 (anconeus), respectively. By comparing **Figure 6** with **Figures 13A,B**, we found that the motion pattern is correct. We may consider another pattern area, such as channels 23, 24, 15, 16, and 6, owing to muscle contraction and relaxation. When we perform the WF and WE motion, these aforementioned areas show have motion activity.
- P motion produced by pronator teres of the forearm (**Figure 13A**) and S motion produced by the supinator of the forearm (**Figure 13C**) correspond to the area near channel 22 and 30 (pronator teres), and the area of channels 13, 21, and 29 (supinator), respectively. Comparing **Figure 6** with **Figures 13A,C** show that the motion patterns are similar to the anatomy results. Because P/S motion contains the interactive movement between two bones, the motion generates the muscle activity on the opposite side, such as channels 17, 18, or

25 (P) and channels 16 or 11 (S). Thus, we consider the motion patterns to be appropriate.

- HG motion is produced mainly by the flexor digitorum superficialis (FDS, **Figure 13A**) and muscles that produce WF, such as brachioradialis. Similar to HG, the HO motion is a complicated motion produced by many muscles, including the anconeus area of the WE. The FDS corresponds to the area around channels 14 and 15 or the adjacent channels. **Figure 6** shows that the motion patterns of the HG of participant S1 are near the FDS area. The patterns occur in the area near channels 24 and 16 (near the WF pattern area). It is difficult to evaluate the motion pattern of HO, but the HO motion pattern shows the area mainly on the extensor side. When the HO motion was performed, these areas show motion activity; thus, we infer that the motion pattern is appropriate.

Limitation of Our Work

In this work, we used a window of 500 ms to segment sEMG signal data flow as CNN input for offline analysis, and we got high CC result after training the regression model using such window length. However, we did not apply it for real-time control. In real-time control, the estimation window lengths should range from 50 ms to 400 ms (Hargrove et al., 2009), 500 ms window might generate delay, we will reduce the window size in our next topic.

CONCLUSION AND FUTURE WORK

In this study: (1) We proposed a CNN-based regression model to estimate 3-DOFs joint angles (WF/WE and P/S as wrist motion, and HG/HO as finger motion) based on sEMG signal, and it performed the highest when we compared it to another regression models; (2) We used transfer learning with small amount of new dataset to make the model can be calibrated in another day. The model comparison result shows that, compared to LR, SVR, KNN and DT, proposed CNN model significantly performs higher than conventional models with and without transfer learning; (3) We tried to find the reason why the proposed model can learn the motion information from muscle, so we design the convolutional filter as CW-CNN filter to obtain force pattern as feature maps, and we tracked back to check the geometry plots to analyze the motion patterns.

In our future work, we would use this model to predict the mentioned 3-DOFs joint angles (WF/WE, P/S and HG/HO) in real time and send the predicted joint angles to the prosthetic hand control system to achieve real-time control of the prosthetic hand.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of the Tokyo Institute of Technology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZQ designed research and experiment, collected data, analyzed the results and completed the paper manuscript. ZH helped how to use the experimental equipment Perception Neuron for data acquisition, and data collection. SS, NY, and YK made discussion together, analyzed the results, and reviewed manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the JSPS KAKENHI (grant no. 19H05728), Tateishi Science and Technology Foundation (grant no. 2188001), JST PRESTO (Precursory Research for Embryonic Science and Technology) (grant no. JPMJPR17JA), and JST MIRAI (grant no. JPMJM18C8).

ACKNOWLEDGMENTS

We thank all the participants who participated in the experiments. And we would like to thank Editage (www.editage.com) for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbot.2021.685961/full#supplementary-material

REFERENCES

Aiuto Co., Ltd. (2020). *E.coli*. Available online at: https://www.aiuto-jp.co.jp/support/file_93.php (accessed March 12, 2021).

Ameri, A., Akhaee, M., Scheme, E., and Englehart, K. (2020). A deep transfer learning approach to reducing the effect of electrode shift in EMG pattern recognition-based control. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 370–379. doi: 10.1109/TNSRE.2019.2962189

Ameri, A., Akhaee, M. A., Scheme, E., and Englehart, K. (2019). Regression convolutional neural network for improved simultaneous EMG control. *J. Neural Eng.* 16:036015. doi: 10.1088/1741-2552/ab0e2e

Angkoon, Phinyomark, Phukpattaranont, P., and Limsakul, C., et al.. (2012). Feature reduction and selection for EMG signal classification. *Expert Syst. Appl.* 39, 7420–7431. doi: 10.1016/j.eswa.2012.01.102

Atzori, M., Cognolato, M., and Müller, H. (2016). Deep learning with convolutional neural networks applied to electromyography data: a resource for the classification of movements for prosthetic hands. *Front. Neurorobot.* 10:9. doi: 10.3389/fnbot.2016.00009

Bai, D., Liu, T., Han, X., Chen, G., Jiang, Y., and Hiroshi, Y. (2021). "Multi-channel sEMG signal gesture recognition based on improved CNN-LSTM hybrid models," in 2021 IEEE International Conference on Intelligence and Safety for Robotics (Tokoname: ISR), 111–116. doi: 10.1109/ISR50024.2021.9419532

Bao, T., Zaidi, S. A. R., Xie, S., et al. (2021a). Inter-subject Domain Adaptation for CNN-based Wrist Kinematics Estimation using sEMG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering. ISSN* 1534–4320. doi: 10.1109/TNSRE.2021.3086401

Bao, T., Zhao, Y., Zaidi, S., Xie, S., Yang, P., and Zhang, Z., et al. (2021b). A deep Kalman filter network for hand kinematics estimation using sEMG. *Pattern Recognit. Lett*. 143, 88–94. doi: 10.1016/j.patrec,.2021.01.001

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Method*. 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Chen, J., Qiu, J., and Ahn, C. (2017). Construction worker's awkward posture recognition through supervised motion tensor decomposition. *Autom. Constr*. 77, 67-81. doi: 10.1016/j.autcon.2017.01.020

Côté-Allard, U., Fall, C.L., Campeau-Lecours, A., Gosselin, C., Laviolette, F., and Gosselin, B., et al. (2017). "Transfer learning for sEMG hand gestures recognition using convolutional neural networks," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* (Banff, AB: IEEE), 1663–1668. doi: 10.1109/S M. C.2017.8122854

Diamantidis, N.A., Karlis, D., and Giakoumakis, E.A. (2000). unsupervised stratification of cross-validation for accuracy estimation. *Artif. Intell*. 116, 1–16. doi: 10.1016/S0004-3702(99)00094-6

Farina, D., Jiang, N., Rehbaum, H., Holobar, A., Graimann, B., Dietl, H., et al. (2014). The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges. *IEEE Trans Neural Syst Rehabil Eng*. 22, 797–809. doi: 10.1109/TNSRE.2014.2305111

GetBodySmart2021 (2021). *E.coli*. Available online at: https://www.getbodysmart.com/arm-muscles/brachioradialis-muscle (accessed January 28, 2021).

Hargrove, L. J., Li, G., Englehart, K. B., and and, Hudgins, B. S. (2009). Principal components analysis preprocessing for improved classification accuracies in pattern-recognition-based myoelectric control. *IEEE Trans. Biomed. Eng*. 56, 1407–1414. doi: 10.1109/TBME.2008.2008171

Hu, Y., Wong, Y., Wei, W., Du, Y., Kankanhalli, M., et al. (2018). A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition. *PLoS ONE* 13:e0206049. doi: 10.1371/journal.pone.0206049

Huang, D., and Chen, B. (2019). "Surface EMG decoding for hand gestures based on spectrogram and CNN-LSTM," in *2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI),* (Xi'an: IEEE), 123–126. doi: 10.1109/CCHI.2019.8901936

Jiang, W., and Yin, Z. (2015). "Human activity recognition using wearable sensors by deep convolutional neural networks," in *Proceedings of the 23rd ACM international conference on Multimedia* (Brisbane, QLD), 1307–1310. doi: 10.1145/2733373.2806333

Kenhub (2021). *E.coli*. Available online at: https://www.kenhub.com/en/library/anatomy/supinator-muscle (accessed January, 28, 2021).

Kim, H.S., Hong, N., Kim, M., Yoon, S.G., Yu, H., Kong, H., et al. (2019). Application of a Perception Neuron® System in Simulation-Based Surgical Training. *J. Clin. Med*. 8:124. doi: 10.3390/jcm8010124

Koch, P., Dreier, M., Larsen, A., Parbs, T. J., Maaß, M., Phan, H., et al. (2020). "Regression of Hand Movements from sEMG Data with Recurrent Neural Networks," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, (Montreal, QC: IEEE), 3783–3787. doi: 10.1109/E M. B. C.44109.2020.9176278

Koike, Y., and Kawato, M. (1995). Estimation of dynamic joint torques and trajectory formation from surface electromyography signals using a neural network model. *Biol. Cybern*. 73, 291–300. doi: 10.1007/BF00199465

Le, Q. V. (2015). A tutorial on deep learning part 2: autoencoders, convolutional neural networks, and recurrent neural networks. *Google Brain* 2015, 1–20. Available online at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.703.5244&rep=rep1&type=pdf

National Academies of Sciences, Engineering, and Medicine (2017). *The Promise of Assistive Technology to Enhance Activity and Work Participation*. Washington, DC: The National Academies Press. doi: 10.17226/24740

Rehman, M., Waris, A., Gilani, S. O., Jochumsen, M., Niazi, I., Jamil, M., et al.. (2018). Multiday EMG-based classification of hand motions with deep learning techniques. *Sensors (Basel, Switzerland)* 18:2497. doi: 10.3390/s18082497

Rodríguez, J. D., Pérez, A., and Lozano, J. A. (2010). Sensitivity analysis of kappa-fold cross validation in prediction error estimation. *IEEE Trans. Pattern Anal. Mach. Intell*. 32, 569–575. doi: 10.1109/TPAMI.2009.187

Sakhavi, S., Guan, C., and Yan, S. (2018). Learning temporal information for brain-computer interface using convolutional neural networks. *IEEE Trans. Neural Netw. Learn. Syst*. 29, 1–11. doi: 10.1109/TNNLS.2018.2789927

Scheme, E., and Englehart, K. (2011). Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use. *J. Rehabil. Res. Dev*. 48, 643–659. doi: 10.1682/JRRD.2010.09.0177

Stapornchaisit, S., Kim, Y., Takagi, A., Yoshimura, N., and Koike, Y. (2019). Finger Angle Estimation from array EMG system using linear regression model with independent component analysis. *Front. Neurorobot*. 13:75. doi: 10.3389/fnbot.2019.00075

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B. Method*. 36, 111–147. doi: 10.1111/j.2517-6161.1974.tb00994.x

Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *J. Diagn. Med. Sonography* 6, 35–39. doi: 10.1177/875647939000600106

Wang, P., Song, Q., Han, H., and Cheng, J. (2016). Sequentially supervised long short-term memory for gesture recognition. *Cogn. Comput*. 8, 982–991. doi: 10.1007/s12559-016-9388-6

Xia, P., Hu, J., and Peng, Y. (2018). EMG-based estimation of limb movement using deep learning with recurrent convolutional neural networks. *Artif. Organs* 42, E67–E77. doi: 10.1111/aor.13004

Yasuharu, K., Yeongdae, K., Sorawit, S., Zixuan, Q., et al. (2020). Development of Multi-sensor array electrodes for measurement of deeper muscle activation. *Sens. Mater*. 32, 959–966. doi: 10.18494/SAM.2020.2636

Yoshikawa, M., Mikawa, M., and Tanaka, K. (2006). "Real-time hand motion estimation using EMG signals with support vector machines," in *2006 SICE-ICASE International Joint Conference* (Busan), 593–598. doi: 10.1109/SICE.2006.315553

Yuanhui, He, and Yufeng, Tang. (2014). *Neuron Data Reader Runtime API Documentation. DocPlayer [Preprint]*. Available online at: https://docplayer.net/42937137-Neuron-data-reader-runtime-api-documentation.html (accessed May 6, 2021).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 15 September 2021
doi: 10.3389/fnbot.2021.711047

Control of Newly-Designed Wearable Robotic Hand Exoskeleton Based on Surface Electromyographic Signals

Ke Li[1]*, *Zhengzhen Li*[1], *Haibin Zeng*[2] *and Na Wei*[3]

[1] *Laboratory of Rehabilitation Engineering, Research Center of Intelligent Medical Engineering, School of Control Science and Engineering, Shandong University, Jinan, China,* [2] *Department of Radiotherapy, Suzhou Dushu Lake Hospital, Suzhou, China,* [3] *Department of Geriatrics, Qilu Hospital, Shandong University, Jinan, China*

The human hand plays a role in a variety of daily activities. This intricate instrument is vulnerable to trauma or neuromuscular disorders. Wearable robotic exoskeletons are an advanced technology with the potential to remarkably promote the recovery of hand function. However, the still face persistent challenges in mechanical and functional integration, with real-time control of the multiactuators in accordance with the motion intentions of the user being a particular sticking point. In this study, we demonstrated a newly-designed wearable robotic hand exoskeleton with multijoints, more degrees of freedom (DOFs), and a larger range of motion (ROM). The exoskeleton hand comprises six linear actuators (two for the thumb and the other four for the fingers) and can realize both independent movements of each digit and coordinative movement involving multiple fingers for grasp and pinch. The kinematic parameters of the hand exoskeleton were analyzed by a motion capture system. The exoskeleton showed higher ROM of the proximal interphalangeal and distal interphalangeal joints compared with the other exoskeletons. Five classifiers including support vector machine (SVM), K-near neighbor (KNN), decision tree (DT), multilayer perceptron (MLP), and multichannel convolutional neural networks (multichannel CNN) were compared for the offline classification. The SVM and KNN had a higher accuracy than the others, reaching up to 99%. For the online classification, three out of the five subjects showed an accuracy of about 80%, and one subject showed an accuracy over 90%. These results suggest that the new wearable exoskeleton could facilitate hand rehabilitation for a larger ROM and higher dexterity and could be controlled according to the motion intention of the subjects.

Keywords: exoskeleton, surface electromyography, gesture recognition, wearable robots, hand rehabilitation

Edited by:
Dingguo Zhang,
University of Bath, United Kingdom

Reviewed by:
Chen Chen,
Shanghai Jiao Tong University, China
Hui Zhou,
Nanjing University of Science and Technology, China

****Correspondence:***
Ke Li
kli@sdu.edu.cn

Received: *17 May 2021*
Accepted: *11 August 2021*
Published: *15 September 2021*

Citation:
Li K, Li Z, Zeng H and Wei N (2021) Control of Newly-Designed Wearable Robotic Hand Exoskeleton Based on Surface Electromyographic Signals. Front. Neurorobot. 15:711047. doi: 10.3389/fnbot.2021.711047

INTRODUCTION

The human hand plays a role in a variety of daily tasks. This delicate instrument is vulnerable to trauma or neurological or musculoskeletal disorders. Stroke, for example, could heavily affect hand function (Hu et al., 2018; Burns et al., 2019; Chowdhury et al., 2019). Over 85% of the post-stroke individuals reported that they could not control their hand freely for dexterous manipulation over 8 months after the onset of stroke. Hand rehabilitation, typically by intensive motor training for restoring hand function, would be one of the most urgent demands in post-stroke survivors, particularly for those who desire to maintain a high quality of life.

Effective rehabilitation requires repetitive passive and active training, which could not be fully conducted face-to-face or hand-by-hand by therapists. A robotic hand exoskeleton can provide high training intensity, stable working performance, and adaptive movement assistance for functional training; thus it has been proven to be an effective technology for hand rehabilitation (Leonardis et al., 2015; Li et al., 2019). However, due to the compact dimensions, structural complexity, and high flexibility, developing a satisfactory hand exoskeleton and to freely control it in real-life scenarios is still a challenging task (Palm and Iliev, 2007).

From structure design, the hand exoskeleton could be divided into pneumatic, glove-based, and linkage-based exoskeletons. The pneumatic exoskeleton is easy to control, but is difficult to perform flexible finger movement (Gerez et al., 2020; Takahashi et al., 2020). Glove-based exoskeletons are usually more supple and comfortable to wear; but due to the coverage of the glove, it blocks the direct contact of the object and, thus, disturbs tactile feedback (Sarac et al., 2019). The linkage-based exoskeletons use mechanical links to connect the finger components, either by fingertip contact or by full contact with the hand. With fingertip contact, the advantage is that the position of the fingertip can be precisely controlled and, thus, suitable for all hand sizes; but the disadvantage is that the contact areas between the exoskeleton and the fingertip are quite limited and it is, thereby difficult to generate high force output. By full contact there are larger contact surfaces and closer interactions between the human hand and the exoskeleton, enabling more powerful assistance, enhanced working stability, and comfortable feeling during hand rehabilitation. Most existing hand exoskeletons have drawbacks in the clumsy control of the thumb. The human hand has a flexible thumb but most of the hand exoskeletons use only one actuator to control thumb movement (Li et al., 2016; Burns et al., 2019; Gasser et al., 2020). The flexion/extension of the fingers and the abduction/adduction of the thumb are two of the most important exercises to improve hand function (Gerez et al., 2020). An intriguing issue is how to design an exoskeleton with more than one actuator for the thumb to realize both flexion/extension and abduction/adduction.

Another issue is how to control the hand exoskeleton according to the movement intention of the patients. Surface electromyography (sEMG) is a non-invasive technology, which has been widely used in human-robot interaction and clinical examinations (Cote-Allard et al., 2019). The sEMG reflects muscle activations under the modulation of the central nervous system. Recording and processing of sEMG signals may help identify motion intention and provide key information for real-time control of hand exoskeleton (Li et al., 2018). Most exoskeleton hands controlled by sEMG adopt the strategy of mirror therapy principle, which suggests decoding the motor intention of the stroke patients from the non-paretic muscles because of their relatively normal function (Emerson et al., 2016). Several algorithms have been developed for sEMG processing, motion intention decoding, and exoskeleton control. But it is still challenging to realize online sEMG processing, time-varied motion intention, and real-time control for multiactuator exoskeletons.

This study presents a newly-designed wearable robotic hand exoskeleton with more active degrees of freedom (DOFs), larger range of motions (ROMs) for most joints, and the capability of being freely controlled by motion intention. The human finger is in full contact with the mechanical shells, and totally six linear actuators are adopted, generating high output forces for each digit. The thumb is controlled by two actuators to perform circumduction and adduction/abduction. The proximal interphalangeal (PIP) and distal interphalangeal (DIP) joints are driven by coupled links. Motion intentions for controlling the exoskeleton were decoded by the sEMG from the muscles of the non-paretic arm and hand. Patients could use this hand exoskeleton for repetitive training of grasping, pinching, individual finger control, thumb adduction/abduction, and thumb circumduction.

METHOD

Hand Exoskeleton Design

The structure of the hand exoskeleton is shown in **Figure 1**. Thumb and fingers were connected to the palm back platform through a linkage. The finger was designed ergonomically following finger motion trajectories. The design allowed independent movement of the thumb and the four fingers, and circumduction, abduction, and the palm opposite for the thumb. Each finger was driven by a linear actuator, and the thumb was driven by two linear actuators. The mechanical structure was designed and simulated using Solidworks (Dassault Systems, USA), and was made from resin by 3D printing.

The palm back shell covering the wrist and the back of the hand was made of thermoplastic materials. The blank was first shaped and was then soaked in hot water to be softened up, fitting it to the hand of the users. The exoskeleton weighs 500 g, was fixed upon the palm back shell, and is convenient to wear. Six linear actuators (Actuonix, L12-50-210-12-p) with matching control boards were applied to drive the exoskeleton. The operating distance of the linear actuators was controlled by voltages. The linear actuator with a length of 102 mm and weight of 40 g can be bidirectionally driven.

Each finger has three shells, connected by a linkage rod. As shown in **Figure 1B**, when the linear actuator reciprocates, the linkage rod transmits force to the finger shells, driving a motion for abduction/adduction. The exoskeleton fingers were designed following the anatomic characteristics of human fingers. Considering the thumb has higher DOFs than the fingers, the newly-designed exoskeleton adopted a more flexible structure for the thumb that can facilitate the thumb for inward rotation, grasping, and abduction. The force of the actuator can directly act on the carpometacarpal joint (CMC) or the metacarpophalangeal joint (MCP). When the force acts on the CMC joint, the wrist and palm joints move first, and as the CMC joint contacts the object, the movement of the CMC joint can be blocked so that the control of the tip of the thumb is not flexible enough. When the force acts on the MCP joint, the flexibility of the thumb could be increased, but this design is

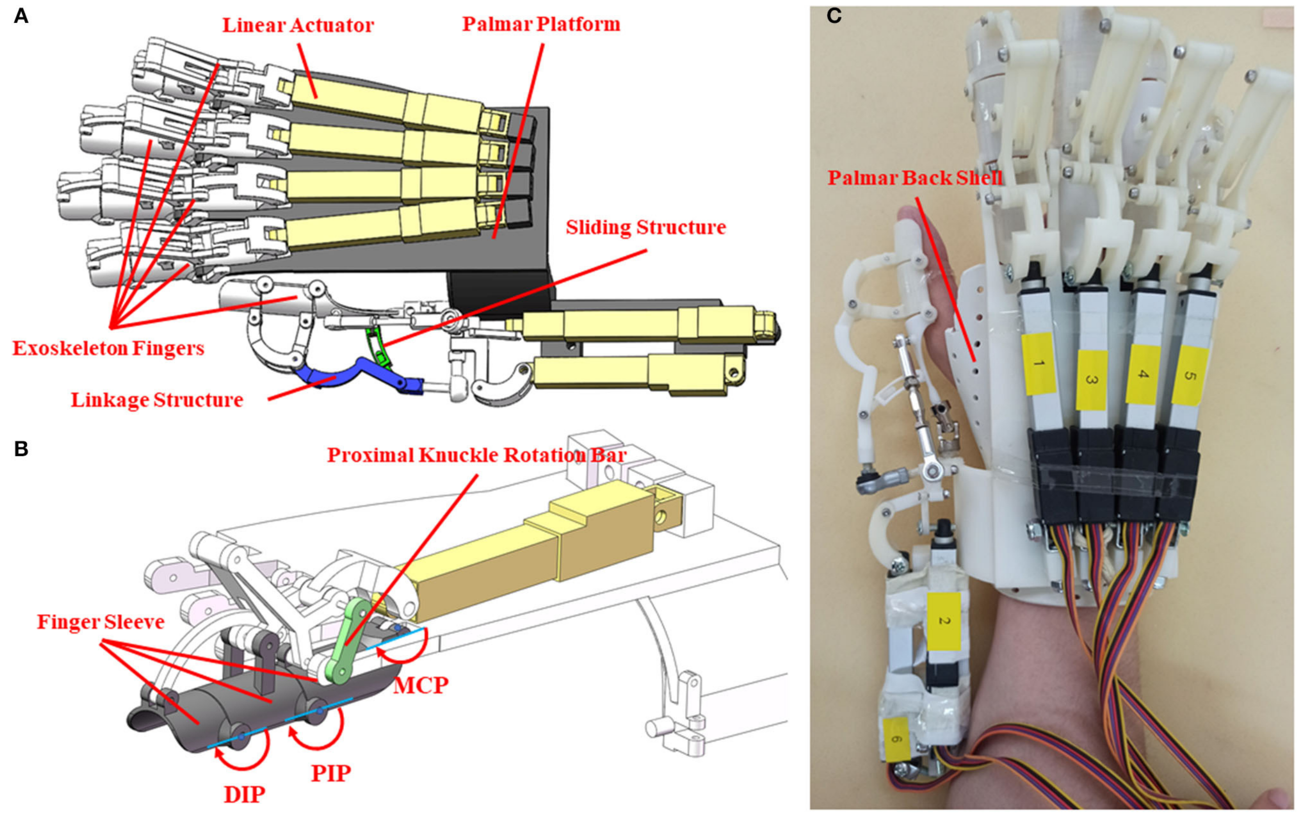


FIGURE 1 | The mechanical design and realization of hand exoskeleton. **(A)** The mechanical design of the exoskeleton; **(B)** mechanical structure of exoskeleton index finger; **(C)** prototype of the hand exoskeleton.

not suitable for grasping relatively bigger objects. The newly-designed exoskeleton did not follow any of these designs. Instead, the force of the new exoskeleton acts on the MCP and PIP joints of the thumb. The joint movement of CMC was driven by the friction of the sliding structure. The ROMs of the MCP and PIP joints were confined to avoid bending the grasped object. When the sliding structure moves, the CMC joint can be pushed forward until blocked by the grasping object. Then the force of the linear actuator accumulates on the MCP and PIP joints. The sliding part of the linkage structure helps extend the ROM of the thumb. Two linear actuators were used to control the movement of the thumb in the vertical planes so that the thumb can complete a circumduction. The schema of the control system for the exoskeleton is shown in **Figure 2**. This exoskeleton can be controlled by sEMG of the non-paretic forearm and hand following the mirror therapy principle. To examine the performance of the hand exoskeleton, the classifiers, and the real-time control, we set a sequence of experiments (specified in Experiment for hand exoskeleton performance to Experiment for real-time control of an exoskeleton). All these experiments were approved by the Institutional Review Board of Shandong University and were in accordance with the Declaration of Helsinki.

Experiment for Hand Exoskeleton Performance

The experiment was performed to examine the performance of the hand exoskeleton. A 3D motion capture system (Opti Track Motive, USA) with six cameras and clusters of reflective markers was used to collect the movement trajectories of the robotic finger joints. There were five reflective markers attached to the exoskeleton fingertips, six markers on the joints of the exoskeleton fingers, and three marks for coordinates of linear actuators (**Figures 3A,B**). Initially, all the digits of the exoskeleton were fully extended. Once a start command was received, the exoskeleton ran for a full movement cycle, that is, the thumb and the four fingers flexed to their extreme positions and then extended back to the initial positions. The test process consists of five cycles (**Figures 3A,B**). A representative subject participated in the experiment. The subject was instructed to perform flexion and extension of the thumb and fingers for five tries. The marker sets were demonstrated in **Figure 3C**. Specifically, five markers were attached to the fingertips, and six markers were attached to the joints of the index finger and little finger (**Figure 3C**). The movement trajectories of the fingertips and the index finger joint angles of the exoskeleton were calculated to evaluate the ROM of the exoskeleton. To

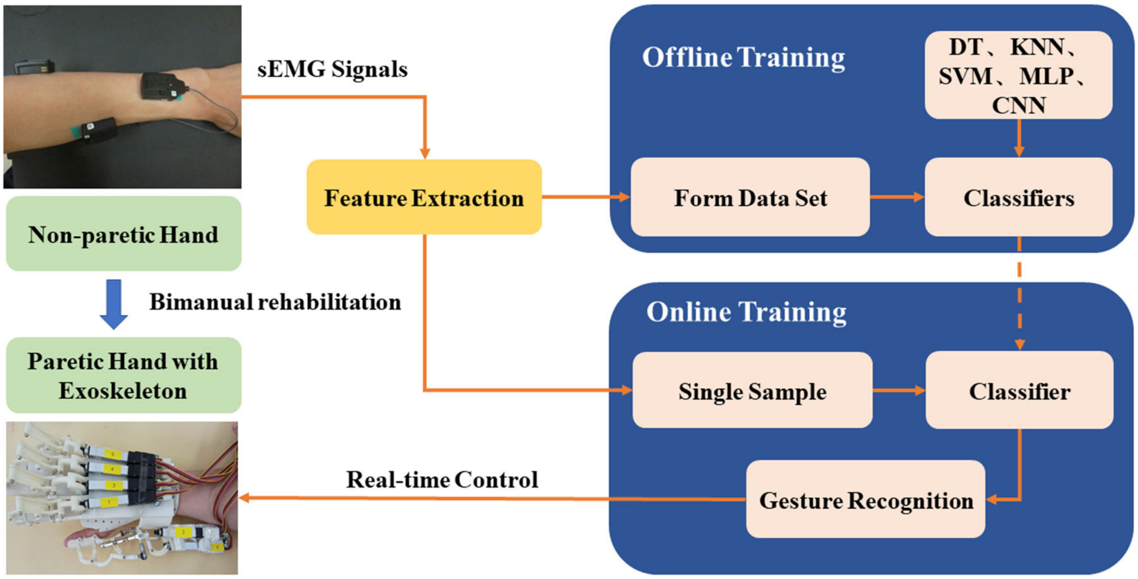


FIGURE 2 | Schema of the control system for the hand exoskeleton based on non-paretic sEMG processing and offline-to-online classifications.

calculate the joint angle, two adjacent markers formed a vector, and the joint angle was the angle between the two adjacent vectors. The ROMs were computed as the changes of joint angles from maximal flexion to full extension and compared with existing exoskeletons.

Experiment for Classifier Selection

Twenty-five healthy right-handed subjects (age = 23.2 ± 1.7 y, 12 women and 13 men) participated in the experiment. The subjects were informed of the purpose of the experiment and were given an informed consent form before the experiment. In total 16 muscles of the left and right hands were selected for the sEMG analysis. The muscles included the following: brachioradialis (BRA), flexor carpi ulnaris (FCU), flexor carpi radialis (FCR), extensor digitorum communis (EDC), flexor digitorum superficialis (FDS), abductor pollicis brevis (APB), first dorsal interosseous (FDI), and abductor digiti minimi (ADM), for both the left and right hands (**Figure 4A**). The sensors were attached to the muscle bellies, and the skin was cleaned with scrub and medical alcohol before attachment. The sEMG signals of the 16 muscles were recorded using the Trigno$^{\text{TM}}$ wireless EMG system (Delsys, USA) at a sampling frequency of 1,000 Hz.

Four hand gestures, clenched fist (CF), thumb opposition (TO), key pinch (KP), and three fingers flexion (TFF) were selected as representative gestures in this experiment (**Figure 4B**). These gestures have been commonly used in hand gesture recognition in human-robot interaction and rehabilitation (He et al., 2017; Yu et al., 2018). The initial position for the four gestures was that the thumb should keep full extension whereas the fingers should be flexed. After hearing a command, subjects were instructed to perform a gesture with both hands and maintain this gesture for 8 s without much effort. The performances of the four gestures were randomized. Each gesture was repeated for 10 trials, with a rest period of 7–10 s between trials and 1–2 min between gestures.

To better understand the muscle activations of different gestures and provide a reference for muscle selection, a co-contraction index (CI) was calculated as follows:

$$CI = \frac{1}{T}\int_{T} A_{ij}(t)dt \tag{1}$$

where, A_{ij} is the overlapping of sEMG envelopes of muscle i and muscle j, and T is the length of the sEMG envelope (Frost et al., 1997). The CI represents the level of the common contraction phase of the two muscles, ranging from 0 (no overlap) to 1 (full overlap) (Hu et al., 2009). The length of the sEMG segment was 500 ms, taken from the initial part of the datasets for each gesture. The sEMG signals subtracted the SDs of the envelopes of the averaged sEMG signals at the relaxed state and were then low-pass filtered using a fourth-order Butterworth filter with a cutoff frequency of 10 Hz. The data were standardized according to the maximum and minimum values of the data segments. The averaged CIs of the muscle pairs were calculated for each gesture, and then an 8×8 CI matrix was formed for the total eight muscles.

Five classifiers were used for hand gesture recognition. The sEMG signals were recorded from the BRA, FCU, FCR, EDC, FDS, APB, FDI, and ADM of the left forearm and hand. Five classifiers including support vector machine (SVM), K-near neighbor (KNN), decision tree (DT), multilayer perceptron (MLP), multichannel convolutional neural networks (multichannel CNN) were applied for classifying the CF, TO, KP, and TFF. A sliding window was used to extract features from the original signals. The window width was 128 ms and the sliding

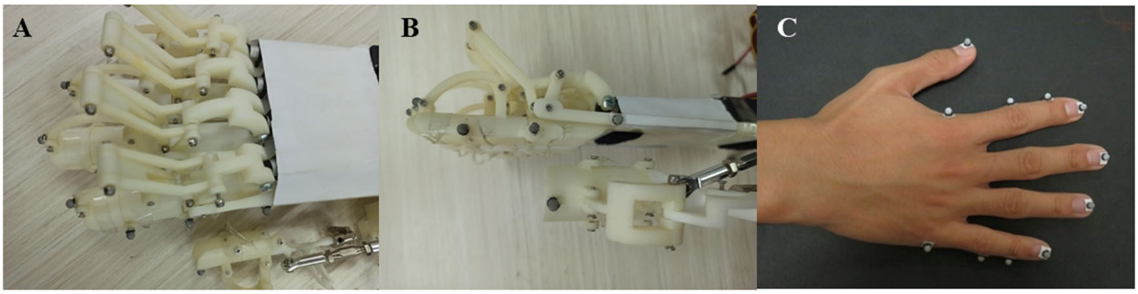


FIGURE 3 | The reflective marker sets for kinematic analysis of the hand exoskeleton. **(A)** The top view of the marker sets; **(B)** the profile view of the marker sets; **(C)** the marker sets for the human hand.

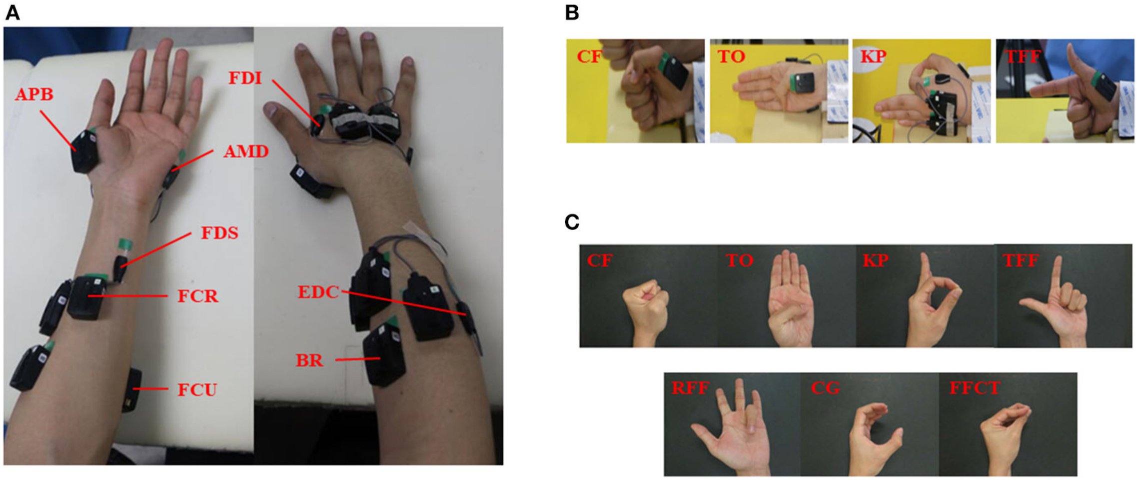


FIGURE 4 | The muscles selection and hand gesture recognition. **(A)** The eight muscles of the bilateral forearms and hands; **(B)** the gestures for classifier selection; **(C)** the gestures for real-time control of exoskeleton.

distance was 78 ms. We selected 13 specific features including zero-crossings (ZC), root mean square (RMS), mean absolute value (MAV), waveform length (WL), variance (VAR), slope sign change (SSC), Willison amplitude (WAMP), mean value (MEAN), the standard value (STD), mean frequency (MNF), median frequency (MF), mean power frequency (MPF), and Lempel-Ziv complexity (LZC). The ZC, RMS, MAV, WL, VAR, SSC, and WAMP have commonly used time-domain features (Hua et al., 2020; Qu et al., 2020; Wu et al., 2020; Duan et al., 2021). In addition to these parameters, we further calculated the MEAN and STD, so that the time-domain features reached nine. Considering the limitation that the time-domain parameters are vulnerable to the noise or interference to amplitudes, the frequency-domain features were selected, including MNF, MF, and MPF. Furthermore, LZC was selected to examine the non-linear characteristics of the signals. All these parameters could provide abundant information with low computing costs and good real-time performance. Among all the sEMG data, we used 80% of the data as the training set and the other 20% for verification. The classifiers were trained based on the sEMG signals for each subject individually.

The SVM with a linear kernel that could easily deal with high dimensional representation was used (De Smedt et al., 2019). The details of SVM are as follows: a one-vs-rest strategy was used and a G-binary-classifier was obtained, where G was the number of different gestures in the experiment. The KNN is a machine learning classification algorithm, which calculates the distance to all the training samples, and selects the k-closest samples (Amin et al., 2019). The DT builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated DT is incrementally developed. According to the DT algorithm, the final result is a tree with a great many decision nodes (Shengchang et al., 2017). In this study, the split criterion was the Gini index and the maximum number of splits was 100.

The network of MLP included an input layer with 104 nodes, two hidden layers constructed by 40 nodes for each layer, and an output layer with six nodes. The MLP followed a back propagation (BP) algorithm. The infrastructure of CNN included a convolution layer and a fully connected layer. The sEMG signals were first transformed into a series of images for the CNN network. For the eight muscles, each sEMG was segmented

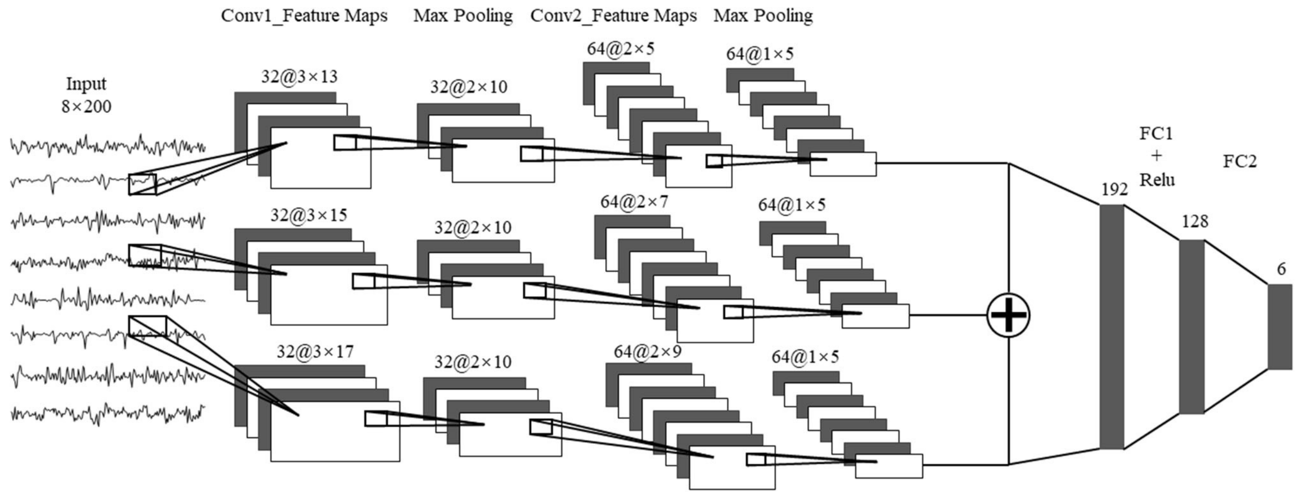


FIGURE 5 | The multi-channel CNN algorithm. Conv1, convolution layer 1; Conv2, convolution Layer 2; FC1, fully connected layer 1; FC2, fully connected layer 2.

with 200-ms windows with a sliding distance of 150 ms, thereby generating 200 × 8 images. The images were input into the convolutional layer to extract features and, thus, reduce the dimension followed by a classification from the fully connected layer. The CNN network could be expanded to multiple channels, which could extract more patterns out of the sEMG signals. The structure of multichannel CNN was shown in **Figure 5**. Each channel has two convolutional layers and two pooling layers, with different convolution kernels. A max-pooling was used for the CNN to reduce the dimension of the convolutional features. A rectified linear unit (ReLU) was used as an activation function for the fully connected layer. The parameter of the dropout layer was 0.5 to avoid over-fitting by randomly deleting the redundant parts of the hidden layer.

Experiment for Real-Time Control of the Exoskeleton

Five healthy subjects (age: 23.4 ± 1.9 year, one woman and four men) participated in the experiment. Six wireless sEMG sensors were attached to six muscles of their left hand, including the EDC, FCU, FCR, BRA, FDS, and FDI. Subjects sat in front of a testing table with their left hand laid on the table in a relaxed state. In total seven hand gestures (as shown in **Figure 4C**) were tested, including CF, TO, KP, and TFF. Four gestures were performed in the prior offline tests, and three new gestures which support grasping and manipulation and are commonly used in hand rehabilitation training were added: ring finger flexion (RFF), cylindrical grip (CG) and fingertips closed together (FFCT) (Zheng et al., 2011; Chen et al., 2017; He et al., 2017; Yu et al., 2018). The performance of each gesture included three phases: a 3-s relax, a 77-s action phase, and a 3-s relax phase. When performing the hand gestures, the subject was not allowed to produce high-level force to maintain the gesture. Subjects were given enough time for rest between trials and sessions. Each gesture was repeated for three trials. After the classifier was trained through offline classification, subjects were instructed to perform two trials for the real-time classification.

The raw sEMG signals were recorded simultaneously at a sampling frequency of 1,000 Hz. The sEMG signals were analyzed using the siding window technique that the window size was 128 ms without overlap at the adjacent windows. About 54 classification results were achieved in each trial. The same 13 features as the abovementioned offline classification were selected. The SVM was used for classification. To distinguish the relaxation state from the gesture execution state, a threshold was set for the relaxation state using the absolute value of Teager Kaiser energy (TKE) (Solnik et al., 2010). The formula for calculating the absolute values after TKE treatment is as follows:

$$x_n = |x_n^2 - x_{n-1}x_{n+1}| \tag{2}$$

where x_n is the sample point of sEMG signals, x_{n-1} and x_{n+1}, are the former and the latter sampling points, respectively. To increase the classification performance, the classification results were verified three consecutive times. If the results were the same all three times, then the classification results could be accepted. Otherwise, the classification analysis was performed again. The programs for sEMG processing, classifiers, and real-time control were realized using MATLAB (MathWorks, USA).

RESULTS

Figures 6A,B demonstrate the joint angles of the MCP, PIP, and DIP of the exoskeleton and the index finger of a human subject during flexion and extension, respectively. Results showed that the MCP, PIP, and DIP joints of the exoskeleton index finger at maximal flexion were 171.26°, 119.57°, and 157.56°, respectively. The ROMs of the MCP, PIP, and DIP joints of the exoskeleton index finger were 8.74°, 60.43°, and 22.44°, respectively. Variations of the joint angles were small, showing

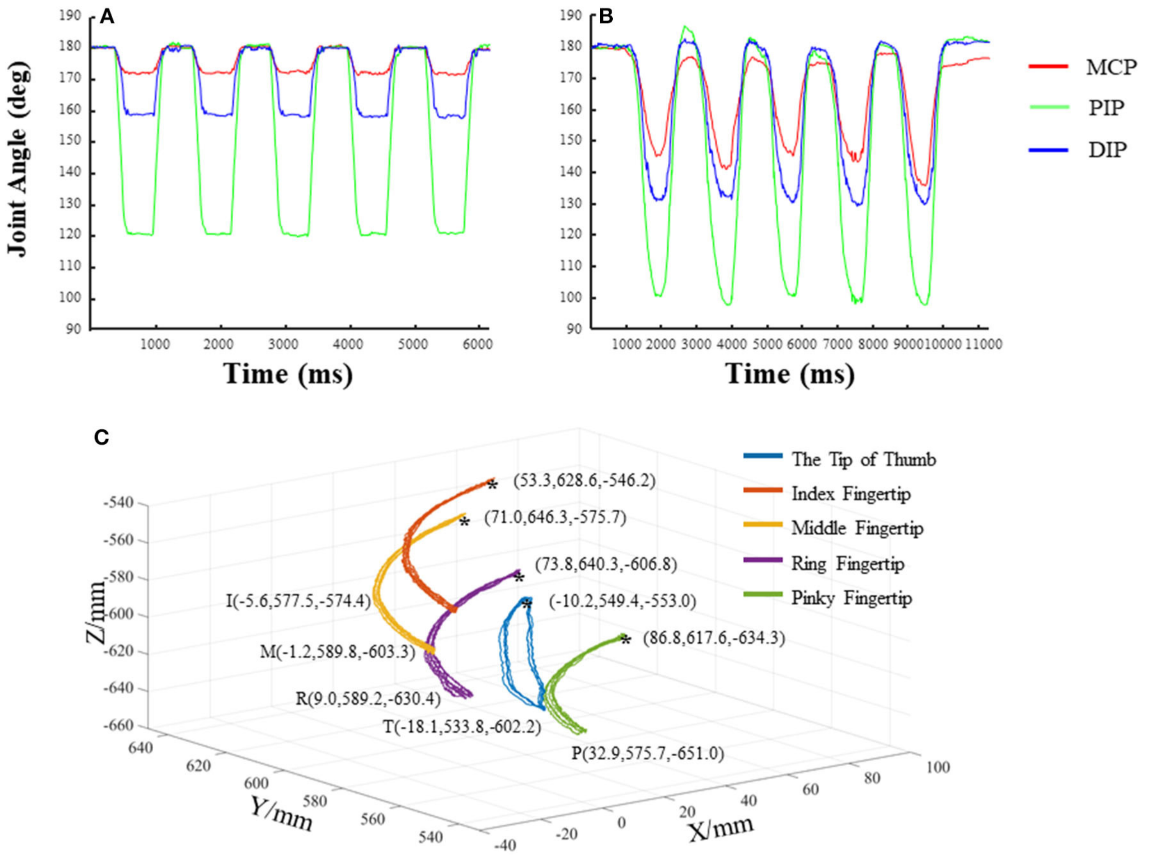


FIGURE 6 | The joint angles and trajectories of the exoskeleton and human digits during flexion and extension. **(A)** The joint angles of the metacarpophalangeal joint (MCP), PIP, and DIP of the exoskeleton index finger; **(B)** the joint angles of the MCP, PIP, and DIP of the index finger of a representative subject; **(C)** the trajectories of the exoskeleton fingertips.

good repeatability. The ROMs of the exoskeleton of the current study were compared with those of previous studies (**Table 1**) (Iqbal et al., 2014; Kim et al., 2017; Refour et al., 2019). The PIP and DIP joints of the new exoskeleton showed larger ROMs, but its MCP joint showed smaller ROM than those of the previously designed exoskeletons. The new exoskeleton set two activate DOFs for the thumb, with one more DOF for the thumb than the previously designed exoskeleton. This new design enables thumb circumduction, in addition to the adduction/abduction for the thumb, compared with the previous exoskeleton. **Figure 6C** shows the movement trajectory of fingertips during digit flexion and extension of the hand exoskeleton. The trajectories of the five testing cycles overlapped, demonstrating a relatively stable motion performance and good repeatability from trial to trial. The trajectory of the thumb presents that the exoskeleton could assist the thumb for circumduction (**Figure 6C**).

Results of sEMG analyses in **Figure 7** showed the following: (a) the averaged coordination matrices of the same hand gesture were similar; (b) the coordination matrices among the forearm muscles showed constant values across the different target gestures than those among the hand muscles; and (c) the matrices among the BRA, FCU, EDC, and FDS were different from those among the muscles of the APB, FDI, and ADM. The CI values of the forearm muscles were higher while performing the CF and TFF gestures compared with the other gestures. The CI values between the APB and the other muscles were lower for the TFF, suggesting that lower intermuscular coordination than the other muscle pairs for this gesture. The FDS had similar results as performing the TO gesture. The BRA and EDC showed greater values for CF, TO, KP, and TFF. There were high correlation coefficients (which were 0.935, 0.903, 0.928, and 0.978 for CF, TO, KP, and TFF) between the coordination matrices of the left and right hands for the same hand gestures, indicating that the muscles of both hands were activated following similar patterns.

Results of offline classification accuracies for the 25 subjects are shown in **Figure 8A**. The SVM and KNN showed classification accuracies up to 99%; the MLP and DT showed classification accuracy over 95%; whereas the CNN showed accuracy just over 80%. Results of real-time control are demonstrated in **Figure 8B**. The offline classification achieved much higher accuracies than the online classification among the five subjects. The online classification accuracy of a subject (H4 subject in **Figure 8B**) was higher than 90%. The online

TABLE 1 | Comparison between the new exoskeleton and some previous results.

Exoskeletons	Finger Number	Total Activate DOFs	Activate DOFs of Thumb	ROM of Index Finger		
				MCP	PIP	DIP
This study	5	6	2	8.74°	60.43°	22.44°
Iqbal et al. (2014)	2	2	1	<30°	NA	NA
Refour et al. (2019)	5	6	1	35°	21°	13°
Kim et al. (2017)	5	5	1	≈13°	≈21°	<5°

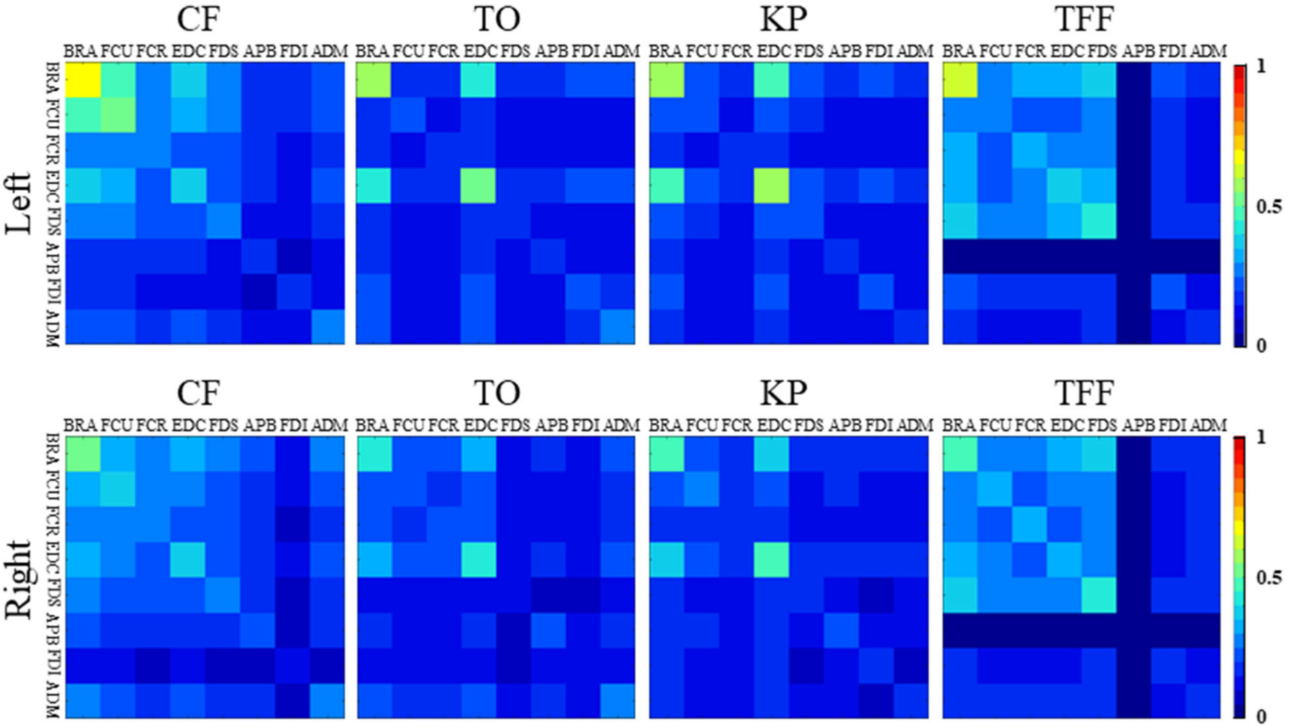


FIGURE 7 | The co-contraction index (CI) matrices for the four gestures of the left and right hands.

classification accuracies for the H2, H3, and H5 subjects were about 80%, but the accuracy for the H1 subject was lower than 50%. **Figure 9** shows the classification accuracies of the seven hand gestures for the five subjects. In general, the classification accuracies for H3, H4, and H5 were better than the other subjects, and the TO and CG gestures achieved better classification performance. Recognitions for FFCT, RFF, and KP gestures of the four subjects were higher than the other gestures except for H1. **Figure 10** demonstrates the original sEMG signals and the TKE signals of the BRA from a representative subject (H4). The final classification based on TKE signals and the three consecutive judgment algorithms had better performance than the classification based on the raw sEMG signals. The threshold was 0.005 mv. By removing the peaks out of the original classification results, although the calculation time for gesture classification increased about 300 ms, the final classification results showed better reliability than the raw classification. The real-time control of the exoskeleton used motion intention extracted from sEMG is shown in a video; URL: https://figshare.com/s/b3a2a1f3ac43172aba76.

DISCUSSION

In this study, we developed a novel wearable robotic hand exoskeleton with multijoints, more active DOFs, larger ROMs for most joints, and the capability of being freely controlled by the motion intention. This hand exoskeleton is capable of driving the thumb and four fingers independently and meets the needs of hand function rehabilitation. Two linear actuators drive the exoskeleton thumb, facilitating a more natural movement of the thumb of the patients for abduction/adduction and circumduction. In addition, the mechanical structure of the exoskeleton could realize the hand functions such as grip and pinch. Finger circumduction is the most difficult movement in exoskeleton design because it requires the coordination of two actuators in different directions. The changes of the exoskeleton

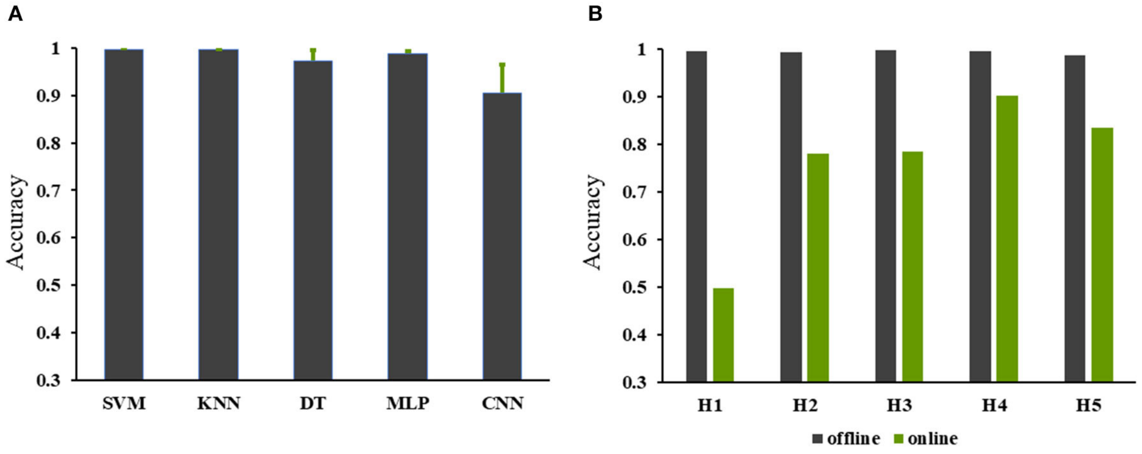


FIGURE 8 | The online and offline classification accuracies. **(A)** The accuracies of the offline classification for classifier selection; **(B)** the accuracies of the online and offline classifications for real-time control of exoskeleton using the support vector machine (SVM).

index finger angles in five cycles were similar to those of a human hand. The similarity of the movement of the five cycles was very high, indicating that the exoskeleton movement could be reliable for motion training.

The exoskeleton in this study showed higher ROMs of the PIP and DIP joints compared with the other exoskeletons, except for the MCP joint that showed a lower ROM than the previously designed exoskeletons. This design was inspired by the human hand anatomic and functional characteristics. Traditionally, exoskeletons mainly exert force on the MCP joint, resulting in relatively larger ROM for the MCP but smaller ROM for the PIP and DIP. During human grasping, the ROM of the PIP and DIP joints are relatively higher than those of the MCP. We, thus, increased the ROMs of the PIP and DIP joints but restricted that of the MCP joint. Although the ROM of the MCP joint of the new exoskeleton was smaller than the other exoskeletons, changes of joint angles during flexion and extension were similar to the human hand.

Results of hand exoskeleton control based on the sEMG signals showed that the classification accuracies were high. Specifically, the H2 subject showed high classification accuracies for the FFCT, CG, RFF, KP, and TO; the H3 subject had high accuracies for the CF,FFCT,TFF,RFF, and TO; the H4 subject had high accuracy for the FFCT, TFF, RFF, KP, and TO; and the H5 subject showed high accuracies for the CFF, FFCT, CG, RFF, and KP. However, for each subject, there were one or two actions that could not reach a high accuracy, such as the CF and TFF for H2, KP for H3, TO for H5, which decreased the overall classification accuracies. The accuracy of online classification for H1 was not ideal, but consistent with the previous studies where there were still three actions (TO, CG, KP) that had high accuracies (Furui et al., 2019). The classification accuracies for each subject were not the same, and thus, different classifiers should be selected and applied individually (Xiloyannis et al., 2017; Dwivedi et al., 2019). For the same subject, the accuracies for classifying different actions could also be quite different. There were between-subject differences in muscle contractions, suggesting that each individual may perform the same action by activating the muscles in quite a different way. The accuracy of predicting the same hand gesture could be different for different individuals (Cote-Allard et al., 2019; Parajuli et al., 2019).

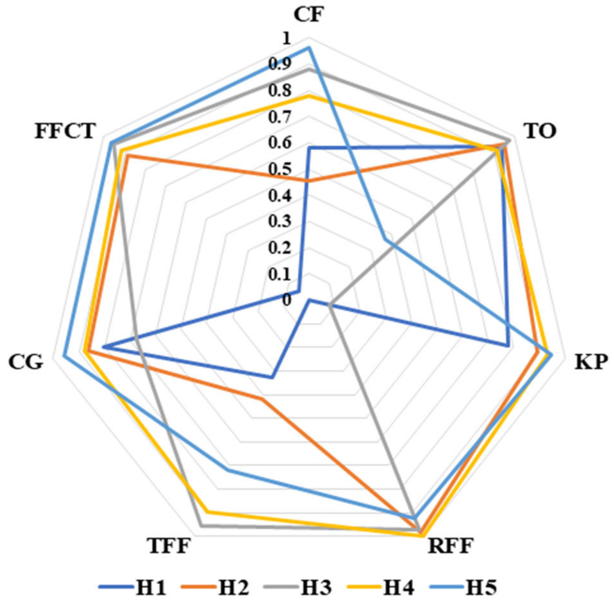


FIGURE 9 | The subject-specific online classification accuracies for recognizing the hand gestures.

According to the results of real-time control, the online accuracy was not as high as the offline classification, which is

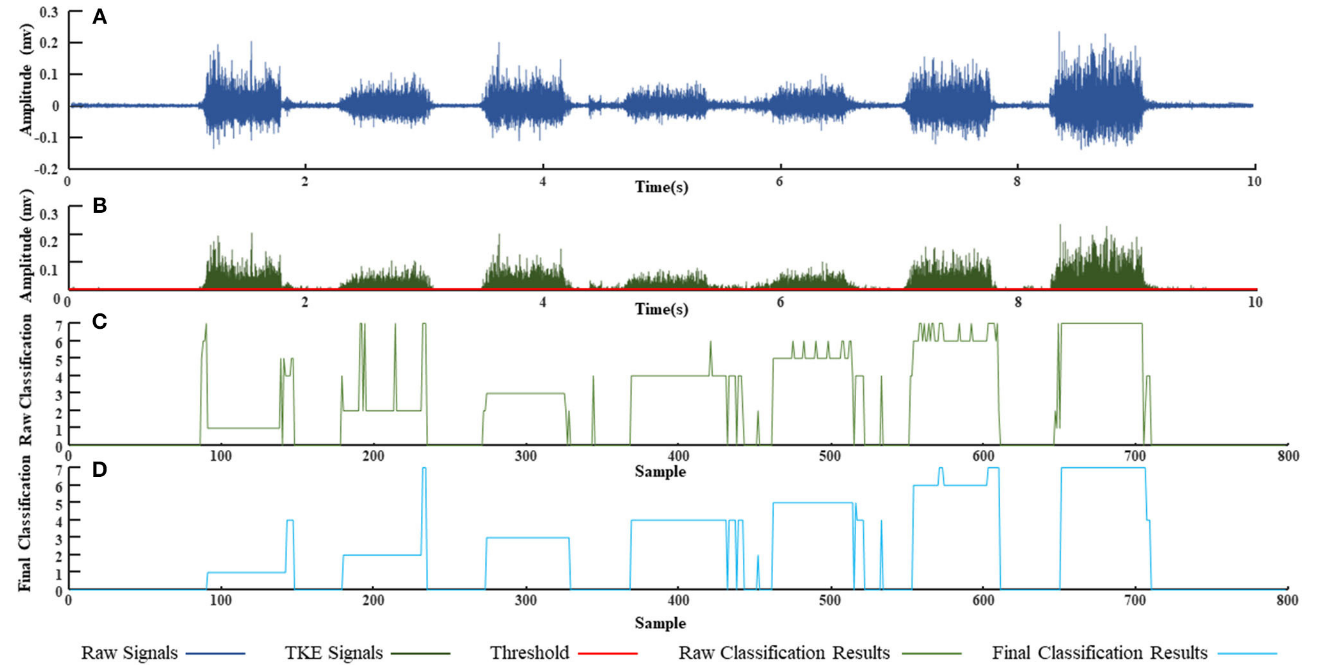


FIGURE 10 | Real-time hand gesture recognition from a representative subject (H4). **(A)** The raw surface electromyography (sEMG) signals recorded from the left brachioradialis (BRA); **(B)** the Teager Kaiser energy (TKE) signals extracted from **(A)** with a threshold 0.001 mv; **(C)** the raw classification based on the SVM; **(D)** final classification based on the TKE and the three consecutive judgment algorithms.

consistent with the previous studies showing that the online accuracy was a more challenging issue of the sEMG controlled hand exoskeleton (Chen et al., 2019; Parajuli et al., 2019). Previous studies showed that in the process of stable grasping, compared with the muscles innervated by different nerves, the muscles innervated by the same nerve showed lower sEMG signal coherence (Pasluosta et al., 2013). For the muscles selected in this study, the BRA and EDC were innervated by the radial nerve, and the FCU and FDI were innervated by the ulnar nerve. Thus, the FCR, FDS, APB, and ADM were innervated by the median nerve. Different from the previous studies, the current study computed the first 500 ms datasets in the execution process instead of the stable grasping data. For all gestures, muscles innervated by the same nerve showed lower CI values compared with the muscles innervated by different nerves (**Figure 7**).

Considering most hand usage in daily activities was under visual supervision, visual feedback was not removed from the experiment. Also, because all the subjects equally received visual feedback during hand performance, the potential effects of visual feedback on results could be further limited. The objective of the current study was to demonstrate a newly-designed wearable robotic hand exoskeleton with more active DOFs, larger ROMs for most joints, and the capability of being freely controlled by motion intention. However, because this is a preliminary study showing a novel design of an exoskeleton, more work is needed prior to any clinical tests. We aim to perform a clinical study in the near future to show the performance of this new exoskeleton for patients with neuromuscular disorders, such as in stroke patients.

CONCLUSION

In this study, we developed a new wearable robotic hand exoskeleton with multiple joints, more DOFs for the thumb, and larger ROM. We also investigated the control of the hand exoskeleton based on the sEMG signals. The former provides a platform and the later builds up its control system. Considering that the post-stroke patients have difficulty in controlling their paretic hands, we adopted the strategy of mirror therapy principle, by which the motion intention was decoded based on the sEMG signals of the non-paretic upper limb and hand. We applied machine learning and deep learning methods to verify the sEMG offline classification. The exoskeleton engaged six linear actuators, in which two were for the thumb and four for the fingers, and can realize independent movement by each digit and the coordinative movement by multiple fingers for grasp and pinch. The joint angles of the exoskeleton index finger were comparable to those of the human index finger, and the circumduction of the thumb was maintained stably. For the real-time control, three out of the five subjects showed an accuracy of about 80%, and one subject showed an accuracy over 90%. The control strategy based on sEMG classification has been integrated with the newly-designed exoskeleton system. This new wearable exoskeleton may play a role in hand rehabilitation in post-stroke patients and may advance the dexterous exoskeleton control according to the motion intention of the patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: URL: https://figshare.com/s/b3a2a1f3ac43172aba76.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Shandong University Qilu Hospital Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KL: original motivation and idea, resources and financial support, design and realization, and writing and revision of the manuscript. ZL: writing of the manuscript, control of hand exoskeleton, signal processing, and classification. HZ: design of the hand exoskeleton, experiment and data analysis, and writing of the manuscript. NW: resources, financial support, and revision of the manuscript. All the authors contributed to the article and approved the submission.

FUNDING

This study was supported by the National Natural Science Foundation of China (62073195), the National Key Research and Development Program (2020YFC2007904), and the Key Research & Development Programs of Guangdong Province (2020B0909020004) and Shandong Province (2019GSF108164, 2019GSF108127 and 2019JZZY021010).

ACKNOWLEDGMENTS

The authors appreciated the participation of all the subjects in the experiment.

REFERENCES

Amin, M. G., Zeng, Z., and Shan, T. (2019). "Hand gesture recognition based on radar micro-doppler signature envelopes," in *2019 IEEE Radar Conference* (Boston, MA). doi: 10.1109/RADAR.2019.8835661

Burns, M. K., Pei, D., and Vinjamuri, R. (2019). Myoelectric control of a soft hand exoskeleton using kinematic synergies. *IEEE Trans. Biomed. Circuits Syst.* 13, 1351–1361. doi: 10.1109/TBCAS.2019.2950145

Chen, C., Chai, G., Guo, W., Sheng, X., Farina, D., and Zhu, X. (2019). Prediction of finger kinematics from discharge timings of motor units: implications for intuitive control of myoelectric prostheses. *J. Neural Eng.* 16:026005doi: 10.1088/1741-2552/aaf4c3

Chen, M., Cheng, L., Huang, F., Yan, Y., and Hou, Z. G. (2017). "Towards robot-assisted post-stroke hand rehabilitation: fugl-meyer gesture recognition using sEMG," in *2017 Ieee 7th Annual International Conference on Cyber Technology in Automation, Control, and Intelligent Systems*(Honolulu, HI), 1472–1477. doi: 10.1109/CYBER.2017.8446436

Chowdhury, A., Nishad, S. S., Meena, Y. K., Dutta, A., and Prasad, G. (2019). Hand-exoskeleton assisted progressive neurorehabilitation using impedance adaptation based challenge level adjustment method. *IEEE Trans. Haptics* 12, 128–140. doi: 10.1109/TOH.2018.2878232

Cote-Allard, U., Fall, C. L., Drouin, A., Campeau-Lecours, A., Gosselin, C., Glette, K., et al. (2019). Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 760–771. doi: 10.1109/TNSRE.2019.2896269

De Smedt, Q., Wannous, H., and Vandeborre, J.-P. (2019). Heterogeneous hand gesture recognition using 3D dynamic skeletal data. *Comput. Vis. Image Underst.* 181, 60–72. doi: 10.1016/j.cviu.2019.01.008

Duan, F., Ren, X., and Yang, Y. (2021). A Gesture recognition system based on time domain features and linear discriminant analysis. *IEEE Trans. Cogn. Dev. Syst.* 13, 200–208. doi: 10.1109/TCDS.2018.2884942

Dwivedi, A., Kwon, Y., McDaid, A. J., and Liarokapis, M. (2019). A learning scheme for EMG based decoding of dexterous, in-hand manipulation motions. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 2205–2215. doi: 10.1109/TNSRE.2019.2936622

Emerson, I., Potgieter, J., and Xu, W. L. (2016). "Control implementation for an integrated robotic and virtual mirror therapy system for stroke rehabilitation," in *2016 Ieee 14th International Workshop on Advanced Motion Control* (Auckland), 479–485. doi: 10.1109/AMC.2016.7496396

Frost, G., Dowling, J., Dyson, K., and BarOr, O. (1997). Cocontraction in three age groups of children during treadmill locomotion. *J. Electromyogr. Kinesiol.* 7, 179–186. doi: 10.1016/S1050-6411(97)84626-3

Furui, A., Eto, S., Nakagaki, K., Shimada, K., Nakamura, G., Masuda, A., et al. (2019). A myoelectric prosthetic hand with muscle synergy-based motion determination and impedance model-based biomimetic control. *Sci. Robot.* 4:eaaw6339. doi: 10.1126/scirobotics.aaw6339

Gasser, B. W., Martinez, A., Sasso-Lance, E., Kandilakis, C., and Durrough, C. M., Goldfarb, M. (2020). Preliminary assessment of a hand and arm exoskeleton for enabling bimanual tasks for individuals with hemiparesis. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2214–2223. doi: 10.1109/TNSRE.2020.3018649

Gerez, L., Gao, G., Dwivedi, A., and Liarokapis, M. (2020). A hybrid, wearable exoskeleton glove equipped with variable stiffness joints, abduction capabilities, and a telescopic thumb. *IEEE Access* 8, 173345–173358. doi: 10.1109/ACCESS.2020.3025273

He, S., Yang, C., Wang, M., Cheng, L., and Hu, Z. (2017). "Hand gesture recognition using MYO armband," in *2017 Chinese Automation Congress* (Jinan), 4850–4855. doi: 10.1109/CAC.2017.8243637

Hu, W., Wei, N., Li, Z. M., and Li, K. (2018). Effects of muscle fatigue on directional coordination of fingertip forces during precision grip. *PLoS ONE* 13:e0208740. doi: 10.1371/journal.pone.0208740

Hu, X. L., Tong, K. Y., Song, R., Zheng, X. J., Lui, K. H., Leung, W. W. F., et al. (2009). Quantitative evaluation of motor functional recovery process in chronic stroke patients during robot-assisted wrist training. *J. Electromyogr. Kinesiol.* 19: 639–650. doi: 10.1016/j.jelekin.2008.04.002

Hua, S., Wang, C., Xie, Z., and Wu, X. (2020). A force levels and gestures integrated multi-task strategy for neural decoding. *Complex Intell. Syst.* 6, 469–478. doi: 10.1007/s40747-020-00140-9

Iqbal, J., Khan, H., Tsagarakis, N. G., and Caldwell, D. G. (2014). A novel exoskeleton robotic system for hand rehabilitation - conceptualization to prototyping. *Biocybern. Biomed. Eng.* 34, 79–89. doi: 10.1016/j.bbe.2014.01.003

Kim, S., Lee, J., and Bae, J. (2017). Analysis of finger muscular forces using a wearable hand exoskeleton system. *J. Bionic Eng.* 14, 680–691. doi: 10.1016/S1672-6529(16)60434-1

Leonardis, D., Barsotti, M., Loconsole, C., Solazzi, M., Troncossi, M., Mazzotti, C., et al. (2015). An EMG-controlled robotic hand exoskeleton for bilateral rehabilitation. *IEEE Trans. Haptics* 8, 140–151. doi: 10.1109/TOH.2015.2417570

Li, K., Wei, N., Cheng, M., Hou, X., and Song, J. (2018). Dynamical coordination of hand intrinsic muscles for precision grip in diabetes mellitus. *Sci. Rep.* 8:4365. doi: 10.1038/s41598-018-22588-z

Li, K., Wei, N., and Yue, S. (2016). Effects of tactile sensitivity on structural variability of digit forces during stable precision grip. *Biomed Res. Int.* 2016:8314561. doi: 10.1155/2016/8314561

Li, M., He, B., Liang, Z., Zhao, C.-G., Chen, J., Zhuo, Y., et al. (2019). An attention-controlled hand exoskeleton for the rehabilitation of finger extension and flexion using a rigid-soft combined mechanism. *Front. Neurorobotics* 13:34doi: 10.3389/fnbot.2019.00034

Palm, R., and Iliev, B. (2007). "Learning of grasp behaviors for an artificial hand by time clustering and Takagi-Sugeno modeling," in *2006 IEEE International Conference on Fuzzy Systems* (Vancouver, BC), 291–298. doi: 10.1109/FUZZY.2006.1681728

Parajuli, N., Sreenivasan, N., Bifulco, P., Cesarelli, M., Savino, S., Niola, V., et al. (2019). Real-time EMG based pattern recognition control for hand prostheses: a review on existing methods, challenges and future implementation. *Sensors* 19:459.456 doi: 10.3390/s19204596

Pasluosta, C. F., Domalain, M. M., Fang, Y., Yue, G. H., and Li, Z.-M. (2013). Influence of nerve supply on hand electromyography coherence during a three-digit task. *J. Electromyogr. Kinesiol.* 23, 594–599. doi: 10.1016/j.jelekin.2013.01.006

Qu, Y., Shang, H., and Teng, S. (2020). "Reduce sEMG channels for hand gesture recognition," in *2020 IEEE 3rd International Conference on Information Communication and Signal Processing (ICICSP) (*Shanghai), 215–220. doi: 10.1109/ICICSP50920.2020.9232078

Refour, E. M., Sebastian, B., Chauhan, R. J., and Ben-Tzvi, P. (2019). A general purpose robotic hand exoskeleton with series elastic actuation. *J. Mechanisms Robotics* 11:060902. doi: 10.1115/1.4044543

Sarac, M., Solazzi, M., and Frisoli, A. (2019). Design requirements of generic hand exoskeletons and survey of hand exoskeletons for rehabilitation, assistive, or haptic use. *IEEE Trans. Haptics* 12, 400–413. doi: 10.1109/TOH.2019.2924881

Shengchang, L., Zonglong, H., Haoyu, T., Kai, Y., and Wenshuang, Y. (2017). "A hand gesture recognition system based on 24GHz radars," in *2017 International Symposium on Antennas and Propagation (ISAP)* (Phuket), 1–2.

Solnik, S., Rider, P., Steinweg, K., DeVita, P., and Hortobagyi, T. (2010). Teager-Kaiser energy operator signal conditioning improves EMG onset detection. *Eur. J. Appl. Physiol.* 110, 489–498. doi: 10.1007/s00421-010-1521-8

Takahashi, N., Furuya, S., and Koike, H. (2020). Soft exoskeleton glove with human anatomical architecture: production of dexterous finger movements and skillful piano performance. *IEEE Trans. Haptics* 13, 679–690. doi: 10.1109/TOH.2020.2993445

Wu, C., Yan, Y., Cao, Q., Fei, F., Yang, D., Lu, X., et al. (2020). sEMG measurement position and feature optimization strategy for gesture recognition based on ANOVA and neural networks. *IEEE Access* 8, 56290–56299. doi: 10.1109/ACCESS.2020.2982405

Xiloyannis, M., Gavriel, C., Thomik, A. A. C., and Faisal, A. A. (2017). Gaussian process autoregression for simultaneous proportional multi-modal prosthetic control with natural hand kinematics. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1785–1801. doi: 10.1109/TNSRE.2017.2699598

Yu, H., Fan, X., Zhao, L., and Guo, X. (2018). A novel hand gesture recognition method based on 2-channel sEMG. *Technol. Health Care* 26, S205–S214. doi: 10.3233/THC-174567

Zheng, X., Guo, Y., and Wang, H. (2011). "Pattern recognition of hand gesture based on LVQ neutral network," in *Artificial Intelligence and Computational Intelligence, Pt Iii*, eds H. Deng, D. Q. Miao, J. S. Lei, and F. L. Wang (Berlin: Springer), 575–580. doi: 10.1007/978-3-642-23896-3_71

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 25 October 2021
doi: 10.3389/fnbot.2021.693110

Improved Motion Classification With an Integrated Multimodal Exoskeleton Interface

Kevin Langlois[1,2]*, *Joost Geeroms*[1,3], *Gabriel Van De Velde*[1], *Carlos Rodriguez-Guerrero*[1,3], *Tom Verstraten*[1,3], *Bram Vanderborght*[1,2] and *Dirk Lefeber*[1,3]

[1] *Robotics & Multibody Mechanics Research Group, MECH Department, Vrije Universiteit Brussel, Brussel, Belgium,* [2] *IMEC, Leuven, Belgium,* [3] *Flanders Make, Lommel, Belgium*

Human motion intention detection is an essential part of the control of upper-body exoskeletons. While surface electromyography (sEMG)-based systems may be able to provide anticipatory control, they typically require exact placement of the electrodes on the muscle bodies which limits the practical use and donning of the technology. In this study, we propose a novel physical interface for exoskeletons with integrated sEMG- and pressure sensors. The sensors are 3D-printed with flexible, conductive materials and allow multi-modal information to be obtained during operation. A K-Nearest Neighbours classifier is implemented in an off-line manner to detect reaching movements and lifting tasks that represent daily activities of industrial workers. The performance of the classifier is validated through repeated experiments and compared to a unimodal EMG-based classifier. The results indicate that excellent prediction performance can be obtained, even with a minimal amount of sEMG electrodes and without specific placement of the electrode.

Keywords: human-machine interface, classification, exoskeletons, machine learning, intention recognition, electromyogram, wearable sensor

Edited by:
R. A. R. C. Gopura,
University of Moratuwa, Sri Lanka

Reviewed by:
Suncheol Kwon,
National Rehabilitation Center,
South Korea
Thilina Dulantha Lalitharatne,
Imperial College London,
United Kingdom

Correspondence:
Kevin Langlois
kevin.langlois@vub.be

Received: *09 April 2021*
Accepted: *23 September 2021*
Published: *25 October 2021*

Citation:
Langlois K, Geeroms J, Van De Velde G, Rodriguez-Guerrero C, Verstraten T, Vanderborght B and Lefeber D (2021) Improved Motion Classification With an Integrated Multimodal Exoskeleton Interface. Front. Neurorobot. 15:693110. doi: 10.3389/fnbot.2021.693110

1. INTRODUCTION

Upper body exoskeletons for industrial workers have been developed at an increasing pace over the past years and have shown promising results in a controlled lab environment, yet more nuanced for in-field experiments (De Looze et al., 2016; De Bock et al., 2020). Passive devices such as the Paexo (Ottobock, Duderstadt, Germany) (Maurice et al., 2019) or the Mate (Comau, Grugliasco, Italy) (Pacifico et al., 2020) provide assistance by storing energy in elastic elements. This energy is harvested through human motion and can be used to support a specific motion or posture such as overhead working. Due to the passive nature of the device, the assistance level cannot be dynamically controlled limiting the versatility of these devices. Active devices, on the other hand, comprise actuators (such as electric motors or other types, Gopura et al., 2016) which have the potential to deliver different assistive profiles for different tasks (Gull et al., 2020). Providing the right assistance is quite challenging, since the range of motions and tasks humans perform with the upper body is virtually infinite.

To solve this problem, researchers are developing intention recognition and task classification strategies to enable natural control. These recognition algorithms are most often based on unimodal sensing strategies, most often comprised of myoelectric signals (Kiguchi and Hayashi, 2012; Novak and Riener, 2015; Bi et al., 2019).

Classically these signals are acquired from wet Ag/AgCl electrodes. This method requires to first locate the ideal position for the sensor using body landmarks, then clean and shave the skin at this location and apply a gel to the muscle body from which the signal is collected. For high signal quality, this process requires expertise and a relatively long setup time. When used in combination with an exoskeleton, the optimal locations for the sensors are often obstructed by the interface. In addition, the comfort of the user could be compromised because many of the devices associated with this acquisition technique are bulky and several cables are attached to the arms of the users. These issues reduce the practical applicability of the method for an upper body exoskeleton for industrial workers. Recently, dry electrodes were developed that do not need gel, reduce setup time and allow for more portability. They however come at the cost of a lower signal to noise ratio (Hakonen et al., 2015), which will inevitably reduce classification accuracies. A promising method to overcome the shortcomings of EMG sensors is to combine information from different sensor modalities (Novak and Riener, 2015). In that context, a popular control method for exoskeletons is to combine the information from EMG- and mechanical sensors such as inertial measurement units (accelerometer and gyroscope) or force- and torque sensors.

A sensor that has yet to be combined with EMG are pressure sensors located at the physical interface of exoskeletons. The integration of pressure sensors in physical interfaces has been investigated in the robotics community (De Rossi et al., 2011; Tamez-Duque et al., 2015; Wilcox et al., 2016; Langlois et al., 2021) and have shown their relevance for the unimodal detection of the user motion intention (Lenzi et al., 2011). Additionally, pressure sensors can ensure a more safe and comfortable operation of such devices (He et al., 2017).

Pressure and EMG sensors were combined in wearable bands developed for the detection of hand and wrist motions (Connan et al., 2016; Jiang et al., 2020). The band comprises sEMG electrodes and force sensitive resistors that measure volume changes induced by muscular activity. Fusion of both modalities showed promising results for the performance of gesture recognition. In the domain of upper body exoskeletons, several muscle groups are usually targeted to achieve good recognition (Trigili et al., 2019). However, in the context of industrial workers, the application of electrodes on multiple muscles can be problematic since clothing is usually covering most of the muscles. Wearing t-shirts is still conceivable for certain applications such as logistics. In that regard, only the biceps and triceps muscles are potential sources of EMG information.

In this manuscript we propose a novel integrated, multimodal interface comprising EMG electrodes that measure activity of the biceps brachii, and pressure sensors that monitor the interaction between the user and the exoskeleton. The novelty of this research lies in the combination of EMG- and pressure-signals that this interface can obtain. Moreover, the EMG electrodes as well as the pressure sensors are all 3D printed. This allows to develop the interface for other body regions as well, or to customize the design to a specific user (Langlois et al., 2018).

Experiments on human subjects are carried out to explore the potential of the interface to classify lifting and reaching tasks. The analysis is performed in a test bench consisting of a torque controlled cobot. A classifier based on K-Nearest Neighbours (KNN) algorithm is trained to recognize lifting and reaching tasks in an off-line manner.

2. METHODS

2.1. Sensorized Interface

The physical interface is an upper-arm orthosis with integrated pressure sensors and EMG-electrodes. The objective of such an interface is to ensure the correct placement of the exoskeleton relative to the body, achieve effective force transmission, and most of all, support safe and comfortable interactions. Interface dynamics are known to play a crucial role in the ability of exoskeletons to provide assistance and comfort (Cherry et al., 2016; Langlois et al., 2020).

Four flexible polymer capacitive pressure sensors are integrated along the centre line of the orthosis, shown in **Figure 1**. The pressure sensors allow pressure measurements with a relative accuracy of approx. 10% at a rate of 10 Hz. At the beginning of every trial the sensors are calibrated relative to the force sensing of the cobot. The design of the pressure sensors and the calibration process are described in detail in Langlois et al. (2021).

In the developed interface, the muscular activity of a single muscle group, the biceps brachii, is monitored. The biceps brachii was chosen since the application for which this interface is designed is the assistance of upper body reaching and lifting during industrial workers' tasks. In that context, the muscles around the shoulder are more challenging to access. Monitoring the triceps muscles as these interact with the interface was out of the scope of this paper. Though co-located force and EMG sensors have been developed (Jiang et al., 2020).

The interface's flexible straps conceal an individual electrode pair, consisting of a printed polylactic acid, conductive filament, similar to Wolterink et al. (2020). The electrodes are pressed against the skin by the elastic straps to ensure skin contact. The same straps are also providing the attachment between the human and the robot. These are tightened the same way a non-sensorized strap would.

Specific electrode placement is not required/performed when donning as the straps ensure a similar pose across users. This approach would be beneficial in a commercial application for industrial workers, where workers do not need expertise or support for donning the exoskeleton. Although resulting in a wider variability of the EMG signal, we believe the pressure sensor data can potentially compensate for this effect. However, once the interface is worn, one should avoid the slippage of the electrodes since this will create noise.

2.2. Experimental Setup

The experimental setup comprises the sensorized interface and a torque-controlled cobot (Panda, Franka Emika, Munich, Germany). The cobot is programmed to simulate a passive upper body exoskeleton by means of a joint impedance controller. The impedance controller is set such that an assistive force is exerted onto the interface, effectively pushing the user's arm to an upward

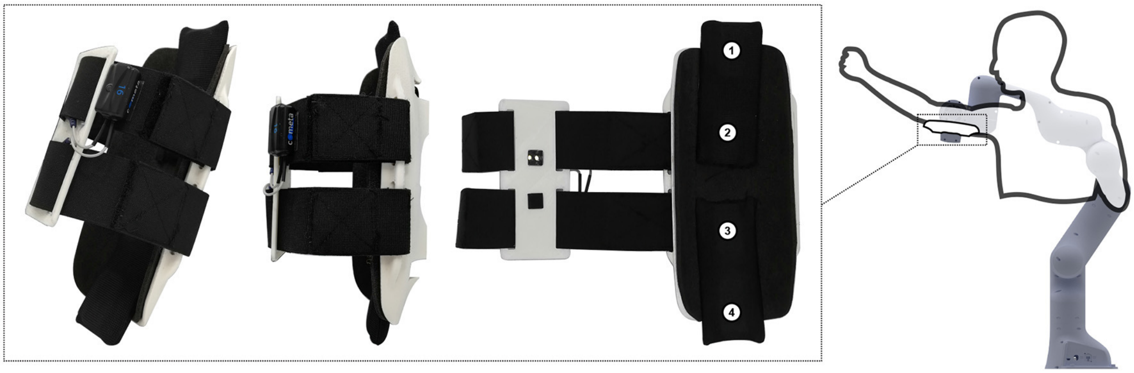


FIGURE 1 | The setup of the experiment consists of a sensorized interface and a torque controlled cobot. The four pressure sensors located on the inside of the interface are shown (number 1–4). The EMG electrodes are the two black squares on the inside of the straps. The cobot simulates an upper body passive exoskeleton using a joint impedance controller.

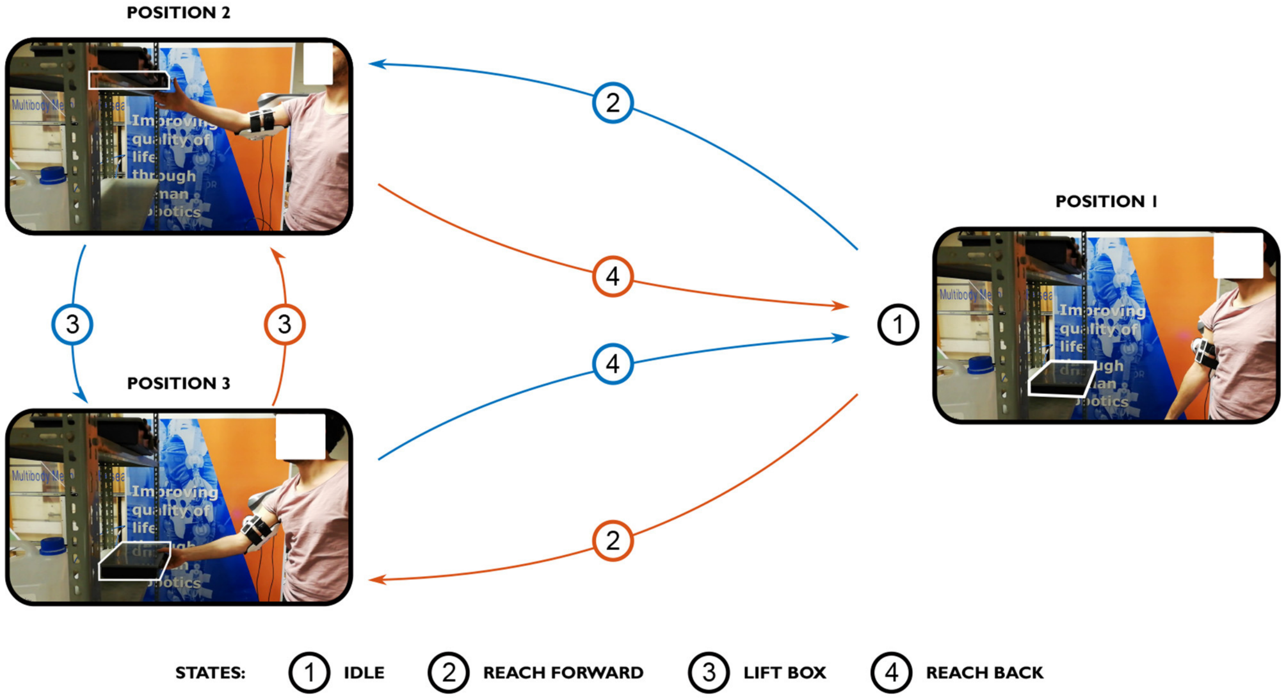


FIGURE 2 | The subject starts in a relaxed position with the arm along the trunk (IDLE). Then, the subject reaches forward (RF) to grab the box, moves the box toward the other shelf (LB), releases the box and reaches backwards (RB). Note that there are two possible initial positions of the box, and thus two paths that can be initially chosen from, as indicated in blue and red.

position. This means the subject exerts a force to pull the arms down. Similarly to an actual passive exoskeleton, the goal is to assist humans by reducing the efforts delivered at the shoulder level when executing lifting tasks.

The experiment consists of two shelves and a box with a mass of 2.2 kg. The user straps him/herself into the interface and performs the task depicted in **Figure 2**. The task consists of four parts: first, the subject starts the exercise by relaxing the arm along the trunk (IDLE). Then, the subject reaches forward to grab the box on the shelf (RF), grabs and lifts the box (LB) and places it onto the other shelf. The initial position of the box is randomized (top or bottom shelf). After placing the box, the participant reaches back (RB) to idle position and repeats the task. IDLE, RF, LB, and RB are the labels of the exercise. This exercise is repeated during 1 min, at self-selected speed. Each participant repeats the exercise seven times.

During this exercise the muscular activity of the biceps, the pressure acting on the inside of the interface and the position

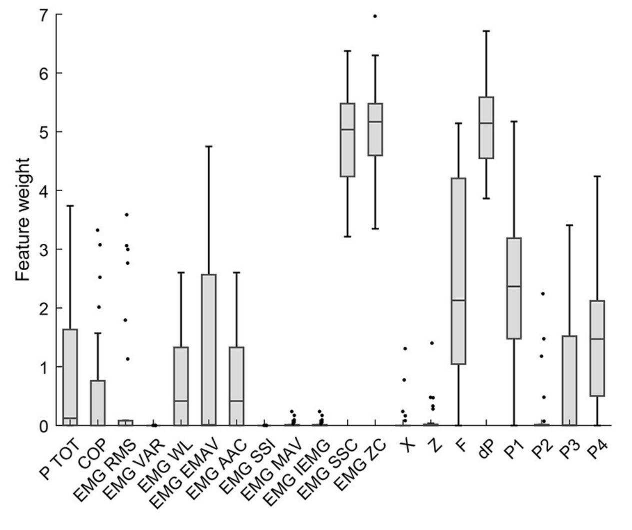


FIGURE 3 | Neighborhood component analysis allows to select features to achieve maximal prediction accuracy. A lower score means a more redundant signal that does not contribute to a higher classification accuracy.

and forces at the end-effector are recorded. The details of the processing methods are explained further below in section 2.3.

2.3. Acquisition System

EMG signals from the biceps brachii are sampled through a Cometa Mini Wave Infinity system (Cometa Srl, Bareggio, Italy) at a frequency of 2 kHz. The signals are band pass-filtered (15–400 Hz) before segmentation. A sliding window segmentation is implemented with a window of 600 samples (300 ms) and an overlap of 150 samples.

The pressure data are captured at a 10 Hz rate and the signals are filtered through a second-order Butterworth filter with a cutoff frequency of 2 Hz. The initial pre-compression/loading of the interface is measured and accounted for: At the start of the experiment, the user keeps his/her arm at the equilibrium position of the cobot, i.e., where forces are small, and this for 20 s. The mean pressure over that time window is subtracted from all subsequent measurements.

The force and position measurements at the end-effector of the cobot are captured at a 1 kHz rate and are filtered through a first-order Butterworth filter at a 2 Hz cutoff frequency. The cobot can measure external forces with a resolution of 0.1 N.

Data analysis is performed in the Matlab environment (MathWorks, Massachusetts, United States).

2.4. Subjects

A total of 4 healthy subjects participated in the experiment, and they all provided written informed consent. The procedures were approved by the Institutional Review Board at The UZ Brussel, Vrije Universiteit Brussel and complied with the principles of the declaration of Helsinki.

2.5. Classification Features

Regarding EMG signals, the 10 following features were selected for further investigation: Root Mean Square (RMS), Wavelength (WL), Enhanced Mean Absolute Value (EMAV), Average Amplitude Change (AAC), Variance (VAR), Simple Square Integral (SSI), Mean Absolute Value (MAV), Integrated EMG (IEMG), Slope Sign Change (SSC), and Zero Crossing (ZC). In terms of pressure, seven features are proposed for further analysis: Pressure values at sensor element 1-4 (P1-P4), Total Pressure (P TOT), Differential Pressure (dP), and Center Of Pressure (COP). The external force (F) acting normal to the interface and the position (X,Z) of the end-effector are also evaluated for further analysis. This constitutes a total of 20 features.

Neighbourhood component analysis (NCA) is performed to reduce the features that are passed on to the classifier. NCA is a non-parametric method for selecting features with the goal of maximizing prediction accuracy of classification algorithms (Yang et al., 2012). The output of the algorithm is a feature weight vector that maximizes the classification accuracy. This algorithm is implemented in Matlab using the *fscnca* function. The results of the analysis are shown in **Figure 3**. Features with a low score are not considered further. These are the RMS, VAR, SSI, MAV, and IEMG features of the EMG signals, and the position signals of the end-effector.

2.6. Classification Protocol

Each subject except one carried out seven trials (due to technicalities, one was discarded for subject four). Each trial is composed of several cycles (ranging from five to eight), as shown in **Table 1**. All the data was manually labeled based on the speed of the end-effector and the video footage of the experiment. To train the classifier, every trial of each subject is divided into three subsets: a training-, validation- and test set. First, a training set and a test set are divided by leave-one-cycle-out partitioning. This means the classifier is trained (and validated) on all but one cycle. The classifier is then tested on the left-out cycle. This process is repeated for each cycle of each trial, and the average accuracy for each subject is reported in the section 3. The validation set is partitioned based on a hold-out fraction of 25% of the training set.

A K-Nearest Neighbour cosine classifier with 1,001 neighbors is chosen for this task. This parameter gave over-all good results and did not overfit the data. Fewer neighbors will result in higher accuracy for a single dataset but entail a less flexible classifier.

2.7. Classification Performances

Four performance metrics are shown in the section 3.

Accuracy is the fraction of predictions that are correct:

$$Accuracy = \frac{TP + TN}{n} \tag{1}$$

with TP the number of True Positives, TN the number of True Negatives and n the number of predictions.

TABLE 1 | Average data across all trials.

Subject	1	2	3	4
IDLE duration [min max] [s]	2.48 [1.60–3.59]	1.25 [0.63–2.14]	0.97 [0.44–1.88]	0.88 [0.29–1.51]
RF duration [min max] [s]	1.70 [1.07–2.28]	2.36 [1.65–3.79]	1.71 [1.17–2.33]	1.77 [0.92–2.53]
LU duration [min max] [s]	3.10 [2.09–3.89]	4.34 [3.55–5.05]	2.76 [2.28–3.40]	3.30 [2.33–4.37]
RB duration [min max] [s]	2.11 [1.51–3.16]	2.48 [1.79–3.45]	1.96 [2.28–1.60]	2.13 [1.46–2.92]
Cycles [min max]	5.14 [5–6]	4.57 [4–5]	7 [7–7]	6.33 [5–8]
Peak force [N]	8.9	26.75	11.95	17.39
Peak pressure [kPa]	8.47	16.2	6.57	8.12
Peak velocity [Vx,Vz] [m/s]	[0.055;0.117]	[0.194;0.119]	[0.215;0.147]	[0.219;0.133]
ROM [range X; range Z] [m]	[0.111;0.279]	[0.232;0.305]	[0.223;0.291]	[0.199;0.279]

Large variations across subjects can be observed in terms of how the exercise was performed. Subject 2 performed slower motions with wider range of motion and higher forces, whereas subject 1 performed smaller range of motions with lower forces.

Sensitivity or true positive rate measures the proportion of positives that are correctly identified:

$$Sensitivity = \frac{TP}{P} \tag{2}$$

with P the number of real positive cases in the data.

Specificity or true negative rate measures the proportion of negatives that are correctly identified:

$$Specificity = \frac{TN}{N} \tag{3}$$

with N the number of real negative cases in the data.

3. RESULTS

3.1. Motion Data, Pressure, and Force

Since the four participants could perform the task at a self-selected speed, we found widely varying executions, both in terms of speed of execution and forces acting on the body as well as muscular activity. First, in terms of external force acting normal to the interface, we observed varying peak forces across the subjects. For subject one, the lowest peak force is found, with a value of 8.9 N. The highest peak force is found for subject 2, with a value of 26.75 N. The forces acting on the interface depend on the equilibrium position of the robot, which had to be slightly adjusted for each participant, as well as the range of motion (ROM) of the participant. Since the robot acts as an impedance, the further away from equilibrium a subject moves the arm, the higher the force. The smallest range of motion is found for subject 1, with a total range across all trials of 0.111 m along the X-direction (moving the hand forward, parallel to ground) and 0.279 m in the Z-direction (parallel to gravity). The highest range of motion is found for subject 2, with a total range of 0.232 m and 0.305 m, in the X- and Z-direction respectively. In terms of pressure, the peak occurs during the lifting motion of the box. This peak pressure is caused by the assistive force of the robot and the change in volume of the arm. The lowest peak pressure is found for subject 3, with a value of 6.57 kPa. The highest peak pressure is found for subject 2, with a value of 16.2 kPa. At the same time, subject 3 spent the least amount of time lifting up the box (LU), on average. While subject 2 spent, on average, the most amount of time lifting up the box. Subject 2 also performed the least amount of cycles per trial, with an average number of cycles of 4.57.

All the average values across all trials are shown in **Table 1**. In **Figure 4** the raw data outcome of a single trial of subject 1 is shown.

3.2. Neighbourhood Component Analysis

The NCA revealed the features that are redundant in the classification of the tasks. The features that were not further used for classification are the RMS, VAR, SSI, MAV, and IEMG features of the EMG signals, and the position signals of the end-effector. The P2 feature was left in the feature pool. The two outer sensors, P1 and P4 scored a higher feature weight (median weight of 2.4 and 1.5, respectively) than the inner pressure sensors, P2 and P3 (median weight of 0.0 for both). The Zero Crossing feature and the Slope Sign Change (SSC) are the two highest rated features of the EMG signals, with a respective median weight of 5.2 and 5.0. The differential Pressure (dP) was the second highest feature of all, with a median score of 5.1.

Based on these results, seven classifiers are further analyzed. First, three unimodal classifiers (i.e., single type of sensor) are constructed: the EMG-classifier (comprising WL, EMAV, AAC, SSC and ZC features), the P-classifier (comprising P TOT, COP, dP, P1, P2, P3, P4 features) and the F-classifier (comprising F feature). Then, multimodal classifiers based on all combinations of sensors are constructed: EMG+P, EMG+F, P+F and EMG+P+F.

3.3. Classifier Performances

3.3.1. EMG, Pressure, and Force

Across all participants and all trials the highest median accuracy is achieved with all the features, as shown in **Figure 5**, with a value of 85.6 %, followed by the EMG+P classifier, with an accuracy of 84.4 %. The P classifier scored better than the EMG classifier with a value of 73.3 % and 69.8 % respectively. The classifier based on external force only scored the worst result with a value of 29.3

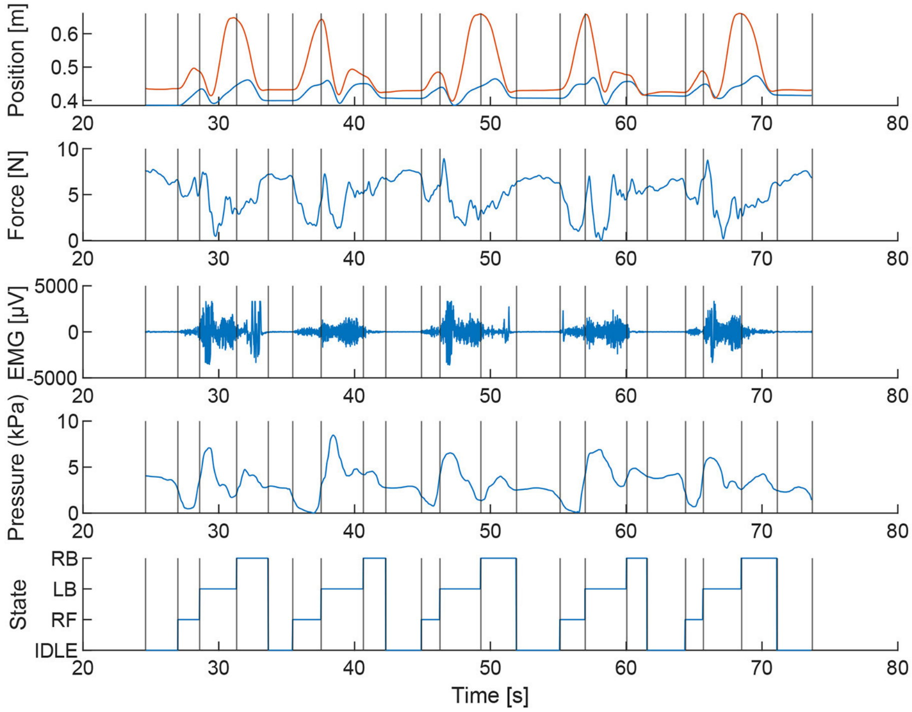


FIGURE 4 | During the task the motion and force of the end-effector, the pressure inside the interface and the muscular activity of the biceps brachii are sampled. The data shown are the results of one of the trials of subject 1.

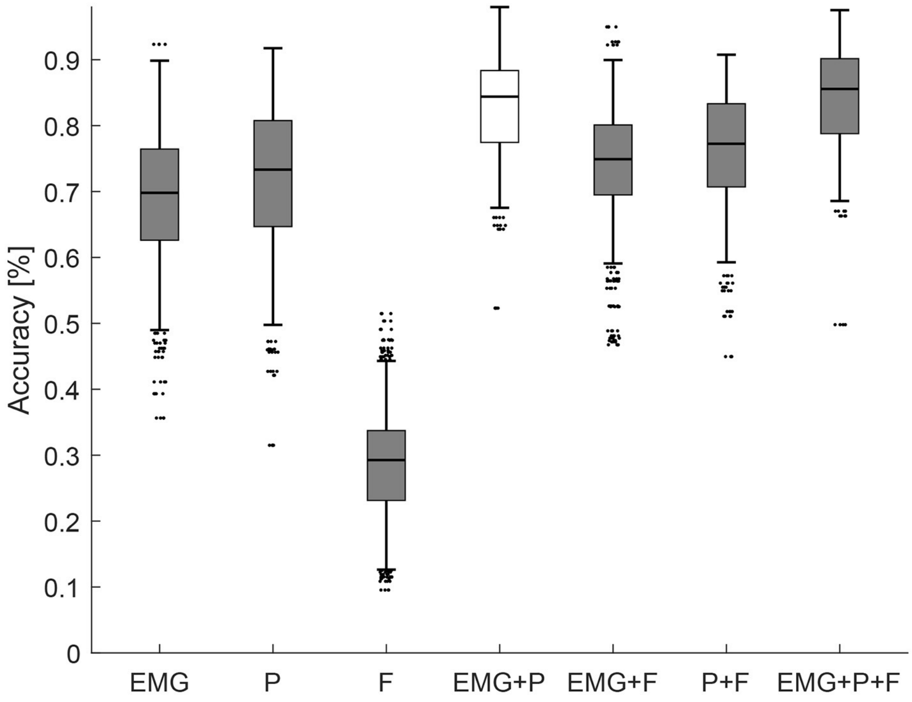


FIGURE 5 | Accuracy of all the classifiers across all subjects and trials trained on different sensory input.

%. The best classifier using only two modalities is the EMG+P classifier which is further analyzed below.

3.3.2. EMG and Pressure

A Tukey-Kramer *post-hoc* test showed a statistically significant better result with the multimodal classifier relative to the unimodal classifiers across all participants and trials ($p \leq 0.0001$). For subject 1 a statistically significant better result was achieved with the EMG classifier, relative to the P classifier ($p \leq 0.0001$). For subject 2 ($p \leq 0.0001$) and subject 3 ($p \leq 0.001$) the opposite is true. For subject 4 the P classifier was not statistically better than the EMG classifier ($p \leq 0.052$). These results are reported in **Figure 6**. In **Figure 7**, the confusion matrices of the EMG+P classifiers are shown for the four subjects. The confusion matrices shown are the results from trial 6, 13, 20, and 23 which represent closely the median accuracies reported in **Figure 6**.

4. DISCUSSION

The main outcome of the analysis is that pressure is a relevant sensory input that can be combined with EMG sensors to recognize lifting and reaching tasks. To the author's knowledge it is the first time this combination of sensors is proposed for the recognition of upper body tasks while wearing an upper body assistive device. The peak classification accuracy found across all subjects and trials is 96.8 %. An important limitation of this outcome is that the classifiers trained in this study are subject specific. For a practical implementation of such devices a non-subjective specific classifier is preferred. Trigili et al. (2019) demonstrated the implementation of such a recognition algorithm through EMG signals alone. As much as seven muscle groups of the upper body were targeted to achieve good classification performances. While it would not be practical to integrate that many electrodes in a commercial device, perhaps a multimodal strategy such as the one described in this manuscript can help reduce the number of necessary electrodes.

Another limitation, are the performances reported in this study, which are achieved through an off-line classification process. It is known from literature that on-line classifiers do not perform as well (Novak and Riener, 2015). For on-line classifiers the processing time, as well as the feature extraction windows become more critical, since dynamic tasks require fast decision making. In that aspect a trade-off between classification accuracy and timing is inevitable. Different lengths of classification windows are known to affect classification performance (Smith et al., 2010). Potentially, in the case of on-line classification a KNN-based classifier might become a bottleneck, since the classifier relies on the calculation of angles between all neighbors for each new feature sample. Consequently, the storage requirements and the computational time proportionally increase with the size of the training set (De Leonardis et al., 2018). A variety of other classifiers such as linear discriminant analysis (LDA), support vector machines (SVM), decision trees (DT), or artificial neural networks (ANN) have been used in the literature (Novak and Riener, 2015; Bi et al., 2019) and are not showing the same disadvantages (De Leonardis et al., 2018).

Since the sample size was small, only subject specific classifiers were developed. For a general classifier, a large sample size will be required, and a different fusion algorithm might be necessary to cope with the variability. More specifically, it is known that the variability of EMG signals can be significant across time due to artefacts and crosstalk (Bi et al., 2019). On the other hand, we found that pressure measurements are generally more stable signals across a particular pattern, which is in line with previous research on the topic (Connan et al., 2016; Langlois et al., 2021). This leads us to expect the pressure sensing would improve a more general motion classification across a large sample size as well. Interestingly, in our experiments the classifier based solely on external force did not perform as well as expected. Most probably, adding the shear component of force (instead of only the normal component) would improve the results. Also adding a derivative or integral component of force to the features might improve the performance, albeit with the necessary filtering.

The same observation holds true for the position information of the end-effector. The neighbourhood component analysis determined the position data to be redundant with regard to the other modalities. Even though, measuring limb position was shown to increase classification accuracy (Fougner et al., 2011), since it resolves the position effect (Radmand et al., 2014). Potentially, the pressure sensors carry information about the position of the limb since the robot is programmed to exert an assistive force by means of a joint impedance controller. Additionally, adding a derivative and second derivative term of the position might result in a different outcome.

Several potential further developments could lead to improved performance of the presented design. Firstly, a similar design to the one presented in Jiang et al. (2020), wherein a co-located force sensor and EMG sensor is implemented would allow the triceps muscles to be monitored as well. While contact might not always be ensured in an exoskeleton, the system can be trained to recognize electrode displacements and mitigate losses in classification accuracy (Hargrove et al., 2008). In that regard, assessing how pressure readings can further improve detection of electrode shift can be interesting. Secondly, to limit complexity, a single electrode pair was printed in the interface. In the future an array of electrodes could be integrated which could compensate for the problem of lower signal to noise ratio. Thirdly, the classification in this experiment was achieved using a KNN algorithm. This type of algorithm was chosen since it is considered a simple and efficient method that yields competitive results compared to state-of-the-art classification methods (Yang et al., 2012; De Leonardis et al., 2018). Previous research has shown that the effects of algorithm type on accuracy is generally small for single time invariant classifiers, and the choice of specific features seems to be more important (Novak and Riener, 2015). However, other types of classification methods such as adaptive or parallel classifiers should be considered in the future, since superior classification accuracy for myoelectric control was shown, albeit at the expense of added complexity (Novak and Riener, 2015).

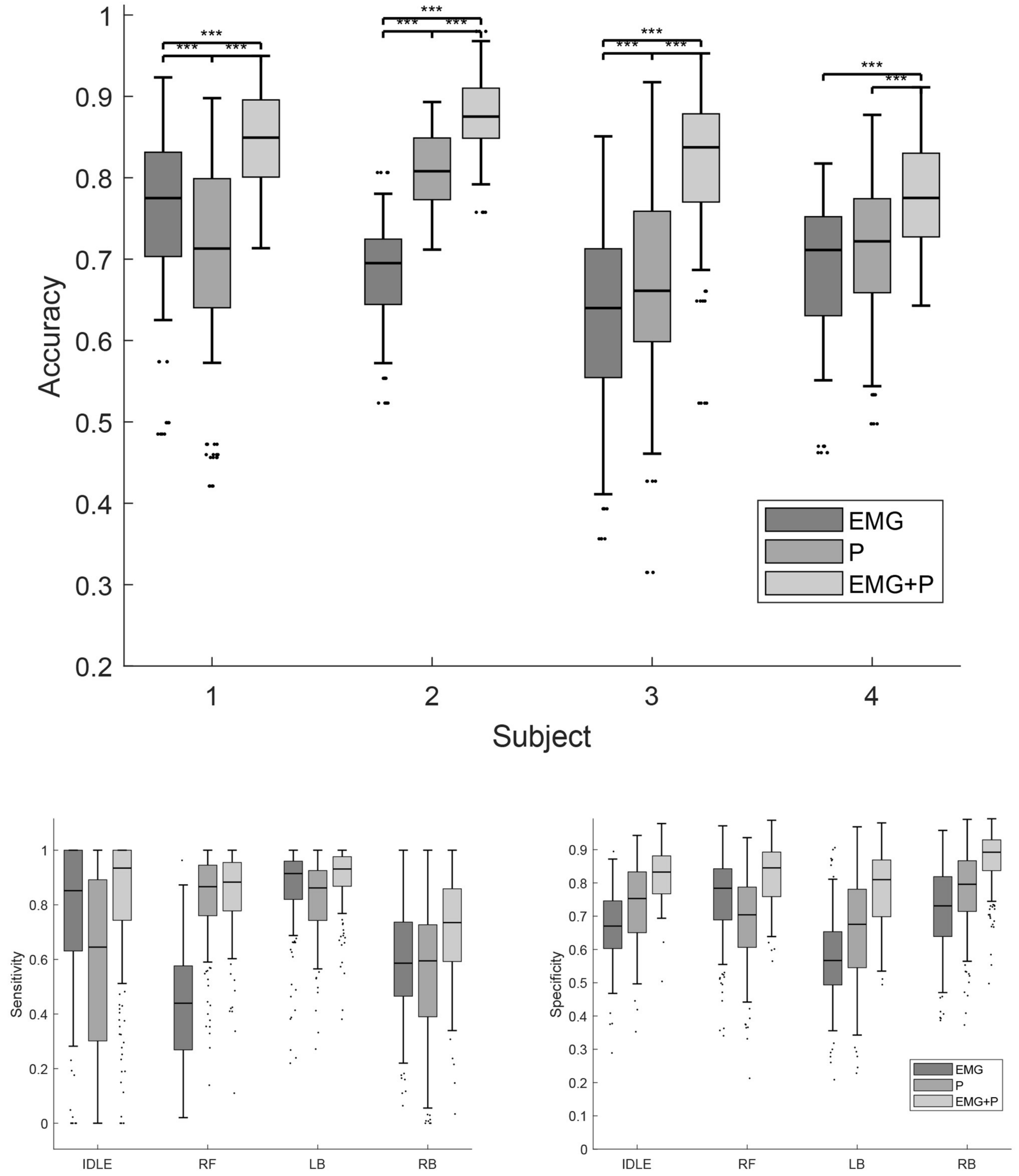
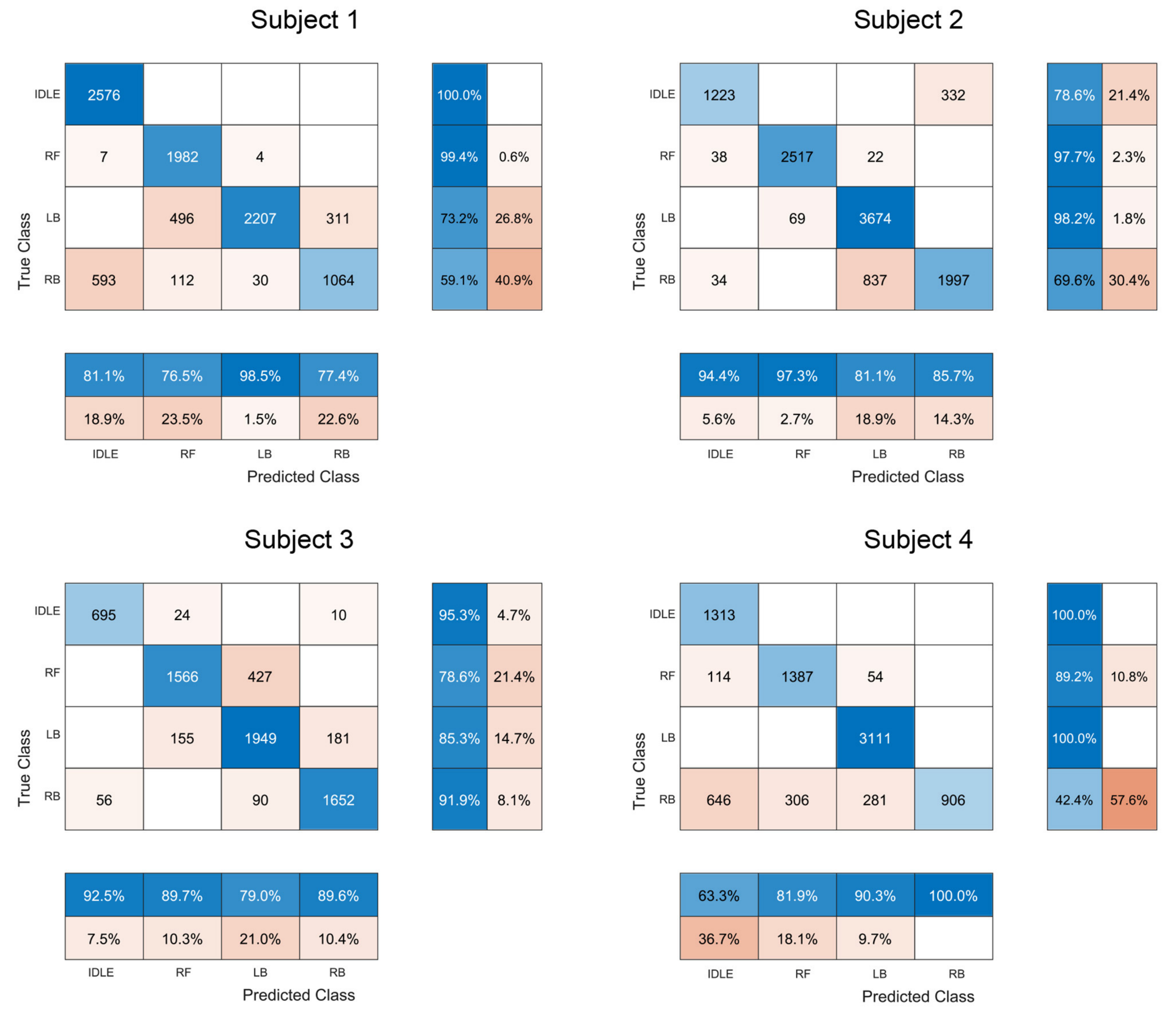


FIGURE 6 | The accuracy, sensitivity, and specificity of the EMG, P, and EMG+P classifiers. The dots indicate outliers, the stars indicate levels of significance.



Subject 1

	IDLE	RF	LB	RB		
IDLE	2576				100.0%	
RF	7	1982	4		99.4%	0.6%
LB		496	2207	311	73.2%	26.8%
RB	593	112	30	1064	59.1%	40.9%
	81.1%	76.5%	98.5%	77.4%		
	18.9%	23.5%	1.5%	22.6%		

Subject 2

	IDLE	RF	LB	RB		
IDLE	1223			332	78.6%	21.4%
RF	38	2517	22		97.7%	2.3%
LB		69	3674		98.2%	1.8%
RB	34		837	1997	69.6%	30.4%
	94.4%	97.3%	81.1%	85.7%		
	5.6%	2.7%	18.9%	14.3%		

Subject 3

	IDLE	RF	LB	RB		
IDLE	695	24		10	95.3%	4.7%
RF		1566	427		78.6%	21.4%
LB		155	1949	181	85.3%	14.7%
RB	56		90	1652	91.9%	8.1%
	92.5%	89.7%	79.0%	89.6%		
	7.5%	10.3%	21.0%	10.4%		

Subject 4

	IDLE	RF	LB	RB		
IDLE	1313				100.0%	
RF	114	1387	54		89.2%	10.8%
LB			3111		100.0%	
RB	646	306	281	906	42.4%	57.6%
	63.3%	81.9%	90.3%	100.0%		
	36.7%	18.1%	9.7%			

FIGURE 7 | Confusion matrices for the EMG+P classifier for each subject. The row summaries display the percentages of correctly and incorrectly classified observations for each true class. The column summaries display the percentages of correctly and incorrectly classified observations for each predicted class. Diagonal and off-diagonal cells correspond to correctly and incorrectly classified observations, respectively.

5. CONCLUSION

In this paper an integrated multimodal interface for upper-body exoskeletons is presented. The analysis shows the relevance of integrating pressure sensors and EMG sensors into interfaces with the aim of improving classification of upper body lifting and reaching tasks. The performed neighbourhood component analysis revealed that the WL, EMAV, AAC, SSC and ZC features of the EMG signal, the dP, P TOT, COP, P1-P4 features of the pressure signals and the external force provided the most information toward optimal classification. In the future, researchers in the field should look into the possibility of developing smarter interfaces, integrating different sensors to achieve better recognition algorithms.

DATA AVAILABILITY STATEMENT

The data presented in this study are available on request from the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by UZ Brussel, Vrije Universiteit Brussel. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KL, JG, CR-G, BV, and DL developed the concept. KL developed the technology with inputs from TV, BV, and DL. KL, JG, and GV conceived and designed the experiments. TV, JG, GV, CR-G, and BV contributed to the design and layout of the article, tables, additions to the bibliography, and extensive revisions. CR-G, BV, and DL helped to obtain the funding for the project that financed this research. All authors contributed to the article and approved the submitted version.

FUNDING

The work presented in this paper was supported by the Research Foundation - Flanders (FWO) under grant no. S000118N SBO Exo4Work project. TV was a postdoctoral fellow of the Research Foundation Flanders (FWO).

ACKNOWLEDGMENTS

The authors would like to thank the participants of the study.

REFERENCES

Bi, L., Feleke, G., and Guan, C. (2019). A review on EMG-based motor intention prediction of continuous human upper limb motion for human-robot collaboration. *Biomed. Signal Process. Control* 51, 113–127. doi: 10.1016/j.bspc.2019.02.011

Cherry, M. S., Kota, S., Young, A., and Ferris, D. P. (2016). Running with an elastic lower limb exoskeleton. *J. Appl. Biomech.* 32, 269–277. doi: 10.1123/jab.2015-0155

Connan, M., Ruiz Ramirez, E., Vodermayer, B., and Castellini, C. (2016). Assessment of a wearable force-and electromyography device and comparison of the related signals for myocontrol. *Front. Neurorobot.* 10:17. doi: 10.3389/fnbot.2016.00017

De Bock, S., Ghillebert, J., Govaerts, R., Elprama, S. A., Marusic, U., Serrien, B., et al. (2020). Passive shoulder exoskeletons: more effective in the lab than in the field? *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 173–183. doi: 10.1109/TNSRE.2020.3041906

De Leonardis, G., Rosati, S., Balestra, G., Agostini, V., Panero, E., Gastaldi, L., et al. (2018). "Human activity recognition by wearable sensors: comparison of different classifiers for real-time applications," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)* (Rome: IEEE), 1–6. doi: 10.1109/MeMeA.2018.8438750

De Looze, M. P., Bosch, T., Krause, F., Stadler, K. S., and O'Sullivan, L. W. (2016). Exoskeletons for industrial application and their potential effects on physical work load. *Ergonomics* 59, 671–681. doi: 10.1080/00140139.2015.1081988

De Rossi, S. M. M., Vitiello, N., Lenzi, T., Ronsse, R., Koopman, B., Persichetti, A., et al. (2011). Sensing pressure distribution on a lower-limb exoskeleton physical human-machine interface. *Sensors* 11, 207–227. doi: 10.3390/s110100207

Fougner, A., Scheme, E., Chan, A. D., Englehart, K., and Stavdahl, Ø. (2011). Resolving the limb position effect in myoelectric pattern recognition. *IEEE Trans. Neural Syst. Rehabil. Eng.* 19, 644–651. doi: 10.1109/TNSRE.2011.2163529

Gopura, R., Bandara, D., Kiguchi, K., and Mann, G. K. (2016). Developments in hardware systems of active upper-limb exoskeleton robots: a review. *Robot. Auton. Syst.* 75, 203–220. doi: 10.1016/j.robot.2015.10.001

Gull, M. A., Bai, S., and Bak, T. (2020). A review on design of upper limb exoskeletons. *Robotics* 9:16. doi: 10.3390/robotics9010016

Hakonen, M., Piitulainen, H., and Visala, A. (2015). Current state of digital signal processing in myoelectric interfaces and related applications. *Biomed. Signal Process. Control* 18, 334–359. doi: 10.1016/j.bspc.2015.02.009

Hargrove, L., Englehart, K., and Hudgins, B. (2008). A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control. *Biomed. Signal Process. Control* 3, 175–180. doi: 10.1016/j.bspc.2007.11.005

He, Y., Eguren, D., Luu, T. P., and Contreras-Vidal, J. L. (2017). Risk management and regulations for lower limb medical exoskeletons: a review. *Med. Dev.* 10:89. doi: 10.2147/MDER.S107134

Jiang, S., Gao, Q., Liu, H., and Shull, P. B. (2020). A novel, co-located EMG-fmg-sensing wearable armband for hand gesture recognition. *Sens. Actuat. A Phys.* 301:111738. doi: 10.1016/j.sna.2019.111738

Kiguchi, K., and Hayashi, Y. (2012). An EMG-based control for an upper-limb power-assist exoskeleton robot. *IEEE Trans. Syst. Man Cybernet. B* 42, 1064–1071. doi: 10.1109/TSMCB.2012.2185843

Langlois, K., Moltedo, M., Bacek, T., Rodriguez-Guerrero, C., Vanderborght, B., and Lefeber, D. (2018). "Design and development of customized physical interfaces to reduce relative motion between the user and a powered ankle foot exoskeleton," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)* (Enschede: IEEE), 1083–1088. doi: 10.1109/BIOROB.2018.8487706

Langlois, K., Rodriguez-Cianca, D., Serrien, B., De Winter, J., Verstraten, T., Rodriguez-Guerrero, C., et al. (2020). Investigating the effects of strapping pressure on human-robot interface dynamics using a soft robotic cuff. *IEEE Trans. Med. Robot. Bionics*. 3, 146–155. doi: 10.1109/TMRB.2020.3042255

Langlois, K., Roels, E., Van De Velde, G., Espadinha, C., Van Vlerken, C., Verstraten, T., et al. (2021). Integration of 3D printed flexible pressure sensors into physical interfaces for wearable robots. *Sensors* 21:6. doi: 10.3390/s21062157

Lenzi, T., Vitiello, N., De Rossi, S. M. M., Persichetti, A., Giovacchini, F., Roccella, S., et al. (2011). Measuring human-robot interaction on wearable robots: a distributed approach. *Mechatronics* 21, 1123–1131. doi: 10.1016/j.mechatronics.2011.04.003

Maurice, P., Čamernik, J., Gorjan, D., Schirrmeister, B., Bornmann, J., Tagliapietra, L., et al. (2019). Objective and subjective effects of a passive exoskeleton on overhead work. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 152–164. doi: 10.1109/TNSRE.2019.2945368

Novak, D., and Riener, R. (2015). A survey of sensor fusion methods in wearable robotics. *Robot. Auton. Syst.* 73, 155–170. doi: 10.1016/j.robot.2014.08.012

Pacifico, I., Scano, A., Guanziroli, E., Moise, M., Morelli, L., Chiavenna, A., et al. (2020). An experimental evaluation of the proto-mate: a novel ergonomic upper-limb exoskeleton to reduce workers' physical strain. *IEEE Robot. Automat. Mag.* 27, 54–65. doi: 10.1109/MRA.2019.2954105

Radmand, A., Scheme, E., and Englehart, K. (2014). "A characterization of the effect of limb position on EMG features to guide the development of effective prosthetic control schemes," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Chicago, IL: IEEE), 662–667. doi: 10.1109/EMBC.2014.6943678

Smith, L. H., Hargrove, L. J., Lock, B. A., and Kuiken, T. A. (2010). Determining the optimal window length for pattern recognition-based myoelectric control: balancing the competing effects of classification error and controller delay. *IEEE Trans. Neural Syst. Rehabil. Eng.* 19, 186–192. doi: 10.1109/TNSRE.2010.2100828

Tamez-Duque, J., Cobian-Ugalde, R., Kilicarslan, A., Venkatakrishnan, A., Soto, R., and Contreras-Vidal, J. L. (2015). Real-time strap pressure sensor system for powered exoskeletons. *Sensors* 15, 4550–4563. doi: 10.3390/s150204550

Trigili, E., Grazi, L., Crea, S., Accogli, A., Carpaneto, J., Micera, S., et al. (2019). Detection of movement onset using EMG signals for upper-limb exoskeletons in reaching tasks. *J. Neuroeng. Rehabil.* 16, 1–16. doi: 10.1186/s12984-019-0512-1

Wilcox, M., Rathore, A., Morgado Ramirez, D. Z., Loureiro, R. C., and Carlson, T. (2016). Muscular activity and physical interaction forces during lower limb exoskeleton use. *Healthcare Technol. Lett.* 3, 273–279. doi: 10.1049/htl.2016.0063

Wolterink, G., Dias, P., Sanders, R. G., Muijzer, F., van Beijnum, B.-J., Veltink, P., et al. (2020). Development of soft sEMG sensing structures using 3D-printing technologies. *Sensors* 20:4292. doi: 10.3390/s20154292

Yang, W., Wang, K., and Zuo, W. (2012). Neighborhood component feature selection for high-dimensional data. *J. Comput*. 7, 161–168. doi: 10.4304/jcp.7.1.161-168

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Neurorobotics

ORIGINAL RESEARCH
published: 29 October 2021
doi: 10.3389/fnbot.2021.692539

A Multi-Information Fusion Method for Gait Phase Classification in Lower Limb Rehabilitation Exoskeleton

Yuepeng Zhang[1], *Guangzhong Cao*[1*], *Ziqin Ling*[1], *WenZhou Li*[1], *Haoran Cheng*[1], *Binbin He*[1], *Shengbin Cao*[1] *and Aibin Zhu*[2]

[1] *Guangdong Key Laboratory of Electromagnetic Control and Intelligent Robots, Shenzhen University, Shenzhen, China,*
[2] *Institute of Robotics and Intelligent Systems, Xi'an Jiaotong University, Xi'an, China*

Gait phase classification is important for rehabilitation training in patients with lower extremity motor dysfunction. Classification accuracy of the gait phase also directly affects the effect and rehabilitation training cycle. In this article, a multiple information (multi-information) fusion method for gait phase classification in lower limb rehabilitation exoskeleton is proposed to improve the classification accuracy. The advantage of this method is that a multi-information acquisition system is constructed, and a variety of information directly related to gait movement is synchronously collected. Multi-information includes the surface electromyography (sEMG) signals of the human lower limb during the gait movement, the angle information of the knee joints, and the plantar pressure information. The acquired multi-information is processed and input into a modified convolutional neural network (CNN) model to classify the gait phase. The experiment of gait phase classification with multi-information is carried out under different speed conditions, and the experiment is analyzed to obtain higher accuracy. At the same time, the gait phase classification results of multi-information and single information are compared. The experimental results verify the effectiveness of the multi-information fusion method. In addition, the delay time of each sensor and model classification time is measured, which shows that the system has tremendous real-time performance.

Keywords: sEMG, multi-information fusion, gait phase classification, lower limb rehabilitation exoskeleton, convolutional neural network (CNN), real-time

Edited by:
Thilina Dulantha Lalitharatne,
Imperial College London,
United Kingdom

Reviewed by:
Farong Gao,
Hangzhou Dianzi University, China
R. A. R. C. Gopura,
University of Moratuwa, Sri Lanka

****Correspondence:***
Guangzhong Cao
gzcao@szu.edu.cn

Received: *08 April 2021*
Accepted: *28 September 2021*
Published: *29 October 2021*

Citation:
Zhang Y, Cao G, Ling Z, Li W, Cheng H, He B, Cao S and Zhu A (2021) A Multi-Information Fusion Method for Gait Phase Classification in Lower Limb Rehabilitation Exoskeleton.
Front. Neurorobot. 15:692539.
doi: 10.3389/fnbot.2021.692539

INTRODUCTION

Disability of the lower body or related body parts will lead to walking difficulties (Jung et al., 2015). Gait recovery is one of the main goals of patients with lower limb motor dysfunction (Wolbrecht et al., 2008). The traditional rehabilitation process is labor-intensive that several therapists are required throughout the training of one patient (Yang et al., 2019). The wearable lower limb rehabilitation exoskeleton is used for gait rehabilitation of patients with lower limb dysfunction (Yin et al., 2020; Céspedes et al., 2021), such as spinal cord injury, cerebral palsy, and stroke (Hobbs and Artemiadis, 2020; Nolan et al., 2020). A suitable lower limb rehabilitation exoskeleton will improve the life quality of patients with lower limb disorder greatly (Young and Ferris, 2017). In order to realize smooth human-machine coupling and achieve robot-facilitated rehabilitation training, it is necessary to synchronize the action of wearable lower limb rehabilitation exoskeleton with that of the body. Therefore, accurate classification of the gait phase is required. The classification of the gait phases correctly is critical for robots to assist timely (Wei et al., 2021).

The human leg sEMG signal can offer valuable motion information, such as symmetric and periodic motion in human gait (Deng et al., 2020; Gao et al., 2021; Yao et al., 2021), and it is characterized by simple signal acquisition, intuitive data, and the non-invasive acquisition method (Kim et al., 2018; Lin et al., 2020; Ma et al., 2020). Artificial neural networks have made great progress, are widely used in the field of classification, and have shown great performance (Adewuyi et al., 2016; Atzori et al., 2016). Therefore, sEMG of legs is combined with an artificial neural network (Cheng et al., 2020) in human gait phase classification (Lee et al., 2017). Morbidoni et al. (2019) proposed a deep learning method for classifying a swing phase and a stance phase. This method is mainly based on the sEMG signal and does not need to extract features from the signal. Through the test of 12 subjects, the accuracy is up to 92.6%, which proves the effectiveness of the sEMG signal in gait classification. Joshi et al. (2013) obtained sEMG data from human lower limbs and used the machine learning method to classify each stage of the gait cycle, which improved the classification accuracy of each stage of the gait cycle. Ziegier et al. (2018) proposed a method based on EMG data to classify the standing stage and the swing stage of the gait of healthy people by using bilateral leg muscle signals. This method introduces a new EMG feature, which is calculated according to the EMG of muscle pairs on both sides, and the classification accuracy of the proposed method reaches 96%. Di Nardo et al. (2021) studied the influence of different sEMG signal processing specifications and different numbers of sEMG sensors on the performance of the gait phase classification method based on neural network prediction and obtained an average accuracy of 93.4%. However, although the above methods have good performance, the accuracy still needs to be improved.

Plantar pressure is widely used in the research of gait phase classification (Joo et al., 2014; Xie et al., 2020). Luo et al. (2020) arranged plantar pressure sensors at the heel and toe and used the working state of plantar pressure sensors to classify gait stages, and the accuracy of this method reached 94.1%. Nazmi et al. (2019), respectively, arranged plantar pressure sensors under the heel and thumb, and divided the gait phase by analyzing the contact state between the heel and toe and the ground; the accuracy of this method reached 87.5 and 77%. Although the above method is enough to detect gait events, the accuracy of gait phase classification is not high, and the phase classification is relatively rough. In addition, Liu et al. (2016) used a single-joint angle to classify the gait phase, and the accuracy reached 94.45%, which proved the feasibility of the method. Grimmer et al. (2019) used the angle sensor to detect the stance and swing and obtained good results. However, the target achieved only by this method still needs to be improved.

Whether the above information can be fully combined to find out the accurate relationship to improve the accuracy of gait phase classification is an interesting problem. Therefore, a multi-information fusion method for gait phase classification in the lower limb rehabilitation exoskeleton is proposed to improve the classification accuracy in this article. Firstly, the gait phase classification experiments at different speeds were carried out and analyzed, and the accuracy of gait phase classification was significantly improved. Secondly, the gait phase classification results of multi-information and single information are compared. The experimental results show that the gait phase classification method based on multi-information fusion has good performance.

The structure of this paper is as follows: The second section introduces the gait phase classification system, gait information acquisition device, data preprocessing, and the neural network model for gait phase classification. The third section shows the design of the gait acquisition experiment and the software environment of the experiment. The fourth section is the result and discussion of the experiment.

METHODOLOGY

Gait Phase Classification System

A real-time gait classification system based on wireless multi-information fusion is designed and implemented. The gait classification system of human lower limb movement consists of the gait information acquisition part and the gait information processing part. The gait information acquisition part includes plantar pressure acquisition, knee angle acquisition, and sEMG signal acquisition. The acquired knee joint angle information and plantar pressure information are, respectively, transmitted to the single-chip microcomputer (Atmel atmega328p microprocessor, ATMEL Inc., USA) and two linear voltage modules (FRP resistance voltage converter, Telesky Inc., China). The output of the sensor that acquires the knee joint angle and plantar pressure is an analog quantity, and the single-chip microcomputer performs 50 Hz A/D sampling on it and records the time stamp at the same time. In the experiment, the microcomputer and the linear voltage module were integrated into an aluminum metal box with a length of 180 mm, a width of 160 mm, and thickness of 48 mm, and the aluminum was sealed and wrapped with tin foil to shield the interference of space clutter signals. The function of the linear voltage module is to convert the resistance signal of the thin-film pressure sensor into the voltage signal. The part of gait information processing is mainly a computer and the neural network model. The communication between the microcomputer and the linear voltage module and the computer uses the Lora wireless transmission module (Lora-01, Alientek Ltd, China) to transmit data information, which reduces the energy loss of wired transmission, and the redundancy of the connection line and is more convenient to wear and move at any time.

Figure 1 shows the structure of the gait phase classification system. The sEMG acquisition system acquires the sEMG signal of the gait movement of the human leg, communicates wirelessly with the myoRESEARCH software in the computer, and transmits it to the neural network model of the computer. The Hall angle sensor acquires the knee joint angle information of human gait movement. The plantar pressure sensor collects the pressure information of the plantar in the stance phase of the human gait movement. The collected information is synchronized using a synchronization cable. The input to the neural network was the sEMG signal and the knee joint angle. After training with a label of gait events detected by plantar

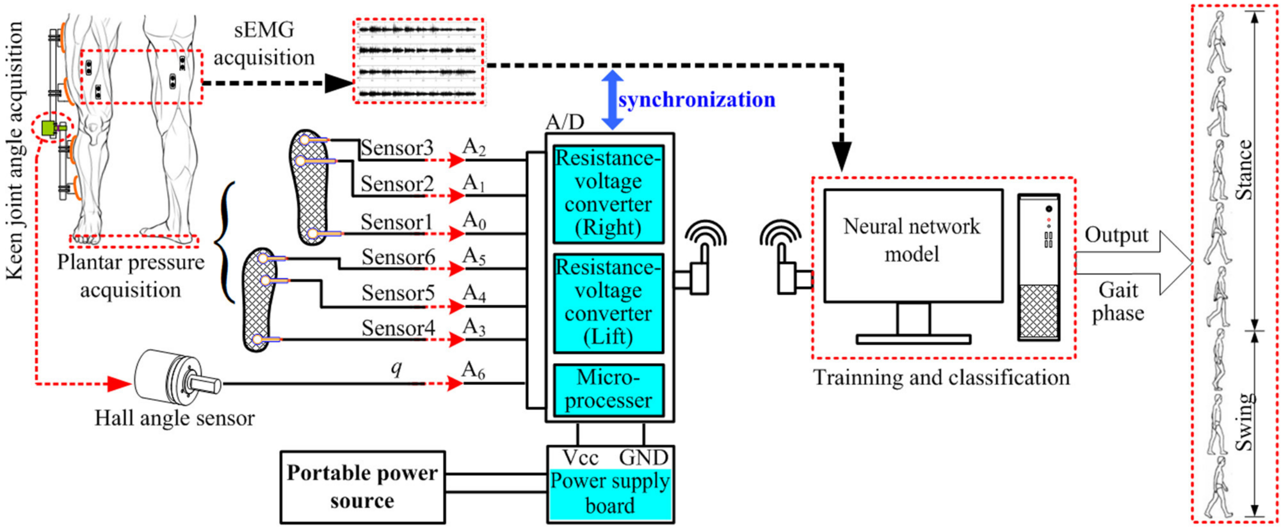
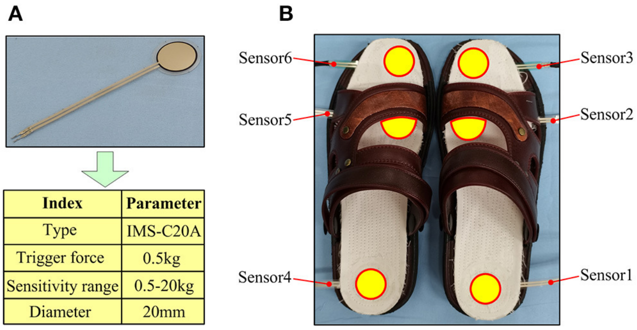


FIGURE 1 | Structure of gait phase classification system.



Index	Parameter
Type	IMS-C20A
Trigger force	0.5kg
Sensitivity range	0.5-20kg
Diameter	20mm

FIGURE 2 | The composition of the plantar pressure collection device; **(A)** the thin-film pressure sensor and its parameters, **(B)** the physical product of the plantar pressure collection device.

pressure information, it can output gait phase classification results in real time.

Gait Information Acquisition

Plantar Pressure Acquisition Device

The information collection of plantar pressure acquisition device is realized by a thin-film pressure sensor (IMS-C20A, Vicos Digital Tech. Ltd, China), which is, respectively, arranged in a multilayer cotton insole, as shown in **Figure 2**. Six identical thin-film pressure sensors are, respectively, placed on the heel, middle, and front of the two insoles to collect plantar pressure information on the heel, sole, and toe of the feet. When the sensor is being compressed, the amplified piezoelectric voltage is saved to the computer in the form of a digital signal through A/D. The specific parameters of the thin-film pressure sensor and the corresponding acquisition position relationship are shown in **Table 1**.

TABLE 1 | Corresponding acquisition position of a thin-film pressure sensor.

Thin-film pressure sensor		Position of acquirement
Left foot	**Right foot**	
Sensor 6	Sensor 3	Toe
Sensor 5	Sensor 2	Sole
Sensor 4	Sensor 1	Heel

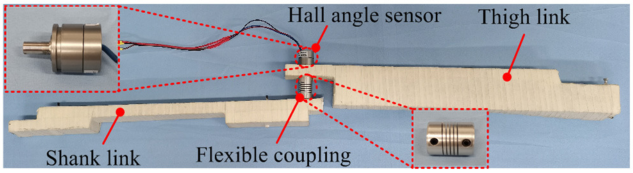


FIGURE 3 | A knee joint angle acquisition device.

Knee Joint Angle Acquisition Device

The knee joint angle acquisition device is composed of a Hall angle sensor (GT-B, Taizhou QT tech. Ltd, China), a 2-link, a flexible coupling, and several straps, as shown in **Figure 3**. The Hall angle sensor is a shaft-type angle measurement sensor. Its effective angle is 180 degrees, and the resolution is 0.18 degrees. The Hall angle sensor was installed on one end of the flexible coupling, which connects the shank and the thigh link. The function of the flexible coupling is to prevent the upper and lower links from being too rigid when the knee joint is moving, causing discomfort to the knee joint movement. The axis of the angle sensor was aligned with the human knee joint according to different individuals in order to fully synchronize the human leg and the knee joint.

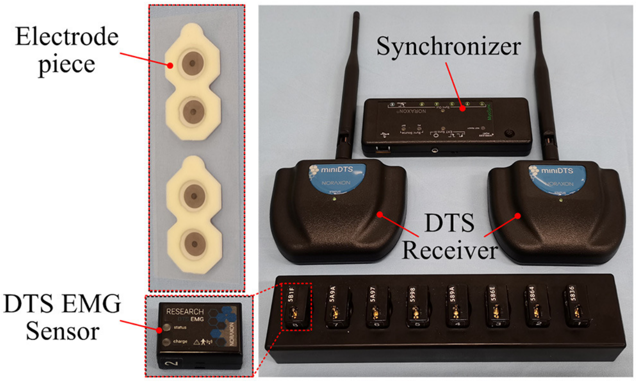


FIGURE 4 | An sEMG acquisition device.

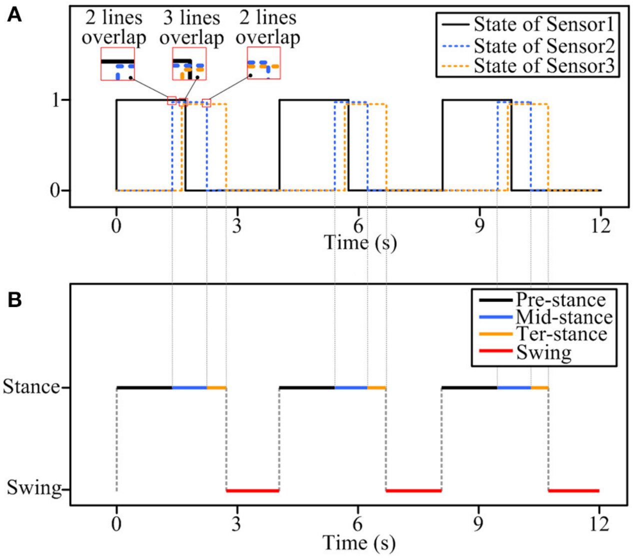


FIGURE 5 | The gait phase classification process, **(A)** 0 and 1 states of the thin-film pressure sensor, **(B)** gait sub-phase classification.

sEMG Data Acquisition

In the gait movement information acquisition experiment, the sEMG acquisition device used is an 8-channel Ultium-EMG sEMG signal acquisition instrument developed by Noraxon, USA, as shown in **Figure 4**. This device is a wireless transmission device that can acquire eight channels of sEMG signals with an acquisition frequency of up to 1,500 Hz. The timestamp of acquiring sEMG signals can be recorded simultaneously. The whole system includes eight sEMG signal sensors to obtain the sEMG signal of the human body, two receivers (Mini DTS Receiver) to transmit the acquired sEMG signal, a synchronizer to synchronize receiver data, and a sensor charger to turn on and off the sensor and charge the sensor. Combined with the myoMUSCLE software platform provided by the company, the sEMG signal can be acquired and simply processed in real time.

Data Preprocessing

The Signal Denoising Method

The human sEMG signal is complex and feeble, and it is susceptible to the influence of many external factors, such as the signal acquisition device, the experimental environment, and the physical condition of the subject, resulting in the acquired sEMG signal containing a lot of external noise. During the experiment, the main noise of the sEMG signal is especially the power frequency interference and motion artifacts. Butterworth filter has the characteristics of a flat frequency response curve in the passband, no fluctuation, frequency response gradually drops to 0 in the stopband, and a steeper frequency response decline curve, which is often used for noise reduction of the sEMG signal (Gui et al., 2019; Li Z. et al., 2020; Ma et al., 2020). Therefore, to retain the useful data information in the collected sEMG signal and eliminate the interference noise during the experiment, the 20–450 Hz 4th-order Butterworth filter is used for filtering, and then the 50 Hz 2nd-order notch filter is used to eliminate the power frequency interference and can obtain effective sEMG data for the subsequent data analysis.

Data Set Construction

The segmented sEMG with gait phase labels and knee joint angle data is required as input to the classification model. Because gait movement is the symmetrical movement of the left and right feet, the right foot is chosen as the research object. In order to simplify the segmentation process, three complete gait cycles are selected, and the acquired sEMG data are divided into gait phases according to the on-off state of the plantar pressure sensor. In the plantar pressure acquisition device, three thin-film pressure sensors measure the force between the heel, sole, and toe, and, according to the working state and working time of the three thin-film pressure sensors of the right foot, the gait is divided into four substages, namely, pre-stance, mid-stance, ter-stance, and swing phase. **Figure 5** shows the classification process of the gait phase. In **Figure 5A**, the sensor is working when it is under pressure, which is represented by "1," and when it is not under pressure, it is represented by "0." In **Figure 5B**, red, black, blue, and orange correspond to the swing phase, the pre-stance phase, the mid-stance phase, and the ter-stance phase, respectively. The states of 0 and 1 of the thin-film pressure sensors reflect different gait substages at different times. When the thin-film pressure sensors sensor1 (4), sensor2 (5), and sensor3 (6) are all "0," the gait phase is in the swing phase. When sensor1 (4) is "1" and sensor2 (5) and sensor3 (6) are "0," it means that the heel touches the ground, the sole and toe do not, and the gait phase is in the pre-stance phase. When more than two of the sensor1 (4), sensor2 (5), and sensor3 (6) in the thin-film pressure sensor are "1," and the thin-film pressure sensor corresponding to the sole remains "1," more than two lines appear to overlap on the image. It shows that there are three situations: heel and sole contact the ground at the same time, but toe does not contact, or heel, sole, and toe contact the ground at the same time, or heel does not contact, sole and toe contact the ground at the same time, at this time, the gait phase is in the mid-stance phase. It is worth noting that, in practice, there is still a state, that is, sensor1 (4) and sensor3 (6) are not working, sensor2 (5) is working, and

TABLE 2 | The relationship between the gait phase and the state of the thin-film pressure sensor.

Gait phase	Plantar pressure sensor		
	Sensor1(4)	Sensor2(5)	Sensor3(6)
Swing	0	0	0
Pre-stance	1	0	0
Mid-stance	1	1	0
	1	1	1
	0	1	0
	0	1	1
Ter-stance	0	0	1

it is also in the mid-stance phase. When sensor1 and sensor2 (5) are "0" and sensor3 (6) is "1," it means that only the toe of the foot contacts the ground, and the gait phase is at the ter-stance phase. **Table 2** shows the corresponding relationship between the working state of the thin-film pressure sensor of the right foot and each substage of the gait phase. After the gait classification is completed, gait data are generated. The gait data, plantar pressure, and joint angle data have the same length.

Select the classic machine learning neural network model support vector machine (SVM) (Li et al., 2015) and deep learning neural network model long short-term memory (LSTM) (Liu et al., 2018) and back-propagation neural network (BPNN) (Chen et al., 2018) to compare with CNN. After filtering the sEMG data, use the sliding window to extract the mean absolute value (MAV) and root mean square (RMS) features of the sEMG signal, which can be expressed as:

$$MAV = \frac{1}{N}\sum_{i=1}^{n}|x_i| \tag{1}$$

where N is the number of sample points in the sampling window, x_i is the amplitude of the i-th sEMG sample point.

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{2}$$

where N is the number of sample points in the sampling window, x_i is the amplitude of the i-th sEMG sample point, $\bar{x}$ represents the average value of sEMG data in this window.

In this article, six-channel sEMG data of human lower limbs and one channel knee joint angle data are acquired. The sliding window method is used to extract RMS and MAV features from the raw sEMG data output by the sEMG sensor. The number of sample points in the sliding window is 30, and the sliding step length is 30. The sEMG feature data can be obtained after feature extraction processing of the sEMG data. At this time, the original sEMG data of each channel will generate two-channel (RMS and MAV) sEMG feature data, and the number of sEMG feature data channels will be changed from the 6 channels to 12 channels. At the same time, the length of sEMG feature data, the length of knee joint angle data, and the length of gait data are the same. **Figure 6** shows the process of data processing.

Since the length of the sEMG feature data is the same as the length of the knee joint angle data, after the feature extraction process is completed, the knee joint angle data and the sEMG feature data are converted into an input feature image matrix. The number of sEMG feature data channels is 12, which, together with one channel of knee joint angle data, forms a 13-dimensional input feature image matrix. The input feature image matrix is slidingly intercepted by the sliding window method, and the length of the sliding window size is set to 20; therefore, the feature image matrix size is 20 × 13 × 1, and then is input into the neural network model. The knee joint angle data and feature data of the sEMG data are used as input. The total sample of sEMG data is 22,500, of which the first 80% is allocated as the training set and the last 20% as the test set.

Since the length of the sEMG feature data is the same as the length of the gait data, the time stamp corresponding to each sample point in the gait data and the sEMG feature data is consistent. When using the sliding window method to intercept the sEMG feature image matrix, the gait data corresponding to the sample points at the end of the sEMG feature image matrix is used as the label of the sEMG feature image matrix.

At this point, the training and testing data sets input to the neural network can be obtained. The neural network input data in the data set include the original sEMG signal after feature processing data and the knee joint angle data of the lower limbs, and the gait data as the label data in the data set. The sEMG signal data and the lower limb joint angle data are features fused through the convolutional neural network (CNN) to realize the gait classification.

A Neural Network Model for Gait Phase Classification

Convolutional Neural Network is a feed forward neural network (Chen et al., 2019) and is the most commonly used network model in the field of deep learning (Zhai et al., 2017). **Figure 7** shows the architecture of a CNN for gait phase classification. In this article, the dimension of the input data into the neural network model CNN is low, and the input data will be lost after adding the dimension reduction operation of the pooling layer, so the pooling layer is removed from the CNN model, and only the convolution layer exists. This will not affect the function of the CNN model and make its structure more concise. It also improves the training speed of the CNN model and the output speed of the gait phase classification results.

The model super parameter epochs are set to 300, the batch size is set to 100, softmax function is used as the activation function of the last layer of the model, Adam optimizer is used to update the model parameters, and the initial learning rate is set to 0.001. During training, the cross-entropy loss function is used to optimize the output. Softmax function and cross-entropy loss function are shown in formula (3) and formula (4)

$$\widehat{y}_i = softmax(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^{m}\exp(x_j)},\ i = 1, 2, ..., m \tag{3}$$

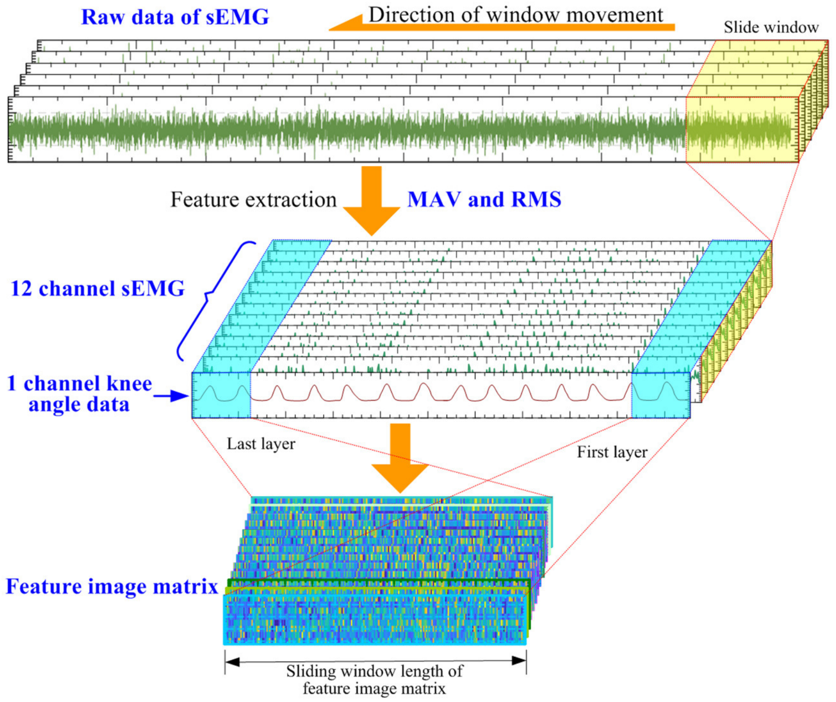


FIGURE 6 | The process of data processing.

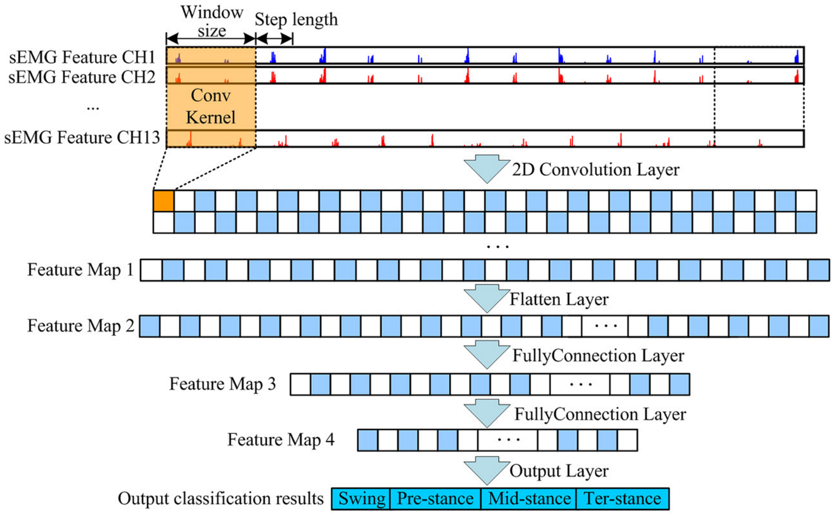


FIGURE 7 | CNN architecture of the gait classification model.

$$J = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\widehat{y}_i) \qquad (4)$$

where x_i is the input value of each node in the output layer, $\widehat{y}_i$ is the probability of actual output,y_i is the category label, m is the number of categories, and N is the number of batches input to the model at one time.

EXPERIMENT

Experimental Design

Five able-bodied subjects took part in this experimental study, aged between 24 and 28, height between 168 cm and 185 cm, and weight between 60 and 70 kg, and had not taken any strenuous exercise before the experiment. Before the experiment, to ensure that the experiment is effective, the following steps should be carried out:

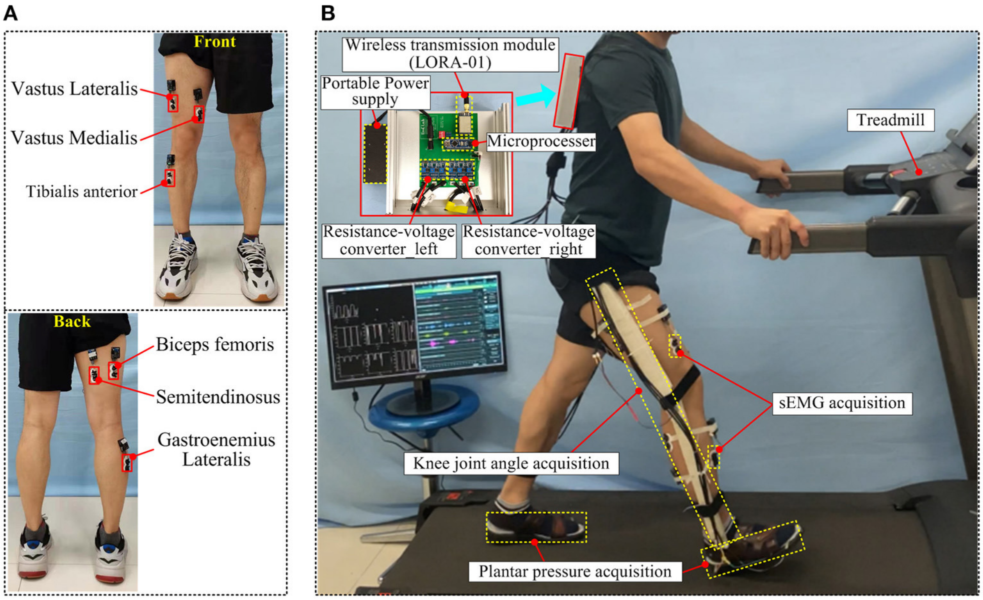


FIGURE 8 | The gait information acquisition device, **(A)** the installation position of the sEMG sensor, **(B)** the wearing effect and information acquisition process of the gait information acquisition device.

- Clean the skin: remove the body hair on the tested muscle and wipe it with medical alcohol.
- Equipment placement: paste the electrode piece at the position of the muscle to be measured, and paste the electrode piece along with the muscle fiber of the leg, which is conducive to signal acquisition. Place the sEMG sensor about 2 cm away from the electrode and connect the electrode correctly. The electrode piece and the sEMG sensor are fixed with medical adhesive tape to prevent falling off during movement.
- Equipment detection: check the paste of the electrode sheet to ensure the paste is tight. Start the sEMG acquisition device; check the transmission status of each channel to ensure the normal transmission of the sEMG signal.

Six sEMG sensors are arranged in muscle positions: vastus medialis (VM), vastus lateralis (VL), semitendinosus (ST), biceps femoris (BF), tibialis anterior (TA), and gastrocnemius lateralis (GA). When the EMG sensor is installed, wear other gait information acquisition equipment. The knee angle acquisition device is arranged on the outside of the thigh with an adhesive bandage, and the position of the Hall angle sensor is on the same axis with the rotation center of the knee joint, to ensure that the thigh rod and leg rod will not affect the rotation of knee joint (motion interference) when they move with the leg. Three thin-film pressure sensors embedded in the front, middle, and back of the insole were used to acquire the pre-stance, mid-stance, and ter-stance of the gait phase. **Figure 8** shows the gait information acquisition device. Each muscle position corresponds to an sEMG acquisition device channel, and the corresponding relationship between the muscle and sEMG sensor channel is shown in **Table 3**.

TABLE 3 | Corresponding channels of the sEMG sensor and muscle.

Muscle location		Channel
Thigh	Vastus Medialis	1
	Vastus Lateralis	2
	Biceps Femoris	3
	Semitendinosus	4
Shank	GastroenemiusLateralis	5
	Tibialis anterior	6

Different Speeds

Five subjects were tested with different gait speeds. Taking into account the conditions of healthy people, lower limb dyskinesia, and the elderly, the walking experiments were carried out at 1, 2, and 3 km/h, respectively, and the gait data were collected at three speeds. Among them, the speed of 3 km/h is close to the daily gait speed of normal people, while the speed of 2 and 1 km/h is gradually lower than the daily gait speed. During the gait walking experiment, the subjects rest 15 min between each gait speed to ensure that the leg muscles are in a relaxed state, and muscle fatigue may cause the distortion of sEMG information and affect the classification results of the gait phase. Each subject was acquired three times of gait data at the same gait speed, and the subjects rest for 5 min in each gait data acquisition to ensure the relaxation of leg muscles and check whether the equipment is loose to avoid affecting the results of gait data acquisition.

Comparison of Multi-Information and Single Information

In order to verify the superiority of the proposed method, the comparative experiments of multi-information and single information were carried out at 1-, 2-, and 3-km/h gait speeds. Plantar pressure is the label of sEMG data. Multi-information is to collect sEMG data of legs and knee joint angle data at the same time and input them into the gait classification model at the same time to classify the gait phase. Single information only collects sEMG information of legs during gait movement and inputs it into the gait classification model to classify the gait phase. At the same time, the multi-information and single information are compared with four gait classification models (SVM, BPNN, LSTM, and CNN), and the classification performance of different classification models with multi-information and single information input is obtained.

Five-Fold Cross-Validation

Cross validation is a common method used to verify the performance of the model in the process of modeling (Jung, 2018). It divides the original data into the training set and the test set. First, the training set is used to train the model, and then the test set is used to test the trained model so as to evaluate the performance of the model. In this paper, the 5-fold cross validation method is used to evaluate the model. Data of each subject are divided into five subsets. Each time, any subset is taken as the test set and the rest as the training set. After that, five models can be obtained. Finally, the average accuracy of the test set is taken as the evaluation index of the subject under the 5-fold cross validation method.

Software Environment

The neural network model CNN of gait phase classification used in this study is compiled on the deep learning network framework Keras 2.3.0. The Keras network framework is an open-source artificial neural network library written in Python language, which can be used as the advanced application program interface (API) of TensorFlow. In this article, the python libraries used include NumPy, Sklearn, SciPy, and Matplotlib. The whole model implementation process is implemented on Pycharm software, and the model training is completed on a computer with an independent GPU. The specific configuration of the computer is shown in **Table 4**.

TABLE 4 | Computer configuration information.

Index	Parameter
Central Processing Unit (CPU)	Intel Core i5 4570
Graphics Processing Unit (GPU)	Nvidia GTX1070 8GB
Operating system	Windows10
Computer memory	DDR3 1600 16GB
Software environment	Python 3.7.6

RESULT AND DISCUSSION

Accuracy is a key index of human gait classification (Gao et al., 2021). In this article, two evaluation indexes are used, accuracy and F1-score, which can be expressed as:

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{5}$$

$$\text{F1-score} = \frac{2T_P}{2T_P + F_P + F_N} \tag{6}$$

where $T_P$indicates correctly identifying positive samples as positive, $F_N$indicates wrongly identifying positive samples as negative, $F_P$indicates wrongly identifying negative samples as positive, and $T_N$indicates correctly identifying negative samples as negative.

Different Speeds

As shown in **Figures 9–11**, the classification results of four models (CNN, LSTM, BPNN, and SVM) of five subjects (P1, P2, P3, P4, and P5) with gait movement of 1 km/h, 2 km/h, and 3 km/h are shown. The classification accuracy and F1-score of the four models are different. In terms of **Figure 9A**, the classification effect of five subjects in the CNN model is the best, the classification accuracy of each subject is higher than that of the other three models, the prediction results of five subjects are between 93 and 98%, and the standard deviation of prediction results each subject is relatively small, indicating that the prediction results of the model are relatively stable. Compared with CNN, the classification effect of LSTM, BPNN, and SVM is unsatisfactory, the LSTM has the highest accuracy of 92% in five subjects, BPNN has the highest accuracy of 90% in five subjects, and SVM has the worst effect of 82%. In addition, the classification accuracy of LSTM, BPNN, and SVM in five subjects fluctuates greatly, and the standard deviation is also large, which indicates that the results of gait phase classification are unstable. **Figure 9B** shows five subjects in four neural network models F1-score. In the F1-score evaluation, the performance of CNN is better than the other three. The F1-score fluctuation of five subjects is relatively small, concentrated in 91–92%, and the standard deviation of the F1-score of each subject is also small, indicating a better classification effect. Compared with CNN, the other three models performed mediocrely in five subjects.

Figures 10, **11** show the classification results of the four models for five subjects at 2 and 3-km/h gait speeds. With the same trend of 1-km/h gait speed, the classification results of CNN are better than those of SVM, BPNN, and LSTM. However, with the increase of gait speed, the accuracy and F1-score of the four models are decreased. At 3 km/h, the classification accuracy of five subjects in the CNN model is about 90%, and the F1 score is up to 82%.

In terms of the analysis in **Figures 9–11**, we can see that, under the 5-fold cross-validation method, the CNN model outperformed the other three. The classification results of the five subjects are better than the other three models, and the classification effect is stable. It can well realize the classification of the gait phase, and the accuracy and F1-score have good performance, which proves the superiority and generalization ability of the CNN model. In addition, different gait speeds also have a great influence on the results of the gait phase classification of the four models. From the perspective of the four models as a whole, the accuracy of gait phase classification of the model is higher at lower gait speed, while the accuracy of gait

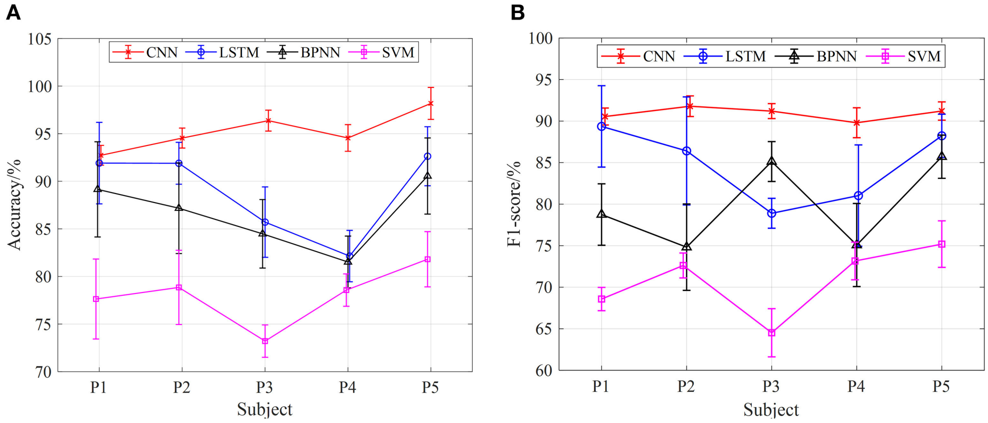


FIGURE 9 | The results of gait phase classification of different models with the 5-fold cross-validation method at 1 km/h, **(A)** the average accuracy of gait phase classification of different models of five subjects, **(B)** the average F1-score of gait phase classification in different models of five subjects.

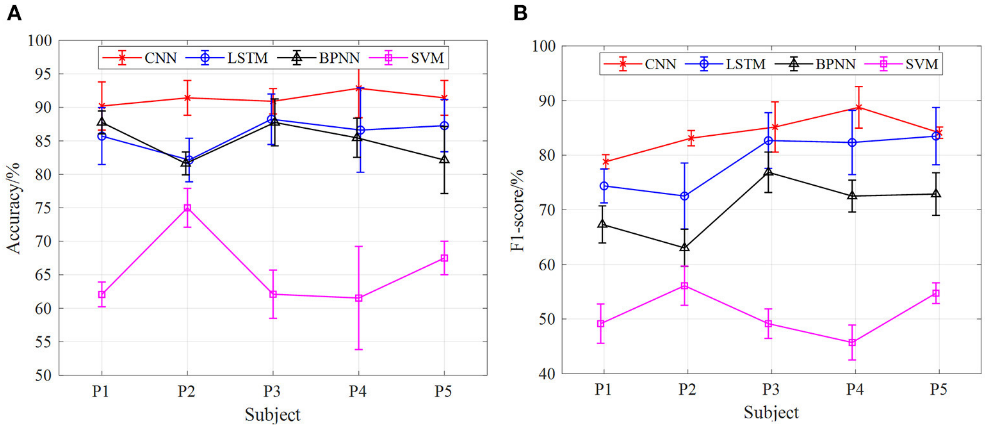


FIGURE 10 | The results of gait phase classification of different models with the 5-fold cross-validation method at 3 km/h, **(A)** the average accuracy of gait phase classification of different models of five subjects, **(B)** the average F1-score of gait phase classification in different models of five subjects.

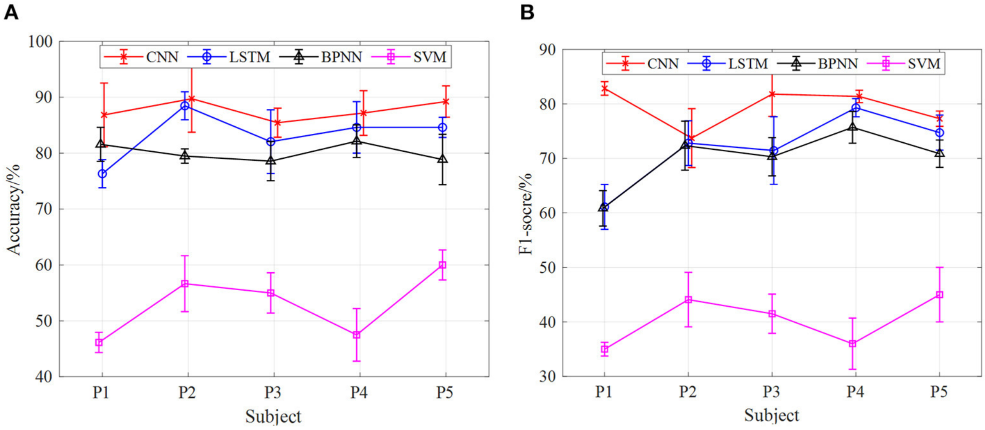


FIGURE 11 | The results of gait phase classification of different models with the 5-fold cross-validation method at 3 km/h, **(A)** the average accuracy of gait phase classification of different models of five subjects, **(B)** the average F1-score of gait phase classification in different models of five subjects.

phase classification of the model decreases significantly with the increase of gait speed. From **Figures 9–11**, when the gait speed is 1 km/h, the accuracy of CNN, LSTM, BPNN, and SVM is 98, 92.5, 90.5, and 82% respectively, and the highest value of F1-score is 92, 89.5, 85, and 75%, respectively; when the gait speed is 3 km/h, the accuracy of CNN, LSTM, BPNN, and SVM is 90, 89, 82, and 90%, respectively, and the highest value of F1-score is 82, 80, 76, and 45%, respectively.

Figure 12 shows the average confusion matrix of the accuracy of the three speeds of the best classification model CNN in the 5-fold cross-validation method. The confusion matrix provides visualization of the classification performance of the gait substages. The vertical axis of the matrix represents the real category of the test data set, and the horizontal axis of the matrix represents the corresponding classification results. The values of these three confusion matrices are the average classification accuracy of all objects under three different gait speeds. In terms of **Figure 12**, the gait phase substages of the three speeds have excellent classification results, and the gait phase substages of each speed can be divided. It can be seen from the figure that the classification accuracy of the swing and the pre-stance is above 99.14 and 92.91%, respectively, under the three speeds. Especially when the speed is 1 km/h, the classification accuracy reaches the highest, 99.50 and 99.14%, respectively. In the mid-stance phase and the ter-stance phase, the classification accuracy performance is undistinguished. Compared with the swing phase and the pre-stance phase, the highest classification accuracy at 1 km/h is only 92.44 and 86.67%, and the highest classification accuracy at 2 km/h is 89.32 and 85.33%, which dropped by 3.12 and 1.34%, respectively; when the gait speed is 3 km/h, the classification accuracy is 85.08 and 79.49%, which dropped by 7.36 and 7.18%, respectively.

In terms of **Figure 12**, gait speed has a great influence on the classification of gait phases. At a gait speed of 1 km/h, the classification effect of the gait phases is better, and the classification accuracy of the substage of the gait phases is high, especially for the swing phase. When the gait speed is 3 km/h, the classification effect of the substage of the gait phase is lower than 1 km/h, and the accuracy also gradually decreases. For the mid-stance and ter-stance, the classification accuracy of each speed is significantly lower than the swing and pre-stance, and the main reason is that the number of samples acquired during the movement is insufficient, which leads to the classification effect becoming mediocre.

Comparison of Multi-Information and Single Information

In addition, under the model of the 5-fold cross-validation method, we carried out comparative experiments of different input signals to the neural network model. At the same speed, the accuracy of gait phase classification and the average value of F1-score of five subjects (P1, P2, P3, P4, and P5) were taken to compare the classification effect of sEMG and sEMG + angle input to SVM, BPNN, LSTM, and CNN neural network models. **Figures 13–15** show the gait classification results of different input signals in the four models under three motion speeds of 1, 2, and 3 km. In terms of **Figure 13A**, when sEMG + angle is used as input, the result of gait phase classification is better than that of sEMG alone, and the classification accuracy of CNN is the highest, and the average accuracy of five subjects is close to 95%. The second was LSTM. When sEMG + angle was used as input, the average accuracy of five subjects was close to 89%. BPNN is worse than CNN and LSTM. When sEMG + angle is used as input, the average accuracy of five subjects is about 87.5%, and the SVM classification effect is the worst, about 78%. When sEMG was used as input alone, the average accuracy of five subjects in the CNN model was close to 90%, and the accuracy of LSTM, BPNN, and SVM was about 82.5, 80.5, and 76%, respectively. It can be seen from the figure that, when sEMG + angle is used as input, compared with sEMG alone, the classification accuracy of SVM, BPNN, LSTM, and CNN is improved by 2.6, 8, 7.3, and 5.6%, respectively. At the same time, F1-score is used as the evaluation index. As shown in **Figure 13B**, when sEMG + angle is used as the input of the neural network model, the classification result is better than that of sEMG alone. As before, taking the average value of F1 score of five subjects, the average value of F1-score of SVM, BPNN, LSTM, and CNN neural network models reaches 70, 87.5, 89, and 95%, respectively, which is 10, 20, 18.5, and 14.7% higher than that of sEMG alone.

Figures 14, **15** are the comparison results of 2- and 3-km/h gait speed, respectively, and the overall trend is the same as **Figure 13**. The classification accuracy and F1-score of the four neural network models are gradually increasing from SVM to CNN, and the classification results with sEMG + angle as input are better than those with sEMG as input alone. It is worth noting that when the gait speed is 3 km/h, the F1-score of LSTM is higher than sEMG + angle, which is different from the results in **Figures 13**, **14**. The reason for this phenomenon is that, among the five subjects, the effect of information collection is not good due to the fast gait speed or unstable walking posture; when data are input into the neural network, the output F1-score is low, resulting in low average F1-score. In general, when sEMG + angle is used as input, the output results of four neural network models are better than that of sEMG alone. Therefore, multi-information has a satisfactory classification effect for gait phase classification.

Table 5 shows the classification time of several classification models on a single sample and the delay time of each sensor of the gait acquisition device. In the gait information acquisition experiment, the delay time of the sEMG sensor is 114 ± 5 ms, and the delay time of the sensor of the knee joint angle acquisition device is about 0.1 ± 0.03 ms, and the delay time of the sensor of plantar pressure acquisition device is about 0.12 ± 0.05 ms. At the same time, in the process of gait phase model classification, the single sample classification time of SVM, BPNN, LSTM, and CNN is 0.18, 0.33, 1.53, and 0.37 ms, respectively. Because sEMG information acquisition, knee angle information acquisition, and plantar pressure information acquisition are carried out at the same time, the delay time of the sEMG sensor includes the sensor delay time of the knee joint angle acquisition device and the plantar pressure acquisition device. It is known from Li K. et al. (2020), the delay time of human sEMG action is about 150 ms, and the delay time of the sEMG sensor and the model for single sample classification is <150 ms, so the gait

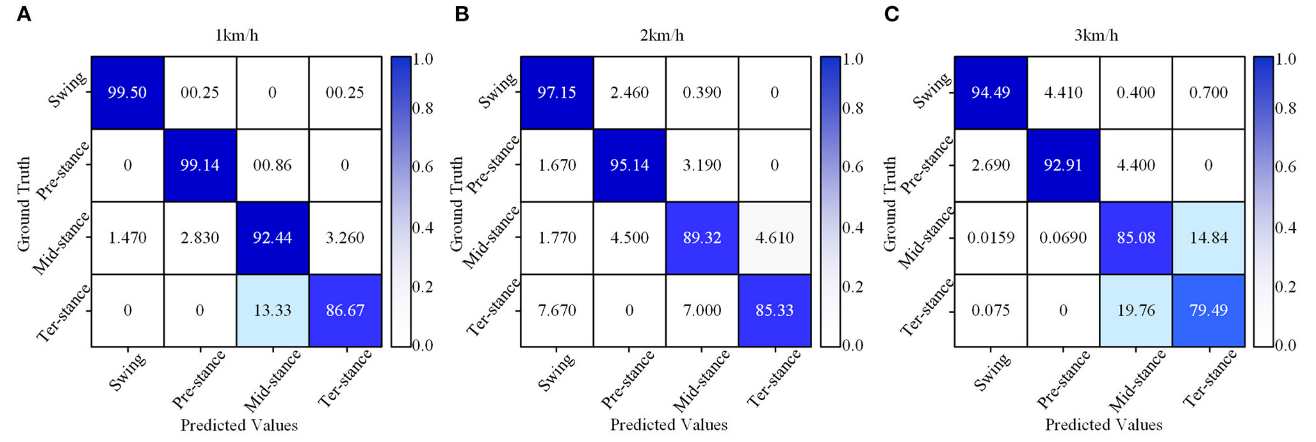


FIGURE 12 | The accuracy confusion matrix of different gait speeds of the CNN model at the 5-fold cross-validation method, **(A)** a confusion matrix of 1 km/h, **(B)** a confusion matrix of 2 km/h, **(C)** a confusion matrix of 3 km/h.

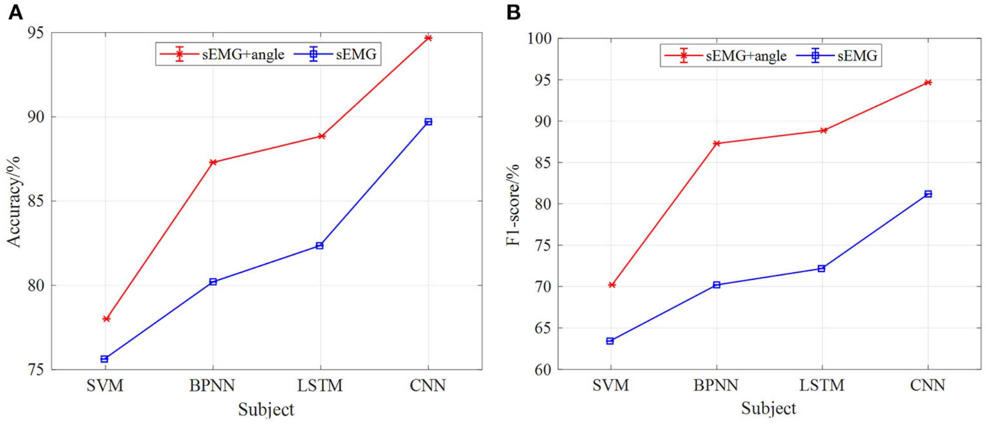


FIGURE 13 | The comparison results of different models of multi-information and single information input with the 5-fold cross-validation method at 1 km/h, **(A)** comparison results of four model accuracy with different input information, **(B)** comparison results of four models F1-score with different input information.

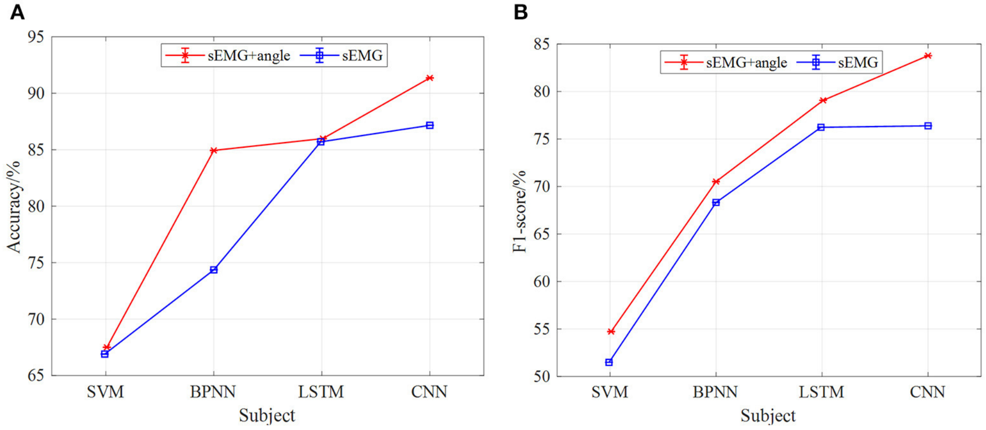


FIGURE 14 | The comparison results of different models of multi-information and single information input with the 5-fold cross-validation method at 2 km/h, **(A)** comparison results of four model accuracy with different input information, **(B)** comparison results of four models F1-score with different input information.

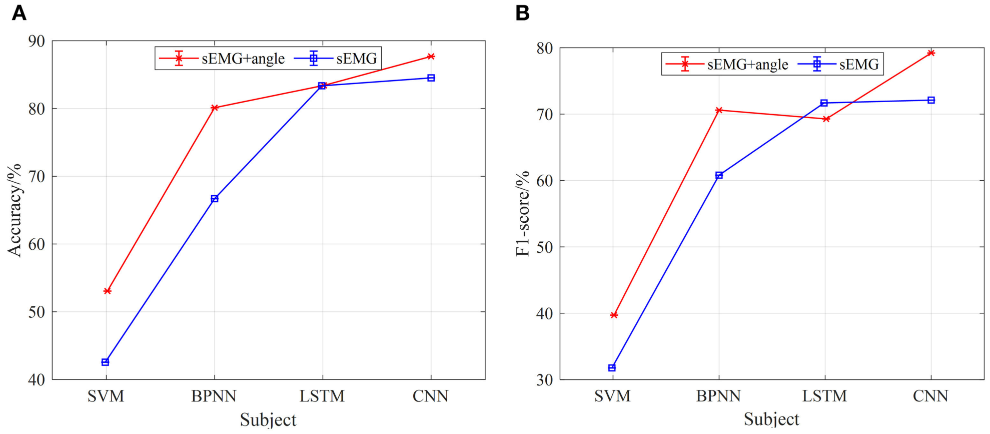


FIGURE 15 | The comparison results of different models of multi-information and single information input with the 5-fold cross-validation method at 3 km/h, **(A)** comparison results of four model accuracy with different input information, **(B)** comparison results of four models F1-score with different input information.

TABLE 5 | Sensor delay time and classification time of a single sample.

Index	SVM	LSTM	BPNN	CNN
sEMGsensor delay time	114 ± 5 ms	114 ± 5 ms	114 ± 5 ms	114 ± 5 ms
Keen angle acquisition sensor delay time	0.1 ± 0.03 ms	0.1 ± 0.03 ms	0.1 ± 0.03 ms	0.1 ± 0.03 ms
Plantar pressure acquisition sensor delay time	0.12 ± 0.05 ms	0.12 ± 0.05 ms	0.12 ± 0.05 ms	0.12 ± 0.05 ms
Classification time of single sample	0.18 ms	0.33 ms	1.53 ms	0.37 ms

phase classification system has real-time performance. Besides, although the classification time of CNN is slightly higher than SVM and BPNN in a single sample, the classification effect of CNN is better than SVM and BPNN, so CNN is selected as the neural network model of gait phase classification.

CONCLUSION

In this article, a gait phase classification method based on multi-information fusion is proposed. The principle of the gait phase classification system, the structure of the gait information acquisition device, and the data preprocessing method are given. The performance and the generalization ability of the model are proved by 5-fold cross-validation. Two task experiments using the proposed multi-information fusion gait phase classification method were carried out. In the 5-fold cross-validation method, the experimental results demonstrated that the average accuracy and the average F1-score of the proposed method reach 98 and 92%, respectively, for the gait phase classification at 1-km/h gait speed. For different input information experiments, in the case of three gait speeds, the classification effect of multi-information is far better than that of single information. At the same time, the delay time of the sEMG sensor, the Hall sensor of knee angle acquisition device, and the thin-film pressure sensor of plantar pressure acquisition device was measured, and the time of the neural network model to classify a single sample was also measured, which proved the real-time performance of gait phase classification system. Due to the integration of the information directly related to gait movement, the proposed multi-information fusion method for gait phase classification is better than the reported (Liu et al., 2016; Luo et al., 2020) gait classification method or system.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Medical Ethics Committee, Department of Medicine, Shenzhen University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YZ, GC, and AZ: methodology. YZ and ZL: software. YZ, ZL, SC, BH, and HC: performed data collection. GC: funding acquisition, project administration. YZ: analysis of the collected data and wrote the first draft of the manuscript. YZ and WL: writing-review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China under Grant NSFC U1813212.

ACKNOWLEDGMENTS

The authors would like to thank the five subjects for participating in this study.

REFERENCES

Adewuyi, A. A., Hargrove, L. J., and Kuiken, T. A. (2016). Evaluating eMg feature and classifier selection for application to Partial-hand Prosthesis control. *Front. Neurorobot.* 10:15. doi: 10.3389/fnbot.2016.00015

Atzori, M., Cognolato, M., and Müller, H. (2016). Deep learning with convolutional neural networks applied to electromyography data: a resource for the classification of movements for prosthetic hands. *Front. Neurorobot.* 10:9. doi: 10.3389/fnbot.2016.00009

Céspedes, N., Raigoso, D., Múnera, M., and Cifuentes, C. A. (2021). Long-term social human-robot interaction for neurorehabilitation : robots as a tool to support gait therapy in the pandemic. *Front. Neurorobot.* 15:612034. doi: 10.3389/fnbot.2021.612034

Chen, J., Zhang, X., Cheng, Y., and Xi, N. (2018). Surface EMG based continuous estimation of human lower limb joint angles by using deep belief networks. *Biomed. Signal Process. Control* 40, 335–342. doi: 10.1016/j.bspc.2017.10.002

Chen, Y., Yang, Y., Wang, W., and Kuo, C. C. J. (2019). "Ensembles of feedforwarddesigned convolutional neural networks," in *IEEE International Conference on Image Processing ICIP* (Taipei: IEEE), 3796-3800. doi: 10.1109/ICIP.2019.8803610

Cheng, H. R., Cao, G. Z., Li, C. H., Zhu, A., and Zhang, X. (2020). "A CNN-LSTM hybrid model for ankle joint motion recognition method based on sEMG," in *2020 17th International Conference on Ubiquitous Robots*. UR 2020 (Kyoto: IEEE), 339–344. doi: 10.1109/UR49135.2020.9144698

Deng, Y., Gao, F., and Chen, H. (2020). Angle estimation for knee joint movement based on PCA-RELM algorithm. *Symmetry* 12, 1–16. doi: 10.3390/sym12010130

Di Nardo, F., Morbidoni, C., Cucchiarelli, A., and Fioretti, S. (2021). Influence of EMG-signal processing and experimental set-up on prediction of gait events by neural network. *Biomed. Signal Process. Control* 63:102232. doi: 10.1016/j.bspc.2020.102232

Gao, F., Tian, T., Yao, T., and Zhang, Q. (2021). Human gait recognition based on multiple feature combination and parameter optimization algorithms. *Comput. Intell. Neurosci*. 2021, 1–14. doi: 10.1155/2021/6693206

Grimmer, M., Schmidt, K., Duarte, J. E., Neuner, L., Koginov, G., and Riener, R. (2019). Stance and swing detection based on the angular velocity of lower limb segments during walking. *Front. Neurorobot.* 13:57. doi: 10.3389/fnbot.2019.00057

Gui, K., Liu, H., and Zhang, D. (2019). A practical and adaptive method to achieve EMG-Based torque estimation for a robotic exoskeleton. *IEEE/ASME Trans. Mechatronics* 24, 483–494. doi: 10.1109/TMECH.2019.2893055

Hobbs, B., and Artemiadis, P. (2020). A review of robot-assisted lower-limb stroke therapy: unexplored paths and future directions in gait rehabilitation. *Front. Neurorobot.* 14:19. doi: 10.3389/fnbot.2020.00019

Joo, S., Bin, Oh, S. E., Sim, T., Kim, H., Choi, C. H., Koo, H., et al. (2014). Prediction of gait speed from plantar pressure using artificial neural networks. *Expert Syst. Appl.* 41, 7398–7405. doi: 10.1016/j.eswa.2014.06.002

Joshi, C. D., Lahiri, U., and Thakor, N. V. (2013). "Classification of gait phases from lower limb EMG: application to exoskeleton orthosis," in *IEEE EMBS Special Topic Conference on Point-of-Care (POC) Healthcare Technologies. Synergy Towards Better Global Healthcare*, PHT 2013 (Bangalore: IEEE), 228–231. doi: 10.1109/PHT.2013.6461326

Jung, J. Y., Heo, W., Yang, H., and Park, H. (2015). A neural network-based gait phase classification method using sensors equipped on lower limb exoskeleton robots. *Sensors* 15, 27738–27759. doi: 10.3390/s151127738

Jung, Y. (2018). Multiple predicting K-fold cross-validation for model selection. *J. Nonparametr. Stat.* 30, 197–215. doi: 10.1080/10485252.2017.1404598

Kim, M., Kim, K., and Chung, W. K. (2018). Simple and fast compensation of sEMG interface rotation for robust hand motion recognition. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 2397–2406. doi: 10.1109/TNSRE.2018.2878439

Lee, S. W., Yi, T., Jung, J. W., and Bien, Z. (2017). Design of a gait phase recognition system that can cope with EMG electrode location variation. *IEEE Trans. Autom. Sci. Eng.* 14, 1429–1439. doi: 10.1109/TASE.2015.2477283

Li, K., Zhang, J., Wang, L., Zhang, M., Li, J., and Bao, S. (2020). A review of the key technologies for sEMG-based human-robot interaction systems. *Biomed. Signal Process. Control* 62:102074. doi: 10.1016/j.bspc.2020.102074

Li, Q. L., Song, Y., and Hou, Z. G. (2015). Estimation of lower limb periodic motions from semg using least squares support vector regression. *Neural Process. Lett.* 41, 371–388. doi: 10.1007/s11063-014-9391-4

Li, Z., Zhang, D., Zhao, X., Wang, F., Zhang, B., Ye, D., et al. (2020). A temporally smoothed MLP regression scheme for continuous knee/ankle angles estimation by using multi-channel sEMG. *IEEE Access* 8, 47433–47444. doi: 10.1109/ACCESS.2020.2979008

Lin, M.-W., Ruan, S.-J., and Tu, Y.-W. (2020). A 3DCNN-LSTM Hybrid framework for sEMG-based noises recognition in exercise. *IEEE Access* 8, 162982–162988. doi: 10.1109/ACCESS.2020.3021344

Liu, D. X., Wu, X., Du, W., Wang, C., and Xu, T. (2016). Gait phase recognition for lower-limb exoskeleton with only joint angular sensors. *Sensors* 16, 1–21. doi: 10.3390/s16101579

Liu, J., Shahroudy, A., Xu, D., Kot, A. C., and Wang, G. (2018). Skeleton-based action recognition using spatio-temporal LSTM network with trust gates. IEEE Trans. *Pattern Anal. Mach. Intell.* 40, 3007–3021. doi: 10.1109/TPAMI.2017.2771306

Luo, R., Member, S., Sun, S., Zhang, X., Tang, Z., and Wang, W. (2020). A low-cost end-to-end sEMG-based gait sub-phase recognition system. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28:267276. doi: 10.1109/TNSRE.2019.2950096

Ma, X., Liu, Y., Song, Q., and Wang, C. (2020). Continuous estimation of knee joint angle based on surface electromyography using a long short-term memory neural network and time-advanced feature. *Sensor* 20, 1–18. doi: 10.3390/s20174966

Morbidoni, C., Principi, L., Mascia, G., Strazza, A., Verdini, F., Cucchiarelli, A., et al. (2019). Gait phase classification from surface EMG signals using neural networks gait phase classification from surface EMG signals using Neural Networks. *XV Mediterr. Conf. Med. Biol. Eng. Comput.* 76, 75–82. doi: 10.1007/978-3-030-31635-8_9

Nazmi, N., Abdul Rahman, M. A., Mohammed Ariff, M. H., and Ahmad, S.A. (2019). "Generalization of ann model in classifying stance and swing phases of gait using EMG signals," in *2018 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2018 - Proceedings* (Sarawak: IEEE), 461–466. doi: 10.1109/IECBES.2018.8626626

Nolan, K. J., Karunakaran, K. K., Chervin, K., Monfett, M. R., Bapineedu, R. K., Jasey, N. N., et al. (2020). Robotic exoskeleton gait training during acute stroke inpatient rehabilitation. *Front. Neurorobot.* 14:581815. doi: 10.3389/fnbot.2020.581815

Wei, P., Zhang, J., Tian, F., and Hong, J. (2021). A comparison of neural networks algorithms for EEG and sEMG features based gait phases recognition. *Biomed. Signal Process. Control* 68:102587. doi: 10.1016/j.bspc.2021.102587

Wolbrecht, E. T., Chan, V., Reinkensmeyer, D. J., and Bobrow, J. E. (2008). Optimizing compliant, model-based robotic assistance to promote neurorehabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 16, 286–297. doi: 10.1109/TNSRE.2008.918389

Xie, H., Li, G., Zhao, X., and Li, F. (2020). Prediction of limb joint angles based on multi-source signals by GS-GRNN for exoskeleton wearer. *Sensors* 20, 1–16. doi: 10.3390/s20041104

Yang, T., Gao, X., Gao, R., Dai, F., and Peng, J. (2019). A novel activity recognition system for alternative control strategies of a lower limb rehabilitation robot. *Appl. Sci.* 9:3986. doi: 10.3390/app9193986

Yao, T., Gao, F., Zhang, Q., and Ma, Y. (2021). Multi-feature gait recognition with DNN based on sEMG signals. *Math. Biosci. Eng.* 18, 3521–3542. doi: 10.3934/mbe.2021177

Yin, G., Zhang, X., Chen, D., Li, H., Chen, J., Chen, C., et al. (2020). Processing Surface EMG Signals for Exoskeleton Motion Control. *Front. Neurorobot.* 14:40. doi: 10.3389/fnbot.2020.00040

Young, A. J., and Ferris, D. P. (2017). State of the art and future directions for lower limb robotic exoskeletons. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 171–182. doi: 10.1109/TNSRE.2016.2521160

Zhai, X., Jelfs, B., Chan, R. H. M., and Tin, C. (2017). Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network. *Front. Neurosci.* 11:379. doi: 10.3389/fnins.2017.00379

Ziegier, J., Gattringer, H., and Mueller, A. (2018). "Classification of gait phases based on bilateral EMG data using support vector machines," in *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics* (Enschede: IEEE), 978–983. doi: 10.1109/BIOROB.2018.8487750

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 03 November 2021
doi: 10.3389/fnbot.2021.700823

Robust Torque Predictions From Electromyography Across Multiple Levels of Active Exoskeleton Assistance Despite Non-linear Reorganization of Locomotor Output

Jacob A. George[1,2]*, Andrew J. Gunnell[3], Dante Archangeli[3], Grace Hunt[3], Marshall Ishmael[3], K. Bo Foreman[4] and Tommaso Lenzi[3]

[1] NeuroRobotics Lab, Department of Electrical and Computer Engineering, College of Engineering, University of Utah, Salt Lake City, UT, United States, [2] NeuroRobotics Lab, Division of Physical Medicine and Rehabilitation, School of Medicine, University of Utah, Salt Lake City, UT, United States, [3] Bionic Engineering Lab, Department of Mechanical Engineering, College of Engineering, University of Utah, Salt Lake City, UT, United States, [4] Motion Analysis Facility, Department of Physical Therapy and Athletic Training, College of Health, University of Utah, Salt Lake City, UT, United States

Robotic exoskeletons can assist humans with walking by providing supplemental torque in proportion to the user's joint torque. Electromyographic (EMG) control algorithms can estimate a user's joint torque directly using real-time EMG recordings from the muscles that generate the torque. However, EMG signals change as a result of supplemental torque from an exoskeleton, resulting in unreliable estimates of the user's joint torque during active exoskeleton assistance. Here, we present an EMG control framework for robotic exoskeletons that provides consistent joint torque predictions across varying levels of assistance. Experiments with three healthy human participants showed that using diverse training data (from different levels of assistance) enables robust torque predictions, and that a convolutional neural network (CNN), but not a Kalman filter (KF), can capture the non-linear transformations in EMG due to exoskeleton assistance. With diverse training, the CNN could reliably predict joint torque from EMG during zero, low, medium, and high levels of exoskeleton assistance [root mean squared error (RMSE) below 0.096 N-m/kg]. In contrast, without diverse training, RMSE of the CNN ranged from 0.106 to 0.144 N-m/kg. RMSE of the KF ranged from 0.137 to 0.182 N-m/kg without diverse training, and did not improve with diverse training. When participant time is limited, training data should emphasize the highest levels of assistance first and utilize at least 35 full gait cycles for the CNN. The results presented here constitute an important step toward adaptive and robust human augmentation via robotic exoskeletons. This work also highlights the non-linear reorganization of locomotor output when using assistive exoskeletons; significant reductions in EMG activity were observed for the soleus and gastrocnemius, and a significant increase in EMG activity was observed for the erector spinae. Control algorithms that can accommodate spatiotemporal changes in muscle activity have broad implications for exoskeleton-based assistance and rehabilitation following neuromuscular injury.

Keywords: powered exoskeleton, hip orthosis, electromyography (EMG) control, adaptive control, wearable robotics, torque prediction, locomotor output

Edited by:
R. A. R. C. Gopura,
University of Moratuwa, Sri Lanka

Reviewed by:
Suncheol Kwon,
National Rehabilitation Center,
South Korea
Sanjaya Bandara,
Kyushu University, Japan

****Correspondence:***
Jacob A. George
jacob.george@utah.edu

Received: *26 April 2021*
Accepted: *11 October 2021*
Published: *03 November 2021*

Citation:
George JA, Gunnell AJ, Archangeli D, Hunt G, Ishmael M, Foreman KB and Lenzi T (2021) Robust Torque Predictions From Electromyography Across Multiple Levels of Active Exoskeleton Assistance Despite Non-linear Reorganization of Locomotor Output.
Front. Neurorobot. 15:700823.
doi: 10.3389/fnbot.2021.700823

INTRODUCTION

Robotic exoskeletons can assist humans with walking by providing supplemental torque at the joint level. Supplemental torque has been provided during hip flexion and/or extension to restore function after neuromuscular disability (Awad et al., 2017, 2020; Ishmael et al., 2019) or to increase metabolic efficiency for healthy individuals (Kim et al., 2019). In the case of healthy individuals, most assistive exoskeletons aim to provide assistive torque proportionate to the user's joint torque. Temporal alignment of the user's joint torque and the assistive torque provided by the exoskeleton is critical to the efficacy of the exoskeleton (Ding et al., 2018).

One approach to temporally align the user's joint torque and the assistive torque provided by the exoskeleton is to use electromyographic (EMG) control algorithms to estimate the user's joint torque directly using real-time EMG recordings from the muscles that generate the torque. Electromyographic activity precedes the resultant joint torque and kinematic motion, thereby allowing joint torque to be estimated before it is generated. Electromyographic control is traditionally established by collecting a dataset of synchronized EMG recordings and known joint torques, and then training an algorithm to predict torque from EMG under a supervised learning paradigm. During run-time operation, live EMG signals are used to predict joint torque in real-time.

However, a fundamental challenge with EMG-control strategies is that EMG activity changes as a result of active exoskeleton assistance. For example, active exoskeleton assistance that reduces metabolic cost is assumed to be reducing muscle effort (Ferris and Lewis, 2009; Kim et al., 2019), and a reduction in muscle effort is detectable with EMG recordings (Gordon et al., 2013). These changes in muscle effort are often non-linear and are observed in muscles which are not associated directly with the joint being assisted; for example, active exoskeleton assistance at the hip joint has been shown to reduce muscle effort at the ankle joint (Lenzi et al., 2013). In essence, the act of controlling the exoskeleton (i.e., assisting the user) changes the relationship of the control signal (EMG) to the desired torque. As a result, traditional EMG control algorithms provide unreliable estimates of the user's joint torque during active exoskeleton assistance.

Before EMG-controlled exoskeletons can be used in a real-world setting with a variety of locomotion modes, EMG control algorithms need to be robust to the non-linear changes in EMG that occur as a result of active exoskeleton assistance. Working toward that goal, here we sought to first identify an EMG-control framework that would provide reliable predictions of joint torque across a variety of different levels of active exoskeleton assistance during treadmill ambulation. We approached this challenge from a data-driven perspective in order to identify the data-collection practices and algorithms that result in the most accurate and robust torque predictions. Data were collected from three participants walking at consistent speeds with varying levels of exoskeleton assistance, and we assessed the error associated with torque predictions from linear and non-linear EMG-control algorithms across a variety of assistance levels.

Our results confirm non-linear reorganization of locomotor output due to active exoskeleton assistance and present a novel EMG-control framework for robotic exoskeletons that provides consistent joint torque estimates across varying levels of exoskeleton assistance. These results constitute an important step toward adaptive exoskeletons capable of assisting individuals reliably across varying levels of exoskeleton assistance.

MATERIALS AND METHODS

Human Participants

Three healthy human participants—P1, P2, and P3—participated in this study. All participants were male, under 30 years old, and had prior training using a hip exoskeleton device. All experiments were carried out with informed consent from the participants and following protocols approved by the University of Utah Institutional Review Board.

Hip Exoskeleton

This study used a powered bilateral hip exoskeleton to assist walking (**Figure 1**). The exoskeleton design has been reported previously (Ishmael et al., 2019) and is summarized here briefly. Hip flexion and extension assistance in the sagittal plane were provided by an actuated joint. The actuator was connected to the participant through passive joints located distal and proximal to the powered joint. The passive joints were used to aid with initial alignment with the user's leg and allow voluntary frontal-plane motions during ambulation. The exoskeleton was attached to a flexible pelvis orthosis fabricated from polyurethane using 3D printing.

Custom control electronics were housed with a six-cell lithium-ion battery in a 3D-printed case on the lower back of the user. Sensing and motor power cables connected each actuator module of the bilateral exoskeleton to the control board. Onboard microcontrollers and an embedded computer performed the middle and high-level control computations to calculate a desired assistive joint torque (Ishmael et al., 2019). The desired torques were sent to onboard motor drivers, which performed low-level current control and output motor commutation signals to the actuation system. The assistive torque profile was generated from the summation of two Gaussian profiles—one for extension and another for flexion. An experimenter tuned the peak flexion and extension assistance timing by communicating with the custom control electronics over Wi-Fi.

Experimental Conditions

Participants walked on a split-belt (each belt was 20-in wide) Bertec Fully Instrumented Treadmill (Bertec, Columbus, OH, USA) at 1.16 m/s. A fixed speed was used in order to force users to achieve similar kinematic profiles despite varying levels of exoskeleton assistance. A speed of 1.16 m/s was selected as approximately halfway between medium and slow walking speeds reported in Bovi et al. (2011). Participants walked for 1 min under each of the five following conditions: (1) *Baseline*—the participant walking normally without wearing the exoskeleton; (2) *No Assistance (Passive)*—the participant walking

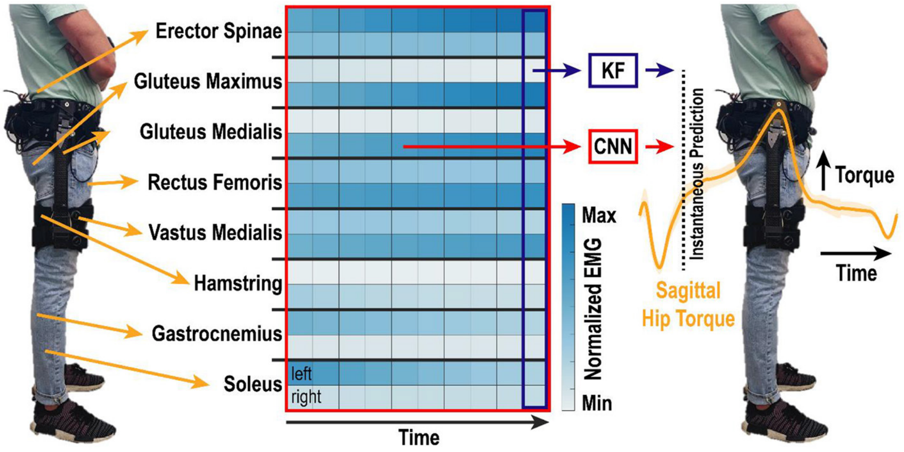


FIGURE 1 | Algorithm overview. EMG was recorded bilaterally from eight lower-limb muscle groups and motion capture was used to determine lower-limb kinematics and kinetics. EMG activity from the one most recent or ten most recent samples in time were used by the KF and CNN, respectively, to predict instantaneous torque of the right and left hips in the sagittal direction. Samples were analyzed at 200 Hz, such that the total time window shown is 50 ms. Heatmap shows example spatiotemporal EMG activity during walking. Differences across muscle groups (rows of two), between left and right legs (top and bottom rows, respectively), and over time (columns) create a rich set of features that are non-linearly correlated to hip torque.

while wearing the exoskeleton in a powered state in which it provided no assistance but actively minimized resistance; (3) *Low Assistance*—the exoskeleton provided 6 N-m of assistive flexion and 2.5 N-m of assistive extension during walking; (4) *Medium Assistance*—the exoskeleton provided 9 N-m of assistive flexion and 3.75 N-m of assistive extension during walking; and (5) *High Assistance*—the exoskeleton provided 12 N-m of assistive flexion and 5 N-m of assistive extension during walking. The assistive torque was provided as absolute torque and was not normalized based on the individual participants mass or body-segment lengths. The order of the five conditions was randomized for each participant to minimize the effect of fatigue toward the latter trials due to extended periods of walking. In the days prior to the experiment, the participants practiced using the exoskeleton with low, medium, and high assistive torque.

Motion Capture and Biomechanics

Each participant wore tight-fitting clothing with reflective markers representing a modified Plug-in-Gait Model with extra redundant markers placed on the pelvis for better tracking. Additional markers were placed on the hip exoskeleton: four on the exoskeleton's frame and four on the exoskeleton's passive degrees of freedom. The participant's pelvis and the hip exoskeleton's belt harness were assumed to move together as a rigid body. The hip exoskeleton's belt harness was kept on during all trial conditions (including the baseline, no-exoskeleton condition) in order for the reflective marker placement to remain the same throughout all trials. The participant's upper body was secured to a harness which was connected to an overhead support system.

Three calibration routines were performed with each participant. First, a static calibration was performed in which the participant was asked to stand still for 5 s with their feet shoulder width apart and arms bent out from their body. The static calibration was used to scale the Vicon Nexus and Visual 3D models for each subject. Second, a functional calibration was performed in which the participant walked at 1.16 m/s on the treadmill for approximately 5 s. The functional calibration was used to improve automatic marker labeling in Vicon Nexus. Third, a joint center calibration was performed in which the participant was asked to swing their legs in a clock pattern and perform two squats to bend their knees. The joint center calibration was used to locate the subject's knee joint centers using Symmetrical Axis of Rotation Analysis (SARA). Following the calibrations, each participant walked under the five experimental conditions described above. Marker trajectories and ground reaction force data were synchronized, recorded, and pre-processed using Vicon Nexus 2 software. The marker trajectory data was collected at 200 Hz and the ground reaction force data at 1,000 Hz. Heel-strike gait events were identified when each ground reaction force exceeded a threshold of 30 N. After pre-processing was completed, the trajectories, analog data from the force plates and EMG signals, and gait events were imported into Visual 3D software. A low-pass Butterworth filter with a cut-off frequency of 6 Hz was applied to the marker trajectories and another low-pass Butterworth filter with a cut-off frequency of 15 Hz was applied for the analog force plate data. Because of inconsistencies between each subject's hip joint center using Symmetrical Center of Rotation Estimation (SCoRE), landmarks were created for each subject's hip joint center of rotations using distances between anterior superior iliac spine pelvis markers and iliac crest markers. Using the joint center of rotations, we computed the kinetics and kinematics of the ankle, knee, and hip for both legs. The kinetics and kinematics were calculated in Visual 3D and then exported as MATLAB files for each participant.

EMG Acquisition

Surface EMG was recorded from 16 channels using the 16-channel MA400 EMG Motion Lab System. The following muscle groups were targeted based on the locations recommended by the SENIAM project (SENIAM, 2021): soleus, gastrocnemius, hamstring, gluteus maximus, gluteus medius, vastus medialis, rectus femoris, and erector spinae. The skin was shaved and cleaned with rubbing alcohol prior to placing the surface electrodes. Electrodes were held in place using self-adherent wrap (Coban, 3M; Saint Paul, MN, USA). Electromyographic signals were sampled at 3,000 Hz and band-pass filtered with cutoff frequencies of 20 and 450 Hz in MATLAB. The rectified EMG signal was then low-pass filtered at 15 Hz in MATLAB.

Data Analysis

Kinetics, kinematics, and EMG data were synchronized for each participant at 200 Hz. Data were segmented into strides from heel strike to heel strike (defined as when the ground reaction force exceeded 30 N). The kinetics, kinematics, and EMG for each gait event were resampled to 1,000 samples for visual overlays. Algorithm training data consisted of continuous EMG data from 16 channels and the continuous combined hip torque of the human-exoskeleton system in the sagittal direction for the right and left legs.

EMG Analysis

For each electrode, the mean and peak EMG were calculated for each individual gait cycle for both the left and right legs. The mean and standard error of the mean were calculated for the mean and peak EMG across all of the gait cycles for each of the five experimental conditions. Electromyographic activity was then normalized relative to the baseline condition. This process was completed independently for each participant.

Algorithms and Training Conditions

Two algorithms were implemented in MATLAB to estimate hip torque from the 16 continuous EMG features: a standard Kalman filter (KF) (George et al., 2019) and a convolutional neural network (CNN) (George et al., 2018) (**Figure 1**). Both algorithms have been used extensively before for real-time myoelectric control of bionic devices (George et al., 2018, 2019, 2020; Brinton et al., 2020; Paskett et al., 2021), and are summarized here briefly.

The KF provides an efficient recursive algorithm to optimally estimate the posterior probability of hip torque when the likelihood model (i.e., the probability of the EMG activity given current hip torque) and prior models (i.e., the state model of how torques change over time) are linear and Gaussian. The inclusion of prior information about the system state enables an efficient recursive formulation of the decoding algorithm and effectively smooths noisy estimates in a mathematically principled way (Wu et al., 2006). In the implementation presented here, the KF predicts the instantaneous torque of the right and left hip in the sagittal plane based on the EMG activity at the current time-point (**Figure 1**).

In contrast, the CNN predicts the instantaneous torque of the right and left hip in the sagittal plane based on a spatiotemporal "image" of EMG activity over the last 10 samples in time (George et al., 2018) (**Figure 1**). The CNN utilizes convolution to learn complex spatiotemporal relations within EMG activity that correlate to torque output. A series of non-linear weights are used to map the spatiotemporal features extracted from the convolution to the instantaneous torque. The architecture of the CNN is defined in (George et al., 2018) and the specific instantiation for this manuscript was as follows: The CNN input was an 16 × 10 image consisting of 16 EMG features sampled at the current time and nine previous time points. The CNN architecture (**Figure 2**) consisted of a single convolutional layer, two fully-connected layers, ReLu activation between layers and a regression output. A 1 × 5 kernel was used for convolution, such that the convolution was only across time and not across the feature set. A total of 10 convolutional filters were used to produce a 16 × 6 × 10 output feature map. The output of the convolutional layer was then passed through a ReLu activation layer before being passed to the first fully-connected layer. The output of the first fully-connected layer was also passed through a ReLu activation layer before being passed to the second fully-connected layer. Both fully-connected layers consisted of 1,056 neurons. The output of the second fully-connected layer was then fed into a final fully-connected layer that produced a total of two regression outputs, for the right and left hip torque in the sagittal plane. The CNN was trained using a Stochastic Gradient Descent with Momentum solver with an initial learning rate of 0.1 and a piecewise learning rate drop factor of 50% every 10 epochs. Training continued until 20 epochs had passed with no increase in performance on the validation data, or after a maximum of 2,000 epochs had passed.

A total of six KFs and six CNNs were trained under six different training conditions: (1–5) using independent data from each of the five experimental conditions, and (6) using combined data from all five of the experimental conditions. The KFs were trained using 75% of the gait cycles for given condition. The CNNs were trained using 65% of the gait cycles and validated using 10% of the gait cycles for a given condition. Each algorithm was tested on the remaining 25% of the gait cycles (not used for training) for each of the six conditions. Performance was measured by the root mean squared error (RMSE) between the algorithm prediction and the true torque of the participant.

To determine the impact of the number of gait cycles trained on, performance was assessed for each algorithm, under each of the five experimental conditions, by iteratively increasing the number of gait cycles from two to the maximum number of gait cycles available from the training data. Additional gait cycles were added in the temporal sequence in which they occurred in the training data. Similarly, to determine the impact of the number of EMG channels using in the training data, performance was assessed for each algorithm, under each experimental condition, by iteratively increasing the number of EMG channels from 1 to 16. Additional EMG channels were added using a stepwise Gram-Schmidt channel-selection algorithm (Nieveen et al., 2017).

Statistical Analyses

A separate one-way ANOVA was performed for each participant and each algorithm to compare the performance of the different training conditions. If any significance was found, subsequent

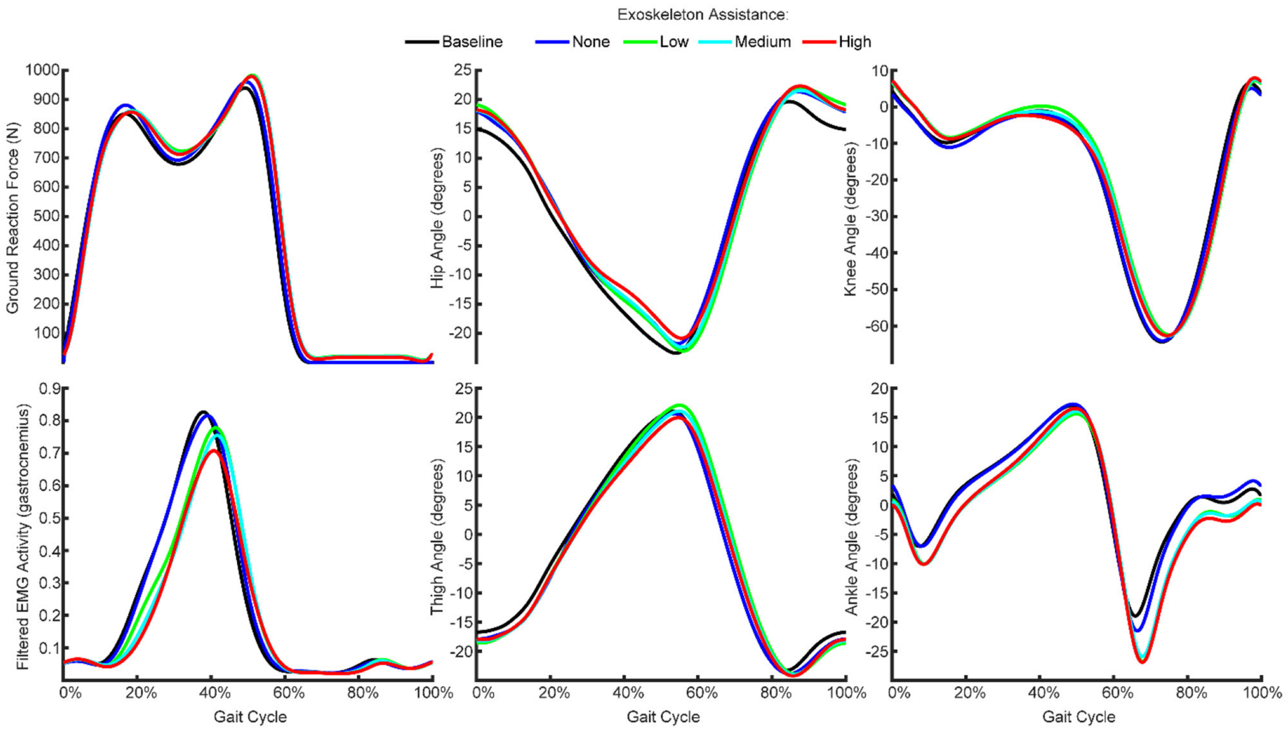


FIGURE 2 | Mean ground reaction force, kinematics and EMG features across all gait cycles from one participant recorded during the baseline condition and with exoskeleton assistance set to none, low, medium, or high. The participants walked at a fixed speed with varying levels of exoskeleton assistance, such that kinematics and gait were generally consistent while EMG activity generally decreased with increasing exoskeleton assistance.

pairwise comparisons (*t*-tests) were made using the Dunn-Sidák correction for multiple comparisons.

RESULTS

Exoskeleton Assistance Caused Non-linear, Participant-specific Changes in EMG With Minimal Changes in Kinematics

We recorded kinematics, kinetics, and EMG activity while the participants walked normally, while wearing a passive exoskeleton, and while wearing the exoskeleton with low, medium, and high assistance (**Figure 1**). We found that the kinematics were generally consistent across the conditions (**Figure 2**). That is, the thigh, hip, knee, and ankle angles throughout the gait cycle did not change meaningfully as a result of wearing the exoskeleton or increasing the exoskeleton assistance.

However, increasing exoskeleton assistance did lead to substantial non-linear changes in EMG activity. Most notably, active exoskeleton assistance resulted in a statistically significant decrease in mean soleus activity relative to the passive exoskeleton and baseline (normal walking) conditions for all three participants (**Figure 3**). Similarly, active exoskeleton assistance resulted in a statistically significant decrease in mean gastrocnemius activity relative to the passive exoskeleton and baseline (normal walking) conditions for two of the three participants. In contrast, active exoskeleton assistance resulted in a statistically significant increase in mean erector spinae activity relative to the passive exoskeleton and baseline (normal walking) conditions for all three participants. Changes in other muscles activity were less noticeable, and highly participant-specific (**Supplementary Figure 1**).

Increasing exoskeleton assistance from low to high did not systematically change EMG activity. Some trends were observed, but they were generally participant-specific and not statistically significant. For example, for P01 increasing exoskeleton assistance from low to high trended toward decreasing mean soleus activity, but the opposite trend was observed for P03, and virtually no difference was seen for P02.

Algorithm Performance Does Not Extrapolate to Conditions Not Explicitly Trained On

We implemented a KF and a CNN to predict hip torque based on EMG activity. We explored the impact of training data on run-time performance of the algorithm by testing all possible training conditions on all possible testing conditions. The participant hip torque (ground truth) was generally consistent within a given condition, and both algorithms were able to accurately recreate the torque when trained using data from that same condition (**Figure 4**). However, when extrapolating to levels of exoskeleton assistance that were not explicitly trained

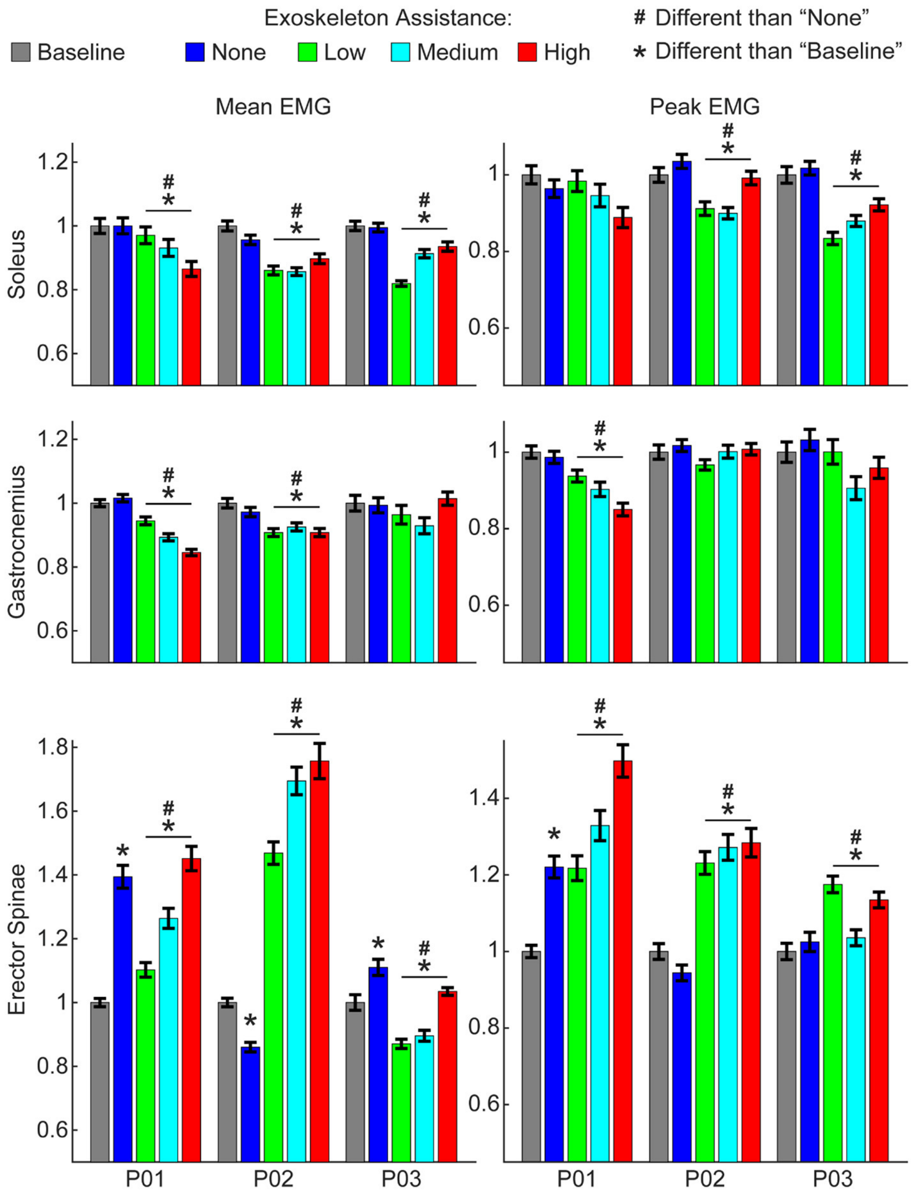


FIGURE 3 | Changes in EMG activity due to exoskeleton assistance. Changes in EMG activity were non-linear and unique to each muscle group and participant. Generally, wearing the exoskeleton in a passive state "none" condition did not alter EMG activity of lower-limb muscles, but did increase activity of the lower-back (erector spinae) muscles. Mean and peak EMG activity for lower-limb muscles were significantly less with active exoskeleton assistance relative to the baseline and relative to the passive exoskeleton. In contrast, mean and peak EMG activity for lower-back muscles were significantly greater with active exoskeleton assistance relative to the baseline and relative to the passive exoskeleton. Trends with increasing exoskeleton assistance were unique to each participant and muscle group. Additional muscle groups are shown in Supplementary Figure 1. Data show EMG activity averaged across each gait cycle for the right and left legs and normalized to the baseline condition.

on, the performance of both algorithms worsened (**Figure 5**). Predicting hip torque while the participant was walking with the exoskeleton in the passive configuration was generally the hardest condition to learn. The overall worst performance occurred when training on low exoskeleton assistance and then predicting hip torque while the participant was walking with the exoskeleton in the passive configuration. Training on the baseline condition or passive exoskeleton condition also generally resulted in poor performance when predicting hip torque while the participant was walking with active exoskeleton assistance.

Training Data Should Be Collected During Active Exoskeleton Assistance

From a practical perspective, torque only needs to be predicted during active exoskeleton assistance—that is, when the

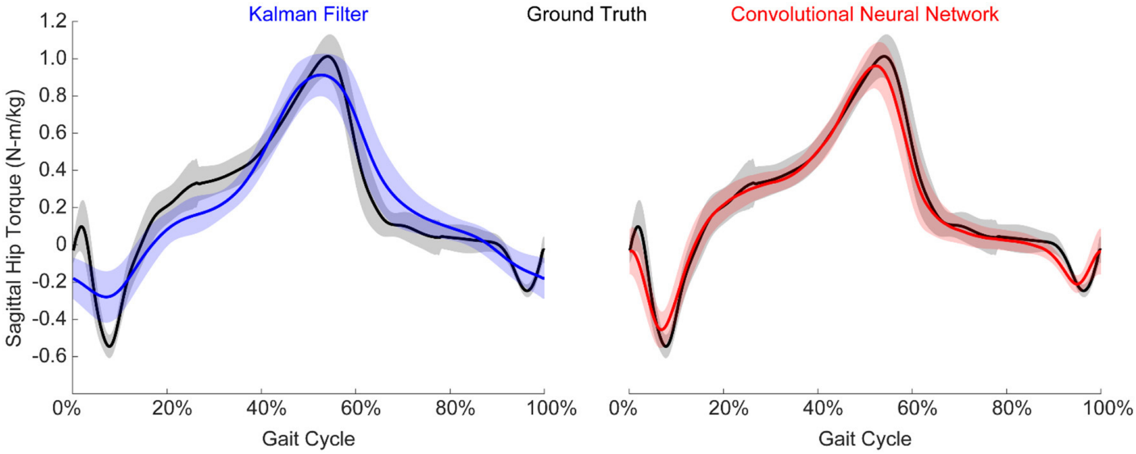
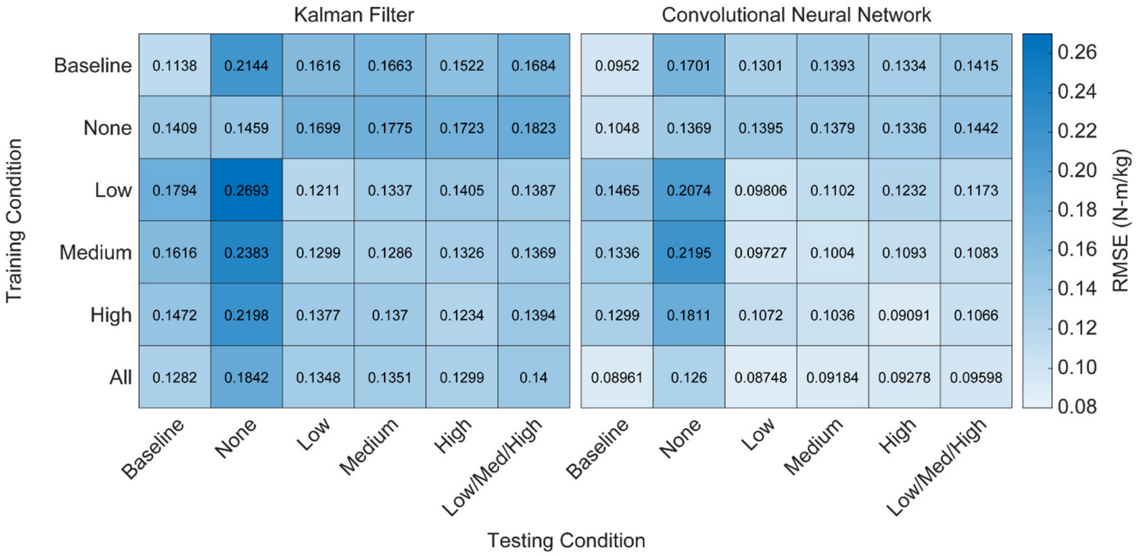


FIGURE 4 | Example normalized torque predictions (N-m/kg) of the hip joint in the sagittal plane. Both algorithms were able to reliably estimate torque, although the CNN generally outperformed the KF. Figure shows mean ± standard deviation.



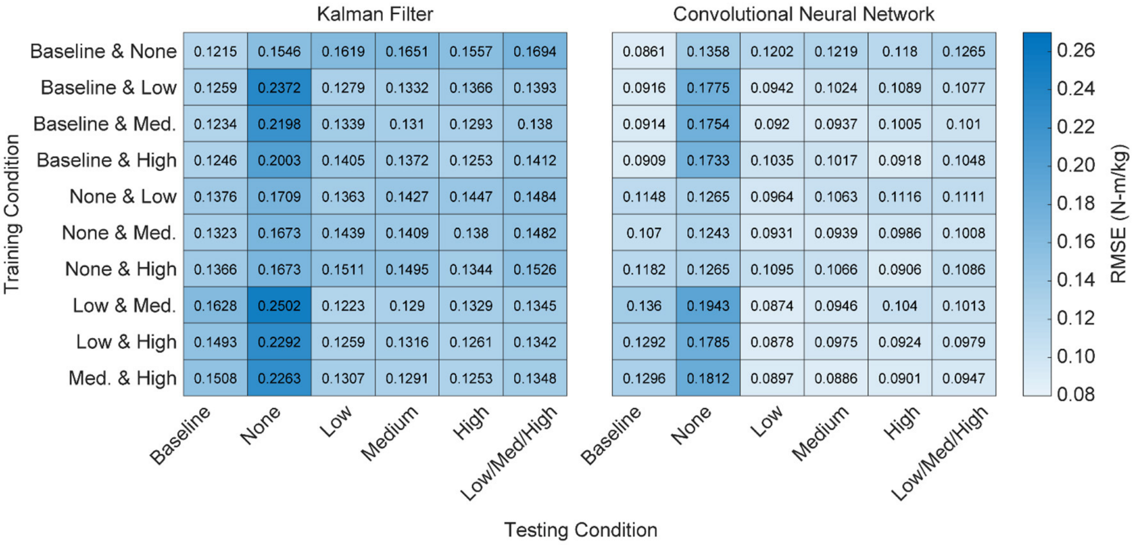
	Kalman Filter						Convolutional Neural Network					
	Baseline	None	Low	Medium	High	Low/Med/High	Baseline	None	Low	Medium	High	Low/Med/High
Baseline	0.1138	0.2144	0.1616	0.1663	0.1522	0.1684	0.0952	0.1701	0.1301	0.1393	0.1334	0.1415
None	0.1409	0.1459	0.1699	0.1775	0.1723	0.1823	0.1048	0.1369	0.1395	0.1379	0.1336	0.1442
Low	0.1794	0.2693	0.1211	0.1337	0.1405	0.1387	0.1465	0.2074	0.09806	0.1102	0.1232	0.1173
Medium	0.1616	0.2383	0.1299	0.1286	0.1326	0.1369	0.1336	0.2195	0.09727	0.1004	0.1093	0.1083
High	0.1472	0.2198	0.1377	0.137	0.1234	0.1394	0.1299	0.1811	0.1072	0.1036	0.09091	0.1066
All	0.1282	0.1842	0.1348	0.1351	0.1299	0.14	0.08961	0.126	0.08748	0.09184	0.09278	0.09598

FIGURE 5 | Heatmap of algorithm performance under various training and testing conditions. Performance is shown as RMSE (N-m/kg) of algorithm predictions relative to the ground truth values of right and left hip torque in the sagittal plane. Lighter colors represent lower RMSE and better performance. Training and testing on the same condition led to strong performance for both algorithms, as indicated by low RMSE values along the diagonals. Testing on conditions that were not explicitly trained on resulted in worse performance, as indicated by high RMSE values off the diagonals. Training on all of the conditions improved the overall performance of the CNN when testing on various levels of exoskeleton assistance, but did not improve the overall performance of the KF. That is, diverse training data increased the adaptability of the CNN, but not the KF.

exoskeleton is providing a low, medium, or high level of assistance. To this end, we assessed the overall performance of the training conditions by looking at the combined RMSE across the low, medium and high levels of assistance (**Figure 6**). For both algorithms, training on the baseline and passive conditions resulted in the worst practical performance (p's < 0.05). For the KF, there was no statistical differences when training on the low, medium, or high conditions. For the CNN, training on the medium and high conditions outperformed training on the low condition (p's < 0.05).

CNNs Trained on Diverse Training Data Perform Well Across All Levels of Assistance

Having demonstrated that algorithm performance does not extrapolate well to untrained levels of exoskeleton assistance, we assessed the performance of the algorithms when training data across all conditions is available—that is, the algorithm was trained using data from the baseline, passive/none, low, medium, and high conditions. We found



Kalman Filter

Training Condition	Baseline	None	Low	Medium	High	Low/Med/High
Baseline & None	0.1215	0.1546	0.1619	0.1651	0.1557	0.1694
Baseline & Low	0.1259	0.2372	0.1279	0.1332	0.1366	0.1393
Baseline & Med.	0.1234	0.2198	0.1339	0.131	0.1293	0.138
Baseline & High	0.1246	0.2003	0.1405	0.1372	0.1253	0.1412
None & Low	0.1376	0.1709	0.1363	0.1427	0.1447	0.1484
None & Med.	0.1323	0.1673	0.1439	0.1409	0.138	0.1482
None & High	0.1366	0.1673	0.1511	0.1495	0.1344	0.1526
Low & Med.	0.1628	0.2502	0.1223	0.129	0.1329	0.1345
Low & High	0.1493	0.2292	0.1259	0.1316	0.1261	0.1342
Med. & High	0.1508	0.2263	0.1307	0.1291	0.1253	0.1348

Convolutional Neural Network

Training Condition	Baseline	None	Low	Medium	High	Low/Med/High
Baseline & None	0.0861	0.1358	0.1202	0.1219	0.118	0.1265
Baseline & Low	0.0916	0.1775	0.0942	0.1024	0.1089	0.1077
Baseline & Med.	0.0914	0.1754	0.092	0.0937	0.1005	0.101
Baseline & High	0.0909	0.1733	0.1035	0.1017	0.0918	0.1048
None & Low	0.1148	0.1265	0.0964	0.1063	0.1116	0.1111
None & Med.	0.107	0.1243	0.0931	0.0939	0.0986	0.1008
None & High	0.1182	0.1265	0.1095	0.1066	0.0906	0.1086
Low & Med.	0.136	0.1943	0.0874	0.0946	0.104	0.1013
Low & High	0.1292	0.1785	0.0878	0.0975	0.0924	0.0979
Med. & High	0.1296	0.1812	0.0897	0.0886	0.0901	0.0947

FIGURE 6 | Heatmap of algorithm performance using various pairwise training approaches. Performance is shown as RMSE (N-m/kg) of algorithm predictions relative to the ground truth values of right and left hip torque in the sagittal plane. Lighter colors represent lower RMSE and better performance. Training with higher levels of exoskeleton assistance generally resulted in better performance. That is, when participant time is limited and only a subset of training conditions can be used, emphasis should be placed on higher levels of exoskeleton assistance.

that training on all of the conditions resulted in a statistically significant improvement for the CNN (p's < 0.05), but not for the KF. Furthermore, training on all of the conditions resulted in the best overall practical performance for the CNN. Interestingly, adding training data not explicitly relevant to the subtask did not degrade the performance of the CNN (**Figure 5**). For example, training on all of the conditions and testing only on the low conditions was not worse than just training on the low condition alone. Thus, training a CNN on diverse training data can yield a multipurpose algorithm capable of performing well at a variety of assistance levels.

When Participant Time Is Limited, Training Should Emphasis High Levels of Active Assistance

Training on all possible conditions is ideal, but may not be practical when participant time is limited. To address this question, we looked at the performance of the algorithms when training on just two conditions and then testing on all possible conditions (**Figure 7**). As noted earlier, we found that training on the baseline and passive conditions resulted in the worst practical performance for both algorithms (**Figure 6**; p's < 0.05). Similarly, for both the CNN and KF, the best performance was seen when training focused on active assistance (i.e., the low + medium, medium + high, or low + high conditions). For the CNN, training on the medium + high condition resulted in the best overall performance (p's < 0.05 compared to all other pairwise conditions except the low + high condition).

When Gait Cycles Are Very Limited, the KF Outperforms the CNN

When participant time is limited and diverse training data is encouraged, the amount of data needed for each condition becomes an important factor. To provide guidance on this, we quantified the performance of the algorithms as a function of the number of gait cycles trained on for each condition (**Figure 8**). For the KF, performance improved greatly within the first three gait cycles, and then leveled off after ∼10 cycles. For the CNN, performance only began improving after 20 cycles and became much more reliable after ∼35 cycles. After 35 cycles, performance leveled off for the baseline and passive conditions, but continued to improve for the active assistance conditions (up to the maximum number of gait cycles). In summary, with very limited data, we recommend a KF with at least three gait cycles, but ideally 10. When at least 35 gait cycles are available, we recommend a CNN—and encourage as much data as possible.

Additional EMG Channels, Including Bilateral Pairs, Improve Accuracy

The analyses thus far utilized eight bilateral pairs of EMG. However, working toward practical implementation, we sought to measure the impact of the number of EMG channels on algorithm performance. To this end, we used a stepwise Gram-Schmidt channel-selection algorithm (Nieveen et al., 2017) to assess algorithm performance when sequentially adding the next best channel one by one. We found that both the algorithms improved with additional channels, although the majority of the improvement comes with the first eight channels (**Figure 9**). Despite the fact that there were eight bilateral pairs, the first eight channels selected were rarely unilateral (**Figure 10**). Selection

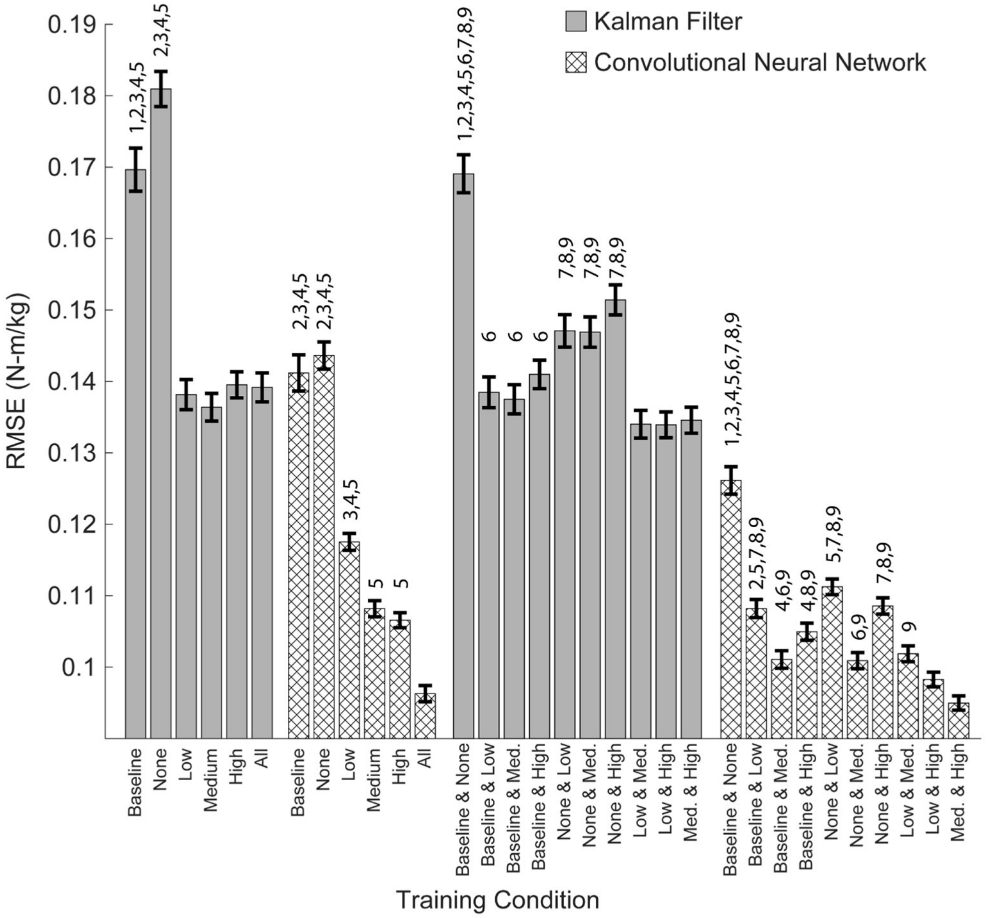


FIGURE 7 | Overall algorithm performance when predicting torque during low, medium and high levels of exoskeleton assistance. Performance is shown as RMSE (N-m/kg) of algorithm predictions relative to the ground truth values of right and left hip torque in the sagittal plane. Training on data from multiple levels of exoskeleton assistance improved the overall performance of the CNN, but not the KF. Bars show mean ± standard error of the mean. Numbers above the bars denote statistical significance for the algorithm within the subset of adjacent conditions (multiple pair-wise comparisons using the Dunn-Sidak correction for multiple comparison). Conditions within the subsets are numbered from 0 to 6, left to right (Baseline = 0, High = 5), or from 0 to 9, left to right (Baseline and None = 0, Med. and High = 9).

order of EMG channels was also highly participant-specific. For example, the right and left soleus muscles were consistently selected last for P01, while the right and left soleus were among the first five selected for P03.

DISCUSSION

This work serves as one of the first demonstrations of adaptive EMG control across varying levels of exoskeleton assistance. Consistent with prior results, we show a non-linear reorganization of locomotor output due to active exoskeleton assistance (Gordon et al., 2013; Lenzi et al., 2013; Sylos-Labini et al., 2014). We also show that, with appropriate training data, a CNN can capture these non-linear changes and accurately predict torque while a KF cannot. This finding is particularly relevant given the extensive use of KFs for predicting joint torque (Menegaldo, 2017; Teramae et al., 2018; Lyu et al., 2019).

Prior work has shown that neural network algorithms trained on mechanical sensor data are capable of generalizing predictions of hip torque across cyclic ambulation modes that were not explicitly trained on when using a fixed level of exoskeleton assistance (Molinaro et al., 2020). Here, we demonstrate that neural network algorithms trained on EMG recordings are *not* capable of generalizing predictions of hip torque across varying levels of exoskeleton assistance without explicit training on those levels of assistance. We saw a significant improvement in algorithm performance when training on additional levels of assistance, while (Molinaro et al., 2020) saw no significance difference when training on additional cyclic ambulation modes. Taken together, this suggests that data-driven machine-learning approaches to control lower-limb exoskeletons should emphasize training on various levels of assistance, but may not necessarily need to train on all possible cyclic ambulation modes (e.g., ascending/descending stairs, walking on level ground). We speculate the ability of control algorithms to generalize across

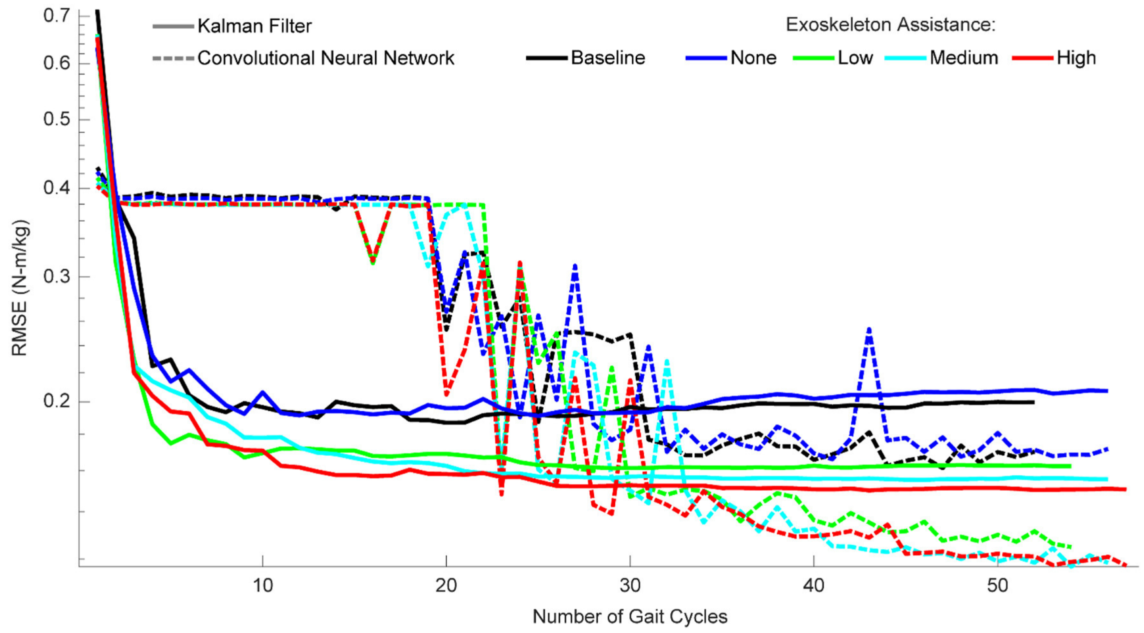


FIGURE 8 | Algorithm performance as a function of the number of gait cycles trained on. Performance is shown as RMSE (N-m/kg) of algorithm predictions relative to the ground truth values of right and left hip torque in the sagittal plane. For the KF, training data should consist of a minimum of three gait cycles, and little improvement is seen after ten gait cycles. For the CNN, at least 20 gait cycles are necessary for the algorithm to begin improving, and at least 35 gait cycles are needed to reliably outperform the KF. CNN performance continues to improve with additional gait cycles, but only if those gait cycles involve some level of exoskeleton assistance.

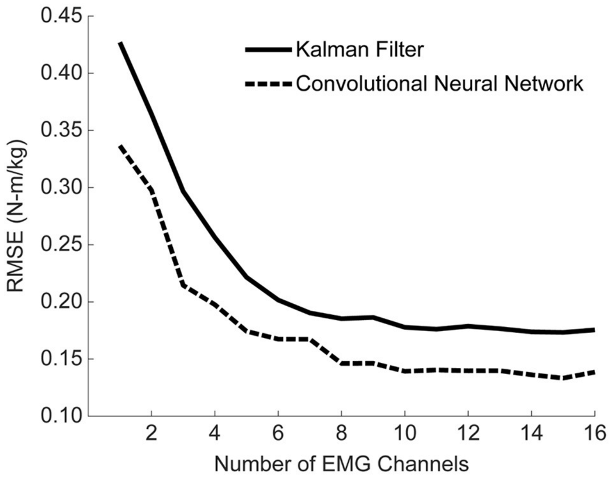


FIGURE 9 | Mean algorithm performance across all training conditions as a function of the number of EMG channels used. Performance is shown as RMSE (N-m/kg) of algorithm predictions relative to the ground truth values of right and left hip torque in the sagittal plane. Channels were selected using a Gram-Schmitt channel selection algorithm. Both the KF and CNN improve with additional channels, although the majority of the improvement comes with the first eight channels. Note that the 16 channels used here consist of eight muscles groups recorded bilaterally, although the eight channels selected were rarely unilateral.

cyclic ambulation modes is due to the fact that changes in ambulation mode result in relatively linear changes in locomotor output (e.g., a shift in the amplitude, density, or offset of the torque profile). In contrast, as demonstrated here, there is a highly non-linear reorganization of EMG activity that occurs with different levels of exoskeleton assistance. Future work should investigate whether or not EMG control algorithms can generalize to acyclic ambulation when trained on cyclic ambulation.

Due to the limited number of participants utilized in this study and prior studies (Molinaro et al., 2020), it is still unclear if neural network models can generalize across participants. That said, the results presented here provide a realistic path forward to participant-specific models of EMG control. When participant time is limited, training should emphasize training on the highest levels of assistance with at least eight EMG channels and at least 35 gait cycles. If <35 gait cycles are available, a KF should be utilized over a CNN. Although we did not use the algorithms presented here for real-time human-in-the-loop control, prior work has demonstrated that the KF and CNN used in this study are fast enough to be run in real-time with minimal computational resources and still provide stable control during complex activities of daily living (George et al., 2018, 2020; Brinton et al., 2020; Paskett et al., 2021). Future work should implement these algorithms for real-time human-in-the-loop EMG control of the hip exoskeleton presented here and address how improvements in torque prediction accuracy translate to real-time control and user comfort.

The error reported with the sagittal hip torque estimates presented here are favorable relative to prior work. Prior RMSEs of sagittal hip torque during level ground walking have

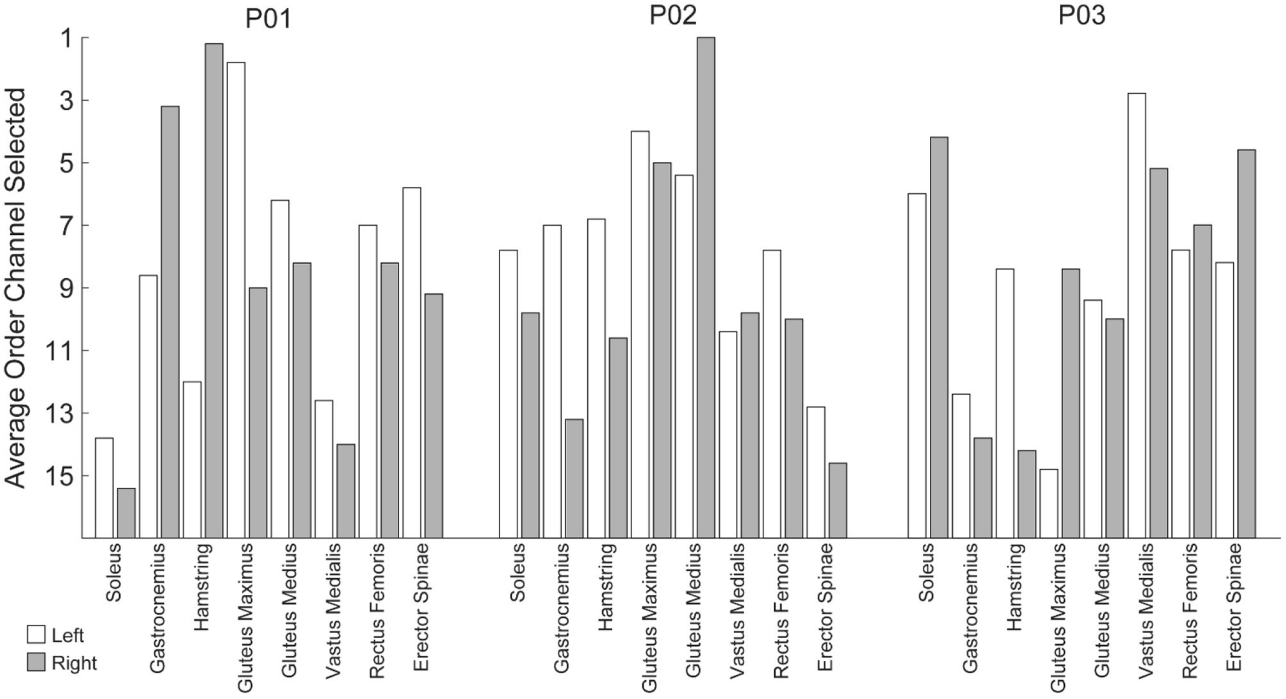


FIGURE 10 | Average order in which EMG channels were selected when using a step-wise Gram-Schmitt channel-selection algorithm. The order channels were selected was unique per participant and generally resulted in bilateral pairs from a subset of muscle groups as opposed to maximizing unilateral independent muscle groups.

been reported as 0.15 N-m/kg using pressure insoles (Forner-Cordero et al., 2006), 0.093 N-m/kg using mechanical sensors (Molinaro et al., 2020), and 0.20 N-m/kg using demographic, anthropometric, kinematic, and EMG data (Hahn and O'Keefe, 2011). In comparison, the best RMSE reported here was 0.095 N-m/kg using strictly EMG data.

Using only EMG data for exoskeleton control offers the unique ability to predict joint torque before any physical movements occur. Traditional approaches, like those mentioned above, have relied on onboard mechanical sensors (e.g., force sensors, inertial measurement units) to detect changes in gait cadence, slope, or task. Use of mechanical sensors is limited in that mechanical sensors can only detect a user's motor intent after their action has occurred, thereby making them prone to considerable mechanical delays (Ferris et al., 2007). In contrast, EMG precedes physical movement and therefore offers the future possibility of detecting novel actions before they happen (e.g., changing direction or starting/stopping walking).

Prior work has shown that users can adapt to simple linear proportional models for EMG control to maintain movement accuracy (Lenzi et al., 2012) and reduce overall energy expenditure (Gordon et al., 2013; Ao et al., 2017). The work presented here builds on these foundational studies by introducing robust data-driven machine-learning approaches that allow EMG control to provide more accurate estimates of torques from non-linear muscle groups. For example, simple linear proportional models work best when there is a primary muscle that is solely responsible for torque generation, like the bicep for the elbow (Lenzi et al., 2012) or the soleus for the ankle (Gordon et al., 2013; Ao et al., 2017). More advanced controllers are required to predict complex torque profiles (e.g., hip torque) from non-linear muscle activity.

Altogether, this work provides practical contributions toward robust EMG control of exoskeletons and has broad implications for the field of rehabilitation robotics. Electromyographic control algorithms that are robust to changes in EMG can ultimately be used to assist individuals when EMG changes as a result of neuromuscular impairment, or to guide rehabilitation when EMG changes as a result of neuromuscular recovery.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Utah Institutional Review Board. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

JG designed experiments, collected data, developed machine-learning algorithms, analyzed data, and wrote the manuscript. AG and GH collected data, developed biomechanical models, and processed motion-capture data. DA and MI collected data, designed, developed, and controlled the hip exoskeleton. KF oversaw data collection and motion capture analyses. TL oversaw all aspects of the work. All authors contributed to the revision of the manuscript.

FUNDING

Research reported in this publication was supported by the Department of Defense (DOD), Congressionally Directed Medical Research Programs (CDMRP), Peer Reviewed Orthopedic Research Program (PRORP), Applied Research Award (ARA), contract number W81XWH-16-1-0701 awarded to TL. Research reported in this publication was also supported by the Office of The Director (OD), Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), and National Institute of Dental and Craniofacial Research (NIDCR) of the National Institutes of Health (NIH) under Award Number DP5OD029571 awarded to JG. Additional trainee support for JG was provided by the University of Utah Center for Clinical and Translational Science (CCTS) Interdisciplinary Spheres of Translation Across the Research Spectrum (STARS) TL1 Training Program, with funding from the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002538 and TL1TR002540.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbot.2021.700823/full#supplementary-material

REFERENCES

Ao, D., Song, R., and Gao, J. (2017). Movement performance of human–robot cooperation control based on EMG-driven hill-type and proportional models for an ankle power-assist exoskeleton robot. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1125–1134. doi: 10.1109/TNSRE.2016.2583464

Awad, L. N., Bae, J., O'Donnell, K., Rossi, S. M. M. D., Hendron, K., Sloot, L. H., et al. (2017). A soft robotic exosuit improves walking in patients after stroke. *Sci. Transl. Med.* 9, eaai9084. doi: 10.1126/scitranslmed.aai9084

Awad, L. N., Kudzia, P., Revi, D. A., Ellis, T. D., and Walsh, C. J. (2020). Walking faster and farther with a soft robotic exosuit: implications for post-stroke gait assistance and rehabilitation. *IEEE Open J. Eng. Med. Biol.* 1, 108–115. doi: 10.1109/OJEMB.2020.2984429

Bovi, G., Rabuffetti, M., Mazzoleni, P., and Ferrarin, M. (2011). A multiple-task gait analysis approach: kinematic, kinetic and EMG reference data for healthy young and adult subjects. *Gait Posture* 33, 6–13. doi: 10.1016/j.gaitpost.2010.08.009

Brinton, M. R., Barcikowski, E., Davis, T., Paskett, M., George, J. A., and Clark, G. A. (2020). Portable take-home system enables proportional control and high-resolution data logging with a multi-degree-of-freedom bionic arm. *Front. Robot. AI* 7, 559034. doi: 10.3389/frobt.2020.559034

Ding, Y., Kim, M., Kuindersma, S., and Walsh, C. J. (2018). Human-in-the-loop optimization of hip assistance with a soft exosuit during walking. *Sci. Robot.* 3, eaar5438. doi: 10.1126/scirobotics.aar5438

Ferris, D. P., and Lewis, C. L. (2009). Robotic lower limb exoskeletons using proportional myoelectric control. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2009, 2119–2124. doi: 10.1109/IEMBS.2009.5333984

Ferris, D. P., Sawicki, G. S., and Daley, M. A. (2007). A physiologist's perspective on robotic exoskeletons for human locomotion. *Int. J. HR* 4, 507–528. doi: 10.1142/S0219843607001138

Forner-Cordero, A., Koopman, H. J. F. M., and van der Helm, F. C. T. (2006). Inverse dynamics calculations during gait with restricted ground reaction force information from pressure insoles. *Gait Posture* 23, 189–199. doi: 10.1016/j.gaitpost.2005.02.002

George, J. A., Brinton, M. R., Duncan, C. C., Hutchinson, D. T., and Clark, G. A. (2018). "Improved training paradigms and motor-decode algorithms: results from intact individuals and a recent transradial amputee with prior complex regional pain syndrome," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Honolulu, HI), 3782–3787. doi: 10.1109/EMBC.2018.8513342

George, J. A., Davis, T. S., Brinton, M. R., and Clark, G. A. (2019). Intuitive neuromyoelectric control of a dexterous bionic arm using a modified Kalman filter. *J. Neurosci. Methods* 330, 108462. doi: 10.1016/j.jneumeth.2019.108462

George, J. A., Radhakrishnan, S., Brinton, M., and Clark, G. A. (2020). "Inexpensive and portable system for dexterous high-density myoelectric control of multiarticulate prostheses," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Toronto, ON), 3441–3446. doi: 10.1109/SMC42975.2020.9283086

Gordon, K. E., Kinnaird, C. R., and Ferris, D. P. (2013). Locomotor adaptation to a soleus EMG-controlled antagonistic exoskeleton. *J. Neurophysiol.* 109, 1804–1814. doi: 10.1152/jn.01128.2011

Hahn, M. E., and O'Keefe, K. B. (2011). A neural network model for estimation of net joint moments during normal gait. *J. Musculoskel. Res.* 11, 117–126. doi: 10.1142/S0218957708002036

Ishmael, M. K., Tran, M., and Lenzi, T. (2019). "ExoProsthetics: assisting above-knee amputees with a lightweight powered hip exoskeleton," in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)* (Toronto, ON), 925–930. doi: 10.1109/ICORR.2019.8779412

Kim, J., Lee, G., Heimgartner, R., Revi, D. A., Karavas, N., Nathanson, D., et al. (2019). Reducing the metabolic rate of walking and running with a versatile, portable exosuit. *Science* 365, 668–672. doi: 10.1126/science.aav7536

Lenzi, T., Carrozza, M. C., and Agrawal, S. K. (2013). Powered hip exoskeletons can reduce the user's hip and ankle muscle activations during walking. *IEEE Trans. Neural Syst. Rehabil. Eng.* 21, 938–948. doi: 10.1109/TNSRE.2013.2248749

Lenzi, T., De Rossi, S. M. M., Vitiello, N., and Carrozza, M. C. (2012). Intention-based EMG control for powered exoskeletons. *IEEE Trans. Biomed. Eng.* 59, 2180–2190. doi: 10.1109/TBME.2012.2198821

Lyu, M., Chen, W.-H., Ding, X., Wang, J., Pei, Z., and Zhang, B. (2019). Development of an EMG-controlled knee exoskeleton to assist home rehabilitation in a game context. *Front. Neurorobot.* 13:67. doi: 10.3389/fnbot.2019.00067

Menegaldo, L. L. (2017). Real-time muscle state estimation from EMG signals during isometric contractions using Kalman filters. *Biol. Cybern.* 111, 335–346. doi: 10.1007/s00422-017-0724-z

Molinaro, D. D., Kang, I., Camargo, J., and Young, A. J. (2020). "Biological hip torque estimation using a robotic hip exoskeleton," in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, 791–796. doi: 10.1109/BioRob49111.2020.9224334

Nieveen, J. G., Zhang, Y., Wendelken, S., Davis, T. S., Kluger, D. T., George, J. A., et al. (2017). "Polynomial Kalman filter for myoelectric prosthetics using efficient kernel ridge regression," in 2017 8th International IEEE/EMBS Conference on Neural Engineering (NER) (Piscataway, NJ), 432–435.

Paskett, M. D., Brinton, M. R., Hansen, T. C., George, J. A., Davis, T. S., Duncan, C. C., et al. (2021). Activities of daily living with bionic arm improved by combination training and latching filter in prosthesis control comparison. *J. Neuroeng. Rehabil.* 18:45. doi: 10.1186/s12984-021-00839-x

SENIAM (2021). Available at: http://www.seniam.org/ (accessed March 18, 2021).

Sylos-Labini, F., La Scaleia, V., d'Avella, A., Pisotta, I., Tamburella, F., Scivoletto, G., et al. (2014). EMG patterns during assisted walking in the exoskeleton. *Front. Hum. Neurosci.* 8:423. doi: 10.3389/fnhum.2014.00423

Teramae, T., Noda, T., and Morimoto, J. (2018). EMG-based model predictive control for physical human–robot interaction: application for assist-as-needed control. *IEEE Robot. Automat. Lett.* 3, 210–217. doi: 10.1109/LRA.2017.2737478

Wu, W., Gao, Y., Bienenstock, E., Donoghue, J. P., and Black, M. J. (2006). Bayesian population decoding of motor cortical activity using a Kalman filter. *Neural Comput.* 18, 80–118. doi: 10.1162/089976606774841585

Author Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Department of Defense.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 05 November 2021
doi: 10.3389/fnbot.2021.706630

The Differences Between Motor Attempt and Motor Imagery in Brain-Computer Interface Accuracy and Event-Related Desynchronization of Patients With Hemiplegia

*Shugeng Chen[1]†, Xiaokang Shu[2]†, Hewei Wang[1], Li Ding[1], Jianghong Fu[1] and Jie Jia[1,3,4]**

[1] Department of Rehabilitation Medicine, Huashan Hospital, Fudan University, Shanghai, China, [2] School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China, [3] National Clinical Research Center for Aging and Medicine, Huashan Hospital, Fudan University, Shanghai, China, [4] National Center for Neurological Disorders, Shanghai, China

Background: Motor attempt and motor imagery (MI) are two common motor tasks used in brain-computer interface (BCI). They are widely researched for motor rehabilitation in patients with hemiplegia. The differences between the motor attempt (MA) and MI tasks of patients with hemiplegia can be used to promote BCI application. This study aimed to explore the accuracy of BCI and event-related desynchronization (ERD) between the two tasks.

Materials and Methods: We recruited 13 patients with stroke and 3 patients with traumatic brain injury, to perform MA and MI tasks in a self-control design. The BCI accuracies from the bilateral, ipsilesional, and contralesional hemispheres were analyzed and compared between different tasks. The cortical activation patterns were evaluated with ERD and laterality index (LI).

Results: The study showed that the BCI accuracies of MA were significantly ($p < 0.05$) higher than MI in the bilateral, ipsilesional, and contralesional hemispheres in the alpha-beta (8–30 Hz) frequency bands. There was no significant difference in ERD and LI between the MA and MI tasks in the 8–30 Hz frequency bands. However, in the MA task, there was a negative correlation between the ERD values in the channel CP1 and ipsilesional hemispheric BCI accuracies ($r = -0.552$, $p = 0.041$, $n = 14$) and a negative correlation between the ERD values in channel CP2 and bilateral hemispheric BCI accuracies ($r = -0.543$, $p = 0.045$, $n = 14$). While in the MI task, there were negative correlations between the ERD values in channel C4 and bilateral hemispheric BCI accuracies ($r = -0.582$, $p = 0.029$, $n = 14$) as well as the contralesional hemispheric BCI accuracies ($r = -0.657$, $p = 0.011$, $n = 14$). As for motor dysfunction, there was a significant positive correlation between the ipsilesional BCI accuracies and FMA scores of the hand part in 8–13 Hz ($r = 0.565$, $p = 0.035$, $n = 14$) in the MA task and a significant

OPEN ACCESS

Edited by:
Dingguo Zhang,
University of Bath, United Kingdom

Reviewed by:
Ke Li,
Shandong University, China
Sébastien Rimbert,
Inria Nancy – Grand-Est Research Centre, France

Correspondence:
Jie Jia
shannonjj@126.com

†These authors have contributed equally to this work

Received: *07 May 2021*
Accepted: *07 October 2021*
Published: *05 November 2021*

Citation:
Chen S, Shu X, Wang H, Ding L, Fu J and Jia J (2021) The Differences Between Motor Attempt and Motor Imagery in Brain-Computer Interface Accuracy and Event-Related Desynchronization of Patients With Hemiplegia.
Front. Neurorobot. 15:706630.
doi: 10.3389/fnbot.2021.706630

positive correlation between the ipsilesional BCI accuracies and FMA scores of the hand part in 13–30 Hz ($r = 0.558$, $p = 0.038$, $n = 14$) in the MI task.

Conclusion: The MA task may achieve better BCI accuracy but have similar cortical activations with the MI task. Cortical activation (ERD) may influence the BCI accuracy, which should be carefully considered in the BCI motor rehabilitation of patients with hemiplegia.

Keywords: BCI accuracies, event-related desynchronization, motor attempt, motor imagery, brain-computer interface

INTRODUCTION

Motor attempt and motor imagery (MI) are two common experimental paradigms in the non-invasive electroencephalogram (EEG)-based brain-computer interface (BCI) system design. Motor imagery is a cognitive rehearsal of physical movements that is defined as the internal reactivation of any first-person motor performance without an overt motor output (Jeannerod, 1995, 2001). There is extensive use of MI for athletes. Action observation combined with MI has been shown to engage the motor system in sports (Di Rienzo et al., 2019). Video observation and MI have been used to improve jumping performance in national rhythmic gymnastics athletes (Battaglia et al., 2014). Motor attempt is defined as attempting to move a paralyzed hand with little or no covert movement, specifically for patients with motor disability (Antelis et al., 2017). A meta-analysis by Bai et al. (2020) suggested that using movement attempts as the trigger task in BCI training appeared to be more effective than using MI. A study (Hotz-Boendermaker et al., 2008) of neural activity using functional MRI (fMRI) in paraplegics showed that during the attempt to move, the primary motor cortex is slightly less engaged than during the imagination of movement, however, the regions of the parietal lobe and cerebellum, well known to be involved in sensorimotor integration, are more activated during the attempt to move. For patients paralyzed in the upper limbs after a stroke, attempted movement is more easily detected in EEG than motor imagination (Muralidharan et al., 2011). A study by Blokland et al. (2015) showed the differences between attempted movement and actual movement using a neuromuscular blocker.

Both motor attempt (MA) and MI tasks can induce cortical activations, which can be applied in BCI decoding. Recent research into MA- and MI-BCI has so far yielded positive results. A randomized controlled trial (RCT) reported that MA-BCI could improve hand function in chronic stroke patients after 4 weeks of training (Ramos-Murguialday et al., 2013). Rathee et al. (2019) tried MA-related EEG-driven hand-exoskeleton on post-stroke patients and found improvement in their Action Research Arm Test (ARAT). Pichiorri et al. (2015) found that 1 month of MI-BCI intervention achieved greater power spectra in the alpha and beta bands in the ipsilesional hemisphere and improved motor function in subacute stroke patients with severe motor deficits. Although MA- and MI-BCI have been both widely but, respectively, researched (Bundy et al., 2017), the overall analysis and research on the BCI accuracy of MA tasks have not been done in patients with hemiplegia, especially compared further with MI tasks. Mizuno et al. (2018) proposed a protocol to compare MA-BCI and MI-controlled treatment to explore the efficacy of MA-BCI in stroke patients but no results have been reported yet. Blokland et al. (2014) used EEG and functional near-infrared spectroscopy (fNIRS) to test the feasibility of using MA instead of MI as a task for brain switch control. Which paradigm to choose from MA and MI tasks in BCI testing and training is still uncertain and is an important question to answer.

Brain-computer interface accuracy is an important parameter in BCI-based intervention. Higher BCI accuracies have been correlated with larger excitability in healthy people (Niazi et al., 2012) and better motor recovery in patients with hemiplegia (Biasiucci et al., 2018). Patients with hemiplegia usually presented different cortical excitability from healthy people (Wong et al., 2013; Agius Anastasi et al., 2017; Li et al., 2019), and their cortical activation patterns changed a lot due to cerebral injury (Shu et al., 2019). As a result, the choices of EEG channels have a great influence on BCI decoding effects. Research on BCI accuracies and control varied among the bilateral, ipsilesional, and contralesional hemispheres in patients with hemiplegia. Lopez-Larraz et al. (2017) analyzed EEG from the whole-brain channels while Ramos-Murguialday et al. and Ono et al. collected EEG signals directly from the ipsilesional hemisphere. Their subjects were asked to perform an MA task with their paralyzed hand (Ramos-Murguialday et al., 2013; Ono et al., 2015). Interestingly, Antelis et al. (2017) and Bundy et al. (2017) both achieved reasonable BCI accuracies by decoding EEG signals from the contralesional hemisphere. The BCI accuracies were different between MA and MI tasks in spinal cord injury (SCI) patients. Lopez-Larraz et al. (2012) reported a higher accuracy of MA than MI in SCI patients. Blokland et al. (2014) reported a significantly higher average accuracy for MA than MI in patients with tetraplegia. Although several studies explored the differences in BCI accuracy in SCI patients, it is still unclear how could it be different concerning MA and MI tasks in patients with hemiplegia.

Additionally, event-related de/synchronization (ERD/ERS) are common indexes extracted from EEG during MA and MI tasks (Pfurtscheller et al., 1999; Müller-Putz et al., 2007). A higher magnitude of ERD activity is related to larger cortical activation during motor tasks (Pfurtscheller et al., 1999; Takemi et al., 2013; Kaiser et al., 2014). It is considered that MI is close to attempting movement by the fact that it is linked to kinesthetic motor imagery (KMI) and kinesthetic feeling (Nikulin et al., 2008). Moreover, the ERD following the KMI after learning is very similar to those generated during motor execution or MA (Rimbert et al., 2019a). It was further reported that BCI accuracy was highly associated with mu-band ERD (Kaplan et al., 2016). Cortical activations vary between different motor tasks. Kraeutner et al. (2014) reported that the strength of ERD was significantly greater in motor execution than in MI in non-disabled participants. Higher motor impairment was reported to be related to stronger ERD in the unaffected hemisphere in MI tasks while it was related to the higher hemispheric asymmetry of ERS in motor execution tasks in stroke patients (Kaiser et al., 2012). However, the MI-induced cortical activity change was significantly augmented and even exceeding that of motor execution tasks in controlling a computer cursor (Miller et al., 2010). In addition, primary motor cortices have a symmetrical organization between the right and left hemispheres, particularly in hand motor control (Cicinelli et al., 1997; Tecchio et al., 1997; Del Gratta et al., 2000). In fMRI-based neuroscience research, the laterality index (LI) (Caria et al., 2011) was used to measure the inter-hemispheric balance in cortical activations (Pivik et al., 1993). Ramos-Murguialday et al. (2013) used LI in an MA-BCI study with fMRI to show the different activations between hemispheres. Johnson et al. (2018) combined repetitive transcranial magnetic stimulation (rTMS) and MI-BCI in stroke and found significant alterations in the interhemispheric inhibition and increased relative ipsilesional cortical activation. However, the inter-hemispheric balance has not yet been compared directly between MA and MI tasks. The differences in ERD/ERS, as well as LI between MA and MI tasks in patients with hemiplegia, need to be further explored. The findings in the current study may provide references on how to choose different BCI experimental paradigms (MA or MI tasks) in BCI training.

Given the lack of studies investigating BCI accuracy and the EEG features between MA and MI tasks in patients with hemiplegia, we aim to explore the cortical difference between MA and MI tasks. We will calculate the BCI accuracy across the hemispheres and ERD values. We hypothesize that there is both difference and relationship in BCI accuracy and ERD of patients with hemiplegia.

MATERIALS AND METHODS

Subject Recruitment

Sixteen patients were recruited from the Department of Rehabilitation Medicine of Huashan Hospital. The EEG data of two patients were contaminated with large artifacts and were discarded. The remaining 14 of the 16 patients (age: 45.7 ± 15.1 years) were enrolled in the further analysis. All the patients met the following inclusion criteria: (1) first-time unilateral stroke who are >2 weeks post-stroke and confirmed by scan or diagnosed with a unilateral traumatic brain injury and in the rehabilitation stage; (2) aged from 25 to 70 years; (3) right-handed; (4) mini-mental state examination (MMSE) ≥ 25; and (5) was able to sit independently in a chair for at least 1 h. The exclusion criteria included: (1) had unilateral neglect or vision problem; (2) receiving non-invasive brain stimulation during the study; (3) allergic to electrode gel; and (4) had previous experience with or knowledge of MA and MI tasks. Participant demographics and clinical characteristics are shown in **Table 1**. This study was approved by the Ethical Committee of Huashan Hospital. Informed consent was signed according to the Declaration of Helsinki.

Experimental Design and Electroencephalogram Data Collection

The patients were asked to sit on a chair/wheelchair in front of a screen in a comfortable posture (**Figure 1A**). An EEG cap (actiCAP, Brain Products, Germany) consisting of 32 channels of Ag/AgCl electrodes was used for EEG recording. The electrodes were distributed according to the 10–20 international system (Klem et al., 1999). The reference channel was placed on the right mastoid process and the ground channel was placed on the forehead. The impedance was kept below 5 kΩ. The signals were amplified with BrainAmp (Brain Products, Gilching, Germany) and recorded at a sampling rate of 200 Hz. The raw EEG signals were filtered with a bandpass filter of [1, 100] Hz.

The patients were asked to perform two sessions of MA or MI tasks of wrist extension. One session included 15 right-hand trials and 15-left hand trials and the trial types (right or left) appeared randomly. In one session, there was only one type of task. The MA and MI tasks were done randomly on different days for the same patient to avoid possible temporal effects from the order of the two tasks. In the MI task, the patients imagined extending the wrists without overt motor output; in the MA task, the patients tried to perform the wrist extension. For both motor tasks, there was a pre-training session before the formal testing. In the pre-training session, the patients were required to perform the MA or MI tasks with both affected and unaffected hands simultaneously. During the MA tasks, visible movements could be observed from the unaffected hand. During the MI tasks, no covert movements were observed from neither affected nor unaffected hands. When they performed the MA tasks with the affected hands, not all wrist extension movements were visible, but they were told to try their best.

Figures 1B,C show the timelines of a single trial during MA/MI tasks. Prior to EEG recording, all the patients practiced the MA or MI tasks and became familiar with the cues on the screen. At the beginning of each trial, a white "+" appeared at the center of the screen to remind the patients to prepare for the task. After 3 s, a red rectangle appeared to inform the patients to perform the motor tasks according to the left or right cue. The red rectangle then disappeared after 1 s and the patients began to perform the MA or MI tasks for 5 s until the white "+" disappeared. During the 5 s, they attempted to perform sustained wrist extension (Cassim et al., 2000). After that, there was a resting period to reduce the chance of the adaptation of the patients. The resting period was randomized to last between 2 and 3 s. The recording session was kept relatively short to minimize discomfort and to ensure the patients were focused on the tasks.

TABLE 1 | Demographics and clinical characteristics of the patients.

Patients	Age (yrs)	AH	TI	SI	Injury location	TSI (m)	MMSE	FMA-UL (max = 66)	FMA-hand (max = 24)
P1	66–70	R	I	S	Basal ganglia	84	29	12	1
P2	61–65	L	I	C+S	Basal ganglia, corona radiata, frontal cortex	4	30	10	0
P3	25–30	R	TBI	C	Parietal cortex	132	30	53	15
P4	31–35	L	I	S	Basal ganglia	16	30	24	2
P5	56–60	R	H	S	Basal ganglia	11	28	4	0
P6	41–45	R	H	S	Basal ganglia	1	30	50	15
P7	61–65	L	I	C+S	Basal ganglia, thalamus, paracele, frontoparietal cortex	8	30	12	1
P8	46–50	R	H	S	Basal ganglia	12	25	18	1
P9	46–50	R	I	C	Frontal, parietal, temporal cortex	1	30	37	13
P10	25–30	L	TBI	C	Subdural	21	30	24	1
P11	31–35	R	TBI	C	Parietal cortex	180	30	34	2
P12	36–40	R	I	S	Insular lobe	5	30	29	1
P13	61–65	R	I	S	Brainstem	3	30	13	1
P14	25–30	R	H	S	Frontoparietal cortex	28	30	49	12

AH, affected hand; TI, type of injury; SI, site of injury; TSI, time since injury; R, right; L, left; S, Subcortical; C+S, Cortical and subcortical; C, cortical; I, ischemia; H, hemorrhage; TBI, traumatic brain injury; yrs, years; m, months.

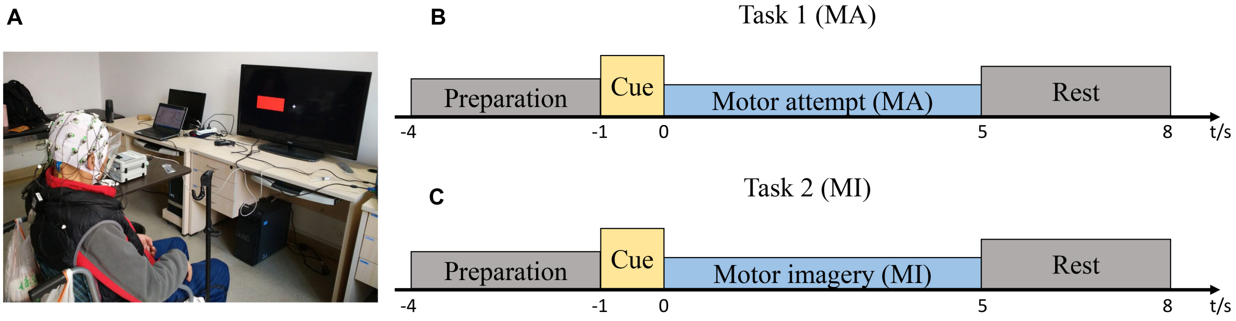


FIGURE 1 | Study setup and experimental protocol. **(A)** The patient was seated in a chair in front of a screen. **(B)** Timeline of a single trial during the motor attempt (MA) task. **(C)** Timeline of a single trial during the motor imagery (MI) task.

The patients were required to look at a stationary fixation point at the center of the screen to minimize eye movement artifacts. They were also instructed to avoid excessive eye blinking, swallowing, or any irrelevant movement.

Data Pre-processing

The left hemisphere was covered with FP1, FZ, F3, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, CP1, PZ, P3, and P7 (15 channels) while the right hemisphere was covered with O2, P4, P8, TP10, CP6, CP2, CZ, C4, T8, FT10, FC6, FC2, F4, F8, and FP2 (15 channels). For the tasks with affected hands (vs. rest), the BCI accuracies of the bilateral hemispheres were calculated with 31 channels except for the reference channel (the 32 channel). The BCI accuracies of the right or left hemisphere were calculated with 15 channels, respectively. The average ERD/ERS in the bilateral, ipsilesional, and contralesional hemispheres were also calculated based on the same number of channels.

The preprocessed EEG data consisted of high-pass filtering at 1 Hz and low-pass filtering at 30 Hz. Then, the datasets were subjected to an independent component analysis (ICA) decomposition by using EEGLAB (Delorme and Makeig, 2004). The ICA components representing eyeblink, head movement, and power line interference were removed from the data. Manual checking was performed in the EEG data of all 31 channels and all trials. No bad channel and bad trial rejection were performed in the data.

Brain-Computer Interface Accuracy Calculation

The offline BCI accuracies were evaluated by the single-trial decoding accuracy between the task and idle states. In every single trial, the task state was defined at [1, 4] s, and the idle state was defined at [−4, −1] s. The EEG features were extracted with the common spatial pattern (CSP) algorithm

(Benjamin et al., 2008). The log-variance of the first and last three components produced by the CSP filters were selected as feature vectors. They were subsequently classified using linear discriminative analysis (LDA). The two classes (MA vs. rest and MI vs. rest) of the affected hand were classified in the offline analysis. The EEG features were extracted from the alpha-beta frequency bands (8–30 Hz).

In the offline analysis, 10-fold cross-validation was conducted with the dataset for each experimental condition. All 31 channels of the EEG signals were used for pattern classification. The 30 trials of the task states and 30 trials of the idle states of the affected hand were randomly divided into 10 sets. Each set was tested with the classifier which was calibrated using the other nine sets. This analysis was repeated 10 times, generating 100 decoding accuracies. The EEG features were also extracted from the alpha-beta frequency bands using the CSP filters. The BCI accuracies of all 14 patients were evaluated with the average classification accuracy and SD. The detailed calculation formulas of the BCI accuracy can be referred to in the published paper (Yao et al., 2013).

Event-Related Desynchronization/Event-Related Synchronization Values Analysis

For each channel, we computed the power spectrum at the alpha-beta frequency bands (8–30 Hz) to identify the ERD/ERS on the motor tasks of the affected hand. The time-frequency distributions (TFDs) of the EEG trials were estimated using a windowed Fourier transform (WFT) (Peng et al., 2019) with a fixed 200-ms Hanning window. The WFT yielded, for each trial, a complex time-frequency estimate F(t,f) at each time-frequency point (t,f), extending from -3,000 to 5,000 ms (in steps of 5 ms) in the time domain, and from 1 to 30 Hz (in steps of 1 Hz) in the frequency domain. The power spectrum (P), $P(\mathrm{t,f}) = |\,\mathrm{F(t,f)}|^2$, was obtained. The percentage of the relative power decrease was calculated to obtain the ERD/ERS with respect to a resting-state baseline ([−3, −1] s). The interest time was set at [1, 4] s, during which the patient was performing the MA or MI tasks. For that, the ERD/ERS in the alpha-beta (8–30 Hz) frequency bands were averaged in the time interval [1, 4] s. The formula of the ERD/ERS (Pfurtscheller et al., 1999) is:

$$\mathrm{ERD/ERS} = \frac{P_{interest} - P_{baseline}}{P_{baseline}} * 100\%$$

By using this definition, ERD was expressed as a negative value and stronger ERD is related to higher cortical activations during the motor tasks (MA or MI) (Pfurtscheller et al., 1999). Laplace transformation was applied when calculating the correlations between the ERD and BCI accuracies. The cortical positions of the patients with injury in the right hemispheres were flipped for calculating the ERD values, simulating that all the patients have an injury in the left hemispheres. The topographies were drawn with an interest time of 1 to 4 s, with respect to a resting-state baseline ([−3, −1] s). The time-frequency maps were drawn with the above-mentioned calculation, representing the signal magnitude as a joint function of time and frequency at each time-frequency point.

The LI, approaching a value of 1 or −1 when the brain activity was either purely ipsilesional or contralesional (Caria et al., 2011), was calculated from the ERD values in both the ipsilesional and contralesional hemispheres during the interest time when the patient was performing the motor tasks. The formula of LI is:

$$\mathrm{LI} = \frac{ERD_{ipsilesional} - ERD_{contralesional}}{|ERD_{ipsilesional}| + |ERD_{contralesional}|}$$

Data Analysis

The statistical analysis was performed with SPSS version 19.0 (SPSS Inc., Chicago, IL, United States) and the figures were drawn with GraphPad Prism 7 Software (GraphPad Software, Inc., San Diego, CA, United States). Two-way repeated-measures ANOVA, taking both task (two levels: MA and MI tasks) and hemisphere (three levels: bilateral, ipsilesional, and contralesional hemispheres) as the within-subject factors, were performed on the BCI accuracies and ERD values. A paired t-test was applied as a *post hoc* analysis and was used to compare the LI values between the MA and MI tasks. Spearman correlation was used between the BCI accuracies and ERD/ERS, between the BCI accuracies and FMA scores, and the ERD/ERS between the tasks. The statistical significance was set at $p < 0.05$. Bonferroni correction was applied in multiple comparisons.

RESULTS

Comparison of Brain-Computer Interface Accuracies Between Motor Attempt and Motor Imagery

The comparison of BCI accuracies of 14 patients in the 8–30 Hz band is shown in **Figure 2A**. The main effect analysis from the two-way repeated-measures ANOVA on the BCI accuracies showed that the tasks had a significant effect on BCI accuracies ($F_{1,13} = 13.293$, $p = 0.003$) while there was no significant effect for the hemispheres on BCI accuracies ($F_{2,26} = 1.49$, $p = 0.244$). There was no significant hemisphere × task interaction ($F_{2,78} = 2.441$, $p = 0.107$). The estimated marginal means showed an average BCI accuracy of 79% (72.3–85.7% in 95% CI) in the MA task and an average BCI accuracy of 66.5% (60.1–72.8% in 95% CI) in the MI task. The BCI accuracy in the MA task was 12.6% (5.1–20% in 95% CI) higher than that in the MI task. **Table 2** shows the BCI accuracy and variance of accuracy for each patient in the MA and MI tasks, respectively.

EVALUATION OF CORTICAL ACTIVATIONS BETWEEN MOTOR ATTEMPT AND MOTOR IMAGERY

The main effect analysis from the two-way repeated-measures ANOVA on the ERD/ERS of the 14 patients showed that the tasks had no significant effect on ERD/ERS and no significant

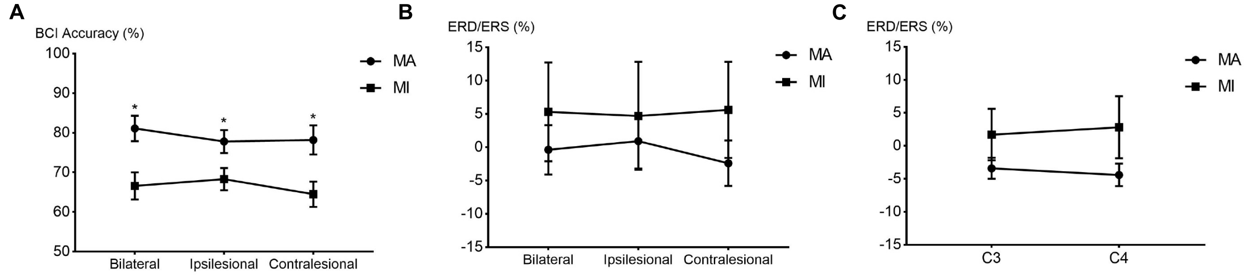


FIGURE 2 | The line chart and scatter plot of brain-computer interface (BCI) accuracies and event-related de/synchronization (ERD/ERS) values in the 8–30 Hz band. **(A)** The average BCI accuracies (mean ± SEM) between the MA and MI tasks in the bilateral, ipsilesional, and contralesional hemispheres. *$p < 0.0167$ after Bonferroni correction. **(B)** The average ERD/ERS values (mean ± SEM) between the MA and MI tasks in the bilateral, ipsilesional, and contralesional hemispheres. **(C)** The average ERD/ERS values (mean ± SEM) between the MA and MI tasks in the C3/C4 channels. SEM, standard error of mean.

TABLE 2 | The BCI accuracy and variance of accuracy for each patient in the MA and MI tasks, respectively.

MA task		**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**
Bilateral H	mean	0.753	0.910	0.885	0.807	0.917	0.891	0.667	0.676	0.651	0.957	0.645	0.930	0.943	0.723
	SD	0.077	0.096	0.058	0.089	0.080	0.065	0.108	0.131	0.087	0.049	0.147	0.067	0.052	0.124
Ipsilesional H	mean	0.709	0.927	0.743	0.793	0.810	0.782	0.693	0.631	0.698	0.943	0.725	0.907	0.910	0.617
	SD	0.111	0.065	0.078	0.067	0.092	0.100	0.129	0.103	0.126	0.058	0.130	0.060	0.086	0.132
Contralesional H	mean	0.820	0.860	0.820	0.758	0.843	0.831	0.597	0.704	0.642	0.957	0.505	0.943	0.957	0.707
	SD	0.076	0.124	0.079	0.096	0.108	0.119	0.142	0.100	0.100	0.042	0.151	0.062	0.049	0.108
MI task		**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**
Bilateral H	mean	0.613	0.850	0.530	0.543	0.567	0.743	0.593	0.645	0.640	0.600	0.560	0.773	0.883	0.837
	SD	0.127	0.087	0.122	0.116	0.102	0.115	0.119	0.108	0.110	0.142	0.128	0.117	0.080	0.098
Ipsilesional H	mean	0.600	0.883	0.570	0.627	0.647	0.870	0.647	0.643	0.645	0.573	0.578	0.777	0.783	0.717
	SD	0.105	0.083	0.129	0.154	0.097	0.072	0.126	0.120	0.122	0.121	0.101	0.089	0.108	0.113
Contralesional H	mean	0.597	0.780	0.533	0.610	0.530	0.717	0.560	0.695	0.585	0.617	0.502	0.703	0.810	0.860
	SD	0.159	0.093	0.102	0.124	0.102	0.113	0.119	0.077	0.086	0.113	0.118	0.099	0.128	0.086

H, hemisphere; SD, standard deviation.

effect for the hemispheres on ERD/ERS. There was also no significant hemisphere × task interaction. There was no significant difference between the MA and MI tasks in the average ERD/ERS with the bilateral or unilateral hemispheres. The main effect analysis from the two-way repeated-measures ANOVA on the C3/C4 ERD/ERS of the 14 patients showed no significant main effect. There was also no significant hemisphere × task interaction. **Figures 2B,C** show the line chart and scatter plot of the ERD/ERS values between the MA and MI tasks in the 8–30 Hz band.

Figure 3 shows the average topographies of all 14 patients in the alpha-beta frequency bands between the MA and MI tasks. In the 8–13 Hz band, the patients presented strong activations (ERD, with color blue) in both tasks while the MA task was stronger. In 8–30 and 13–30 Hz bands, both ERD and ERS existed.

Figure 4 shows the ERD pattern changes over time for one patient with left hemisphere injury during the motor tasks. The ERD was presented in the red rectangular box in the 8–30 Hz frequency bands in channels C3 and C4 of the MA task (**Figure 4A**) and 8–13 Hz frequency bands in channels C3 and C4 of the MI task (**Figure 4B**).

Table 3 shows the channels with the average strongest ERD in the MA and MI tasks of the 14 patients. Most of the electrodes were in or around the sensorimotor areas and they presented stronger ERD than other electrodes. Besides, there was a positive correlation between the MA and MI tasks in the C4 ERD values in the 8–13 Hz ($r = 0.534$, $p = 0.049$, $n = 14$).

Comparison of Hemispheric Balance Between Motor Attempt and Motor Imagery

Figure 5 shows the LI values of all 14 patients in the MA and MI tasks in the 8–30 Hz band. Nine out of 14 patients presented the same positive/negative sign in the LI values (P3, P6, and P13 showed the same LI value between tasks) while five patients (P7, P8, P9, P12, P14) showed different positive and negative values. Eight out of 14 patients (57%) in MA and 7 out of 14 (50%) in MI showed a negative value in LI when they were performing motor tasks of wrist extension of the affected hands. There was no significant difference between the MA and MI tasks after the paired t-test.

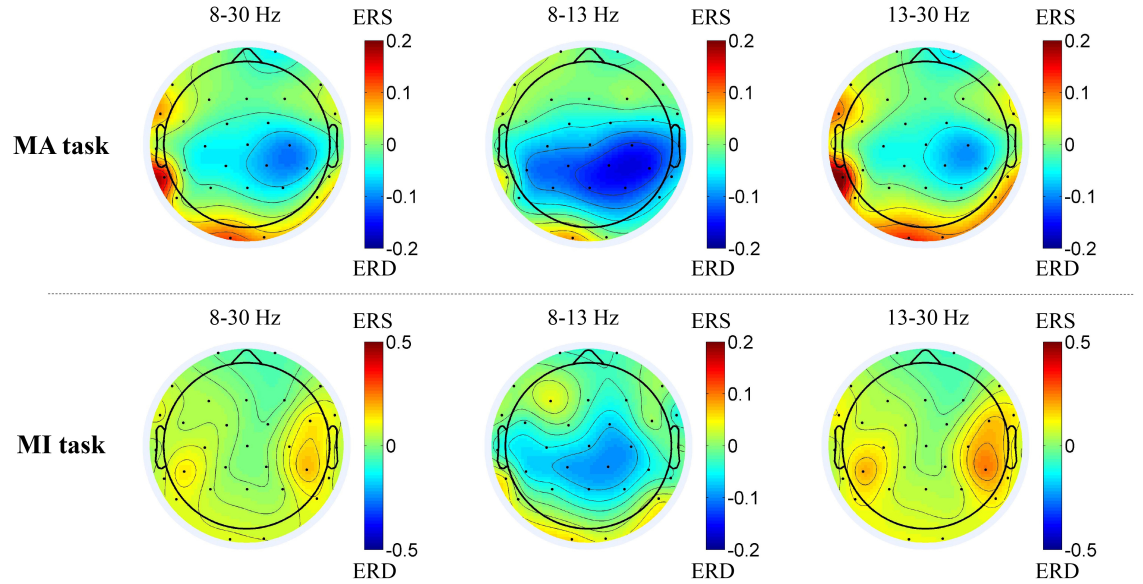


FIGURE 3 | Average topographies of all 14 patients in the 8–30 Hz, 8–13 Hz and 13–30 Hz frequency bands between the MA and MI tasks. The first row shows the average topographies of the 8–13 Hz and 13–30 Hz bands in the MA tasks. The second shows the average topographies of the 8–30 Hz, 8–13 Hz and 13–30 Hz bands in the MI tasks.

Correlations Between Brain-Computer Interface Accuracies and FMA Scores and Event-Related Desynchronization/Event-Related Synchronization Values

Figure 6 shows the correlations between the BCI accuracies and ERD/ERS values in the 8–30 Hz band. In the MA task, there was a negative correlation between the ERD values in channel CP1 and the ipsilesional hemispheric BCI accuracies ($r = -0.552$, $p = 0.041$, $n = 14$) and a negative correlation between the ERD values in channel CP2 and the bilateral hemispheric BCI accuracies ($r = -0.543$, $p = 0.045$, $n = 14$). While in the MI task, there were negative correlations between the ERD values in channel C4 and the bilateral hemispheric BCI accuracies ($r = -0.582$, $p = 0.029$, $n = 14$) as well as the contralesional hemispheric BCI accuracies ($r = -0.657$, $p = 0.011$, $n = 14$).

For all 14 patients, there was a significant positive correlation between ipsilesional BCI accuracies and the FMA scores of the hand part in the 8–13 Hz ($r = 0.565$, $p = 0.035$, $n = 14$) in the MA task and a significant positive correlation between ipsilesional BCI accuracies and the FMA scores of the hand part in the 13–30 Hz ($r = 0.558$, $p = 0.038$, $n = 14$) in the MI task.

DISCUSSION

In this study, we performed two motor tasks (MA and MI) on 16 right-handed patients with hemiplegia with BCI-based experimental paradigms. We compared every subject *via* a self-control design to eliminate the potential effect of handedness. To explore the differences between the MA and MI tasks, the BCI accuracies and ERD/ERS, as well as the LI, were compared between the two tasks. It demonstrated significantly higher BCI accuracies in the MA task. Additionally, similar strength in ERD and no significant difference in the LI between the two tasks were found. The correlations between the BCI accuracies and ERD, as well as the FMA scores, were also observed.

Difference in Brain-Computer Interface Accuracies of Motor Attempt and Motor Imagery Tasks

As it is known to all BCI researchers, MI and MA (execution) are two important experimental paradigms for motor tasks in the BCI system design. Both MI and MA have been explored in healthy subjects, and MI has also been explored in stroke or patients with hemiplegia. Thus, further exploring the BCI tasks of MI and MA is valuable in improving the clinical application of the BCI system. There were some differences between MA and MI tasks. Physically, MA was similar to motor execution and was easily accepted by stroke patients. Mentally, it was reported that MI required the active inhibition of motor neural activation, and the brain patterns during MI were less distinguishable from rest than motor execution patterns (Wolpaw et al., 2000). The patients felt that it was less natural and more difficult to perform MI. In practice, we found that the stroke patients tended to be more focused in the MA than the MI tasks and less likely to fall asleep during the motor task. Although the motor attempt was probably to induce spasticity during movement, long-term BCI studies based on the MA task as a paradigm have reported no significant increase in the spasticity of stroke patients (Biasiucci et al., 2018; Ramos-Murguialday et al., 2019). To optimize BCI

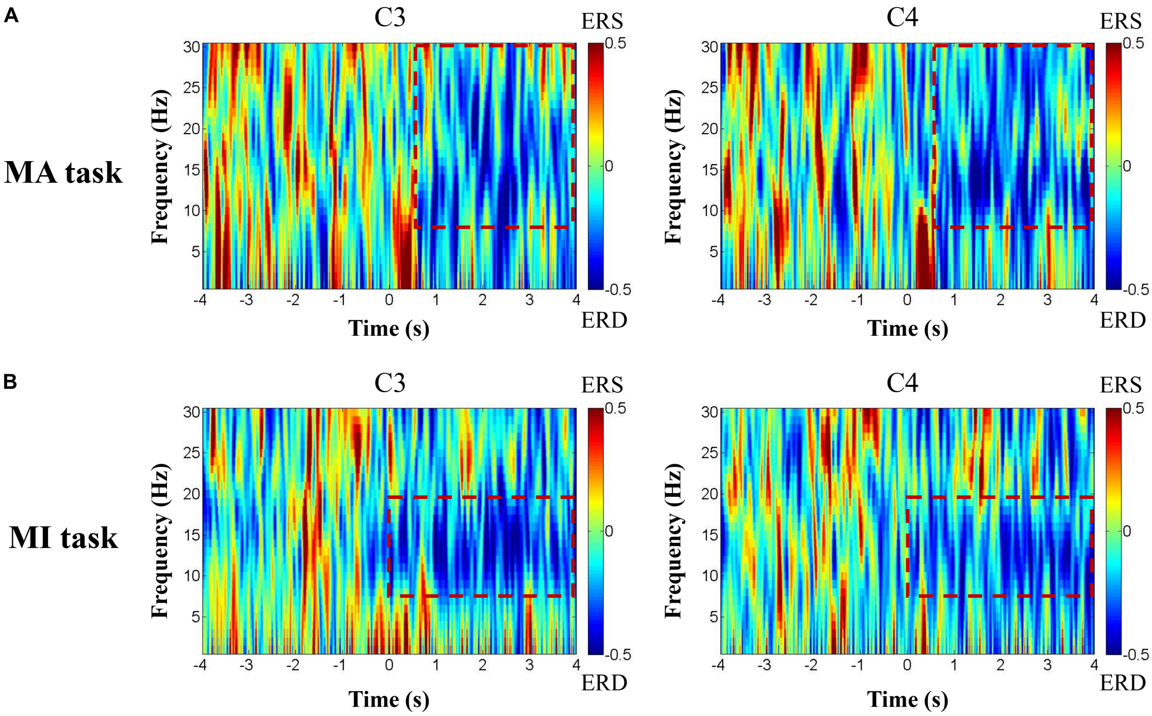


FIGURE 4 | The ERD pattern changes over time of one patient with left hemisphere injury during the motor tasks. The ERD was presented in the red rectangular box in 8–30 Hz frequency bands in channels C3 and C4 of the MA task **(A)** and 8–13 Hz frequency bands in channels C3 and C4 of the MI task **(B)**.

TABLE 3 | Channels with average strongest ERD in the MA and MI tasks of the 14 patients.

		MA					MI		
8–30 Hz	C4	CP2	CP6	P4	8–30 Hz	FP2	FP1	F4	FC2
	−10.2%	−9.3%	−8.0%	−6.4%		−5.7%	−3.1%	−2.5%	−1.9%
8–13 Hz	CP2	CP6	P4	C4	8–13 Hz	CP2	CP1	C4	Cz
	−15.6%	−15.3%	−14.2%	−13.7%		−9.8%	−8.8%	−8.3%	−7.1%
13–30 Hz	C4	CP2	CP6	FP2	13–30 Hz	FP2	FP1	F4	Fz
	−9.2%	−7.5%	−5.9%	−5.8%		−6.4%	−3.1%	−2.2%	−0.5%

application in patients with hemiplegia, we compared their differences in BCI accuracies.

The results in **Figure 2** were in line with previous research (Lopez-Larraz et al., 2012; Blokland et al., 2014). Blokland et al. and Eduardo et al. both reported a significantly higher average BCI accuracy for MA tasks than MI tasks. However, their results were based on SCI patients, who had no cerebral injury. The average BCI accuracy found by Blokland et al. (2014) was 79% for MA and 70% for MI tasks. The results of the current study were similar, which was 79% for MA and 66.5% for MI tasks. The BCI accuracies of the MA task were significantly higher than those of the MI task in patients with hemiplegia. Whereas there was a difference in the BCI accuracy of the MI task in our study and Blokland's. One explanation was that our participants were all patients with hemiplegia, whose cortical activation patterns could be different from patients with spinal cord injury. Theoretically, the BCI accuracies of SCI patients could be higher than those of patients with hemiplegia because the cortical status was relatively and functionally intact for SCI patients. The results in our study showed confidence in the BCI application for the MA task in patients with hemiplegia. Further study needs to be performed to distinguish the cortical variation between cerebral injury and SCI patients.

In recent years, the choices of EEG channels varied in BCI-related research of motor rehabilitation. The EEG signals from the bilateral (Lopez-Larraz et al., 2017), ipsilesional (Ramos-Murguialday et al., 2013; Ono et al., 2015), and contralesional (Antelis et al., 2017; Bundy et al., 2017) hemispheres were all reported to successfully control the BCI system. The different choices of EEG channels might obtain different BCI decoding effects but there was no conclusion for the best application. Our results showed no significant difference between the bilateral and unilateral hemispheres in the MA and MI tasks, but the BCI accuracies of the bilateral hemispheres were higher than those of the unilateral hemisphere in the MA task. This was consistent with Spüler et al. (2018), who explored the accuracies of the

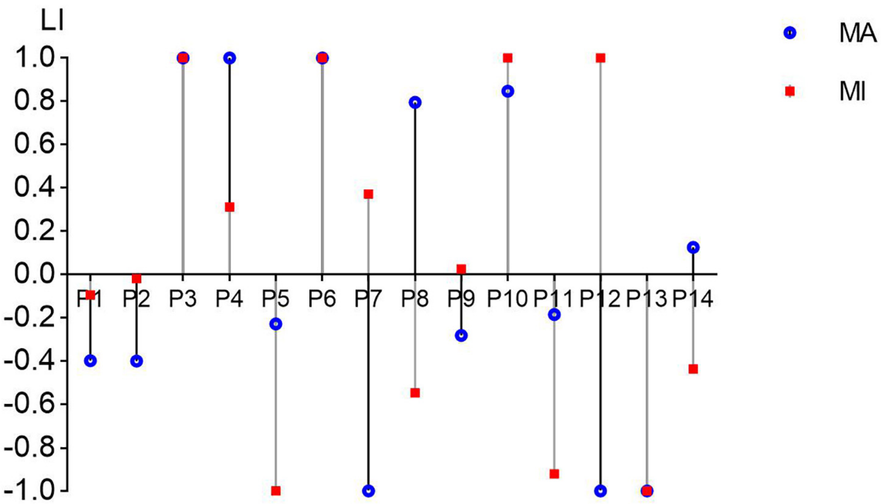


FIGURE 5 | The laterality index (LI) values from C3/C4 channels in all 14 patients in the MA and MI tasks in the 8–30 Hz band. The value range of LI can be from –1 (entirely ipsilesional) to 1 (entirely contralesional). The blue round dots present the LI values of the MA task and the red square dots present the LI values of the MI task.

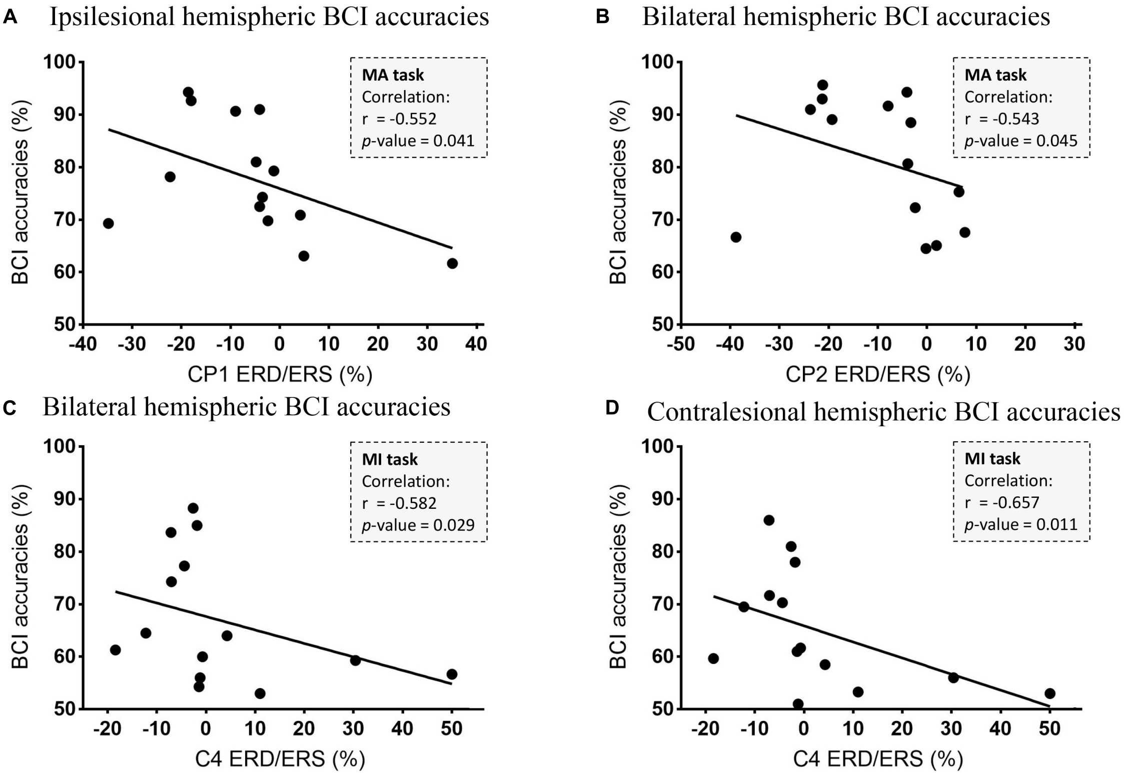


FIGURE 6 | Correlations between the BCI accuracies and ERD/ERS values in the 8–30 Hz band. **(A)** The correlation between CP1 ERD and the ipsilesional hemispheric BCI accuracies ($r = -0.552$, $p = 0.041$, $n = 14$). **(B)** The correlation between CP2 ERD and the bilateral hemispheric BCI accuracies ($r = -0.543$, $p = 0.045$, $n = 14$). **(C)** The correlation between C4 ERD and the bilateral hemispheric BCI accuracies ($r = -0.582$, $p = 0.029$, $n = 14$). **(D)** The correlation between C4 ERD and the contralesional hemispheric BCI accuracies ($r = -0.657$, $p = 0.011$, $n = 14$).

bilateral, ipsilesional, and contralesional hemispheres and found that using bi-hemispheric activity led to the best accuracies in severely paralyzed stroke patients. However, the variations in the injury location and time since the cerebral injury of our recruited patients made it difficult to find a significant difference in the BCI accuracies between the hemispheres.

Difference in Cortical Activations of Motor Attempt and Motor Imagery Tasks

Understanding the cortical differences between MI and MA is of benefit for exploring the brain function plasticity change through BCI training since the sensorimotor rhythm (SMR)-based BCI training is based on these BCI tasks. Event-related desynchronization represents the cortical activation state and stronger ERD suggests better brain function and plasticity (Ono et al., 2015). Kraeutner et al. (2014) found that the ERD was stronger in motor execution tasks than in MI tasks in non-disabled participants. However, in our study, we only found stronger ERD with no significance in the MA task than in the MI task among the bilateral, ipsilesional, and contralesional hemispheres as well as in the C3/C4 channels (**Figure 3**). As MI was reported to require many of the same processes to execute (Blokland et al., 2014), the current results suggested that MA tasks might present similar cortical activity as MI tasks. Besides, the channels around the sensorimotor areas presented a stronger ERD than the average ERD in the bilateral and unilateral hemispheres during the MA task (**Table 3**). It was reasonable because the sensorimotor areas should be involved mostly in motor-related tasks (Li et al., 2014; Wang et al., 2016). Both fMRI and TMS showed activations in motor-related areas. Wang et al. (2016) found that there were activations among M1, bilateral premotor cortex (PMC), and supplementary motor area (SMA) in the fMRI during the motor execution and MI tasks of wrist motor control. Hummel et al. (2002) applied single-pulse TMS over the PMC (M1) and saw motor evoked potentials (MEPs) increasing when ERD occurred. The process of motor function rehabilitation was related to the motor cortex remodeling. Interestingly, as we can see in **Table 3** and **Figure 3**, strong ERD might not only be present in sensorimotor areas (C3, C4, CP1, CP2, and CP6) but also in nearby electrodes (P4), as well as some remote electrodes (FP2), during the MA task. During the MI task, ERD was present in the sensorimotor areas but most of the strong ERD appeared in the frontal areas (F4, FP1, and FP2).

Figure 5 showed that 9 out 14 patients presented the same positive/negative sign in LI values in the MA and MI tasks. Hanakawa et al. (2008) explored different anatomical locations with fMRI in the brain and found that motor execution and MI shared neural substrates to some extent. Referring to the findings of Hanakawa, we considered that MA and MI tasks might present similar cortical excitability patterns and maintain similar activating balance between the hemispheres, although the extent of balance/LI value was not the same. As EEG is a highly non-linear process with high variability (Schomer and Da Silva, 2012), there could be some difference even in the same patient. Eight out of 14 patients (57%) in MA and 7 out of 14 (50%) in MI showed a negative value in LI in **Figure 5**. This finding was similar to a review (Rossini et al., 2003), although we did not apply somatosensory evoked fields detection as previous studies did. A negative/positive LI was considered to indicate a relatively stronger/lower activation in the ipsilesional hemisphere to the affected hand (Kaiser et al., 2012). Among these patients, they showed more activations in the ipsilesional hemisphere than in the contralesional hemisphere. The activations might lead to a higher BCI accuracy in ipsilesional hemispheres although there were cerebral damages on them.

Relationship in Brain-Computer Interface Accuracies and Event-Related Desynchronization of Motor Attempt and Motor Imagery Tasks

In the previous study (Chen et al., 2021), we tried to explore the relationship between sensorimotor rhythm and upper limb motor impairment (motor dysfunction and spasticity) in MA and MI tasks. It suggests that motor dysfunction may be more correlated to ERS in the MI task and to ERD in the MA task while spasticity may be more correlated to ERD in the MA task. In this study, we focused on BCI accuracy, which is an important parameter for the BCI system. During BCI intervention, high BCI accuracies were essential to a good interaction and stronger ERD was the important foundation for cortical plasticity. Several BCI studies tried to decode motor-related signals for motor control by applying the electrodes around the sensorimotor areas (Shu et al., 2018, 2019; Spüler et al., 2018). In our study, negative correlations were found between ERD and BCI accuracies in channels CP1 and CP2 in the MA task, and between ERD and BCI accuracies in channels C4 in the MI task. These results suggested that the stronger ERD of these channels around sensorimotor areas (CP1, CP2, and C4) might lead to higher BCI accuracies. Although it is not that representative, the negative correlations between C4 as well as CP2 seem to provide evidence for those who applied BCI control by using the EEG signals from the contralesional hemispheres (Antelis et al., 2017; Bundy et al., 2017). Unlike the healthy subjects, people with brain injury may exhibit bilateral hemispheric activations due to compensation. One study (Shu et al., 2018) found that contralesional hemispheres in stroke survivors could also present activations.

These channels around the sensorimotor areas could be used to detect if the motor tasks were well performed and if the cortical excitability was modulated. Besides, it has been reported in several BCI studies that BCI accuracies were significantly associated with the improvement of upper limb motor function (Li et al., 2014) or even related to the rehabilitation efficacy of stroke patients (Bundy et al., 2017; Frolov et al., 2017). Interestingly, in the current study, there were significant positive correlations between ipsilesional BCI accuracies and the FMA scores of the hand part. However, the relationship between BCI accuracies and the improvement in motor function has not been explored in this study.

For healthy subjects, an MI classification of around 60% seems relatively low. However, it might be different in stroke survivors or people with brain injury. In a previous study (Shu et al., 2018), we found that contralesional hemispheres could also present activations, which led to a relatively low classification of around 60%. The reason might be that this activation pattern causes difficulties in distinguishing the motor tasks between the left and right hand. Up to now, there are many ways to increase BCI accuracy, among which algorithm and experimental paradigm are two of the most important methods. We think the further step for improving the current BCI accuracy may improve the experimental paradigm such as adding tactile stimulation as assistance for motor attempt tasks. One study (Shu et al., 2019) aimed at improving the SMR-based BCI accuracy by integrating motor tasks with tactile stimulation, which indicated that appropriate tactile stimulation benefited the BCI accuracy in stroke patients. It also suggested that improving the experimental paradigm can be a step to enhance BCI application. Brain-computer interface accuracies and ERD were both very important in a rehabilitative BCI system while the relationship between BCI accuracies and brain function is not totally clear. It should be further explored to get better decoding. Higher BCI accuracy is good for BCI intervention, but for stroke motor rehabilitation, the cortical response could be more important.

Limitations of Our Study

The limitations in the study include the relatively short duration of single-trial time, the lack of long-term detection, and fMRI data. As the patients tend to be less focused than healthy subjects on the trials, the recording session was kept relatively short to minimize discomfort and to ensure that the patients were focused on the tasks. As a result, each part of one trial was set relatively short and the resting time was also short. In the beta band, the ERS might appear 300 to 500 ms after the end of the movement and last for approximately 1 s. Concurrently, in the alpha band, the power returns to the baseline after several seconds. Thus, in this study, there would not be bias in the beta band while there could have been a slight bias in the alpha band referring to the baseline (Rimbert et al., 2018, 2019b). The long-term change of MA and MI tasks has not been compared to see how BCI accuracy and ERD may change over time. Functional MRI can be added in the study to improve the spatial resolution in explaining the detailed cortical activation locations to explore the differences between MA and MI tasks.

CONCLUSION

In this study, we compared the BCI accuracy and ERD/ERS, as well as LI, between MA and MI tasks in patients with hemiplegia in a self-control design. We found that the MA task achieved higher BCI accuracies than the MI task. There was no significant difference in the ERD/ERS and LI between the tasks. Cortical activation (ERD) may influence BCI accuracy, which should be carefully considered in the BCI motor rehabilitation of patients with hemiplegia.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Committee of Huashan Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SC and XS designed and performed the study and analyzed the data. SC organized the database and wrote the manuscript. XS, HW, LD, JF, and JJ reviewed and edited the manuscript. All authors contributed to the manuscript revision and read and approved the submitted manuscript.

FUNDING

This study was supported by the National Key Research and Development Program Project of China (2018YFC2002300 and 2018YFC2002301), National Natural Innovation Research Group Project (82021002), National Natural Integration Project (91948302), and the National Natural Science Foundation of China (51950410602).

ACKNOWLEDGMENTS

We thank all the volunteers for their participation in the study.

REFERENCES

Agius Anastasi, A., Falzon, O., Camilleri, K., Vella, M., and Muscat, R. (2017). Brain symmetry index in healthy and stroke patients for assessment and prognosis. *Stroke Res. Treat.* 2017:8276136. doi: 10.1155/2017/8276136

Antelis, J. M., Montesano, L., Ramos-Murguialday, A., Birbaumer, N., and Minguez, J. (2017). Decoding Upper Limb Movement Attempt From EEG Measurements of the Contralesional Motor Cortex in Chronic Stroke Patients. *IEEE Trans. Biomed. Eng.* 64, 99–111. doi: 10.1109/TBME.2016.2541084

Bai, Z., Fong, K. N. K., Zhang, J. J., Chan, J., and Ting, K. H. (2020). Immediate and long-term effects of BCI-based rehabilitation of the upper extremity after stroke: a systematic review and meta-analysis. *J. Neuroeng. Rehabil.* 17:57.

Battaglia, C., D' Artibale, E., Fiorilli, G., Piazza, M., Tsopani, D., Giombini, A., et al. (2014). Use of video observation and motor imagery on jumping performance in national rhythmic gymnastics athletes. *Hum. Mov. Sci.* 38, 225–234. doi: 10.1016/j.humov.2014.10.001

Benjamin, B., Tomioka, R., Lemm, S., Kawanabe, M., and Müller, K. R. (2008). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process. Mag.* 25, 41–56. doi: 10.1109/MSP.2007.909009

Biasiucci, A., Leeb, R., Iturrate, I., Perdikis, S., Al-Khodairy, A., Corbet, T., et al. (2018). Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke. *Nat. Commun.* 9:2421. doi: 10.1038/s41467-018-04673-z

Blokland, Y., Spyrou, L., Lerou, J., Mourisse, J., Jan Scheffer, G., van Geffen, G. J., et al. (2015). Detection of attempted movement from the EEG during neuromuscular block: proof of principle study in awake volunteers. *Sci. Rep.* 5:12815. doi: 10.1038/srep12815

Blokland, Y., Spyrou, L., Thijssen, D., Eijsvogels, T., Colier, W., Floor-Westerdijk, M., et al. (2014). Combined EEG-fNIRS decoding of motor attempt and imagery for brain switch control: an offline study in patients with tetraplegia. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 222–229. doi: 10.1109/TNSRE.2013.2292995

Bundy, D. T., Souders, L., Baranyai, K., Leonard, L., Schalk, G., Coker, R., et al. (2017). Contralesional Brain–Computer Interface Control of a Powered Exoskeleton for Motor Recovery in Chronic Stroke Survivors. *Stroke* 48, 1908–1915. doi: 10.1161/STROKEAHA.116.016304

Caria, A., Weber, C., Brötz, D., Ramos, A., Ticini, L. F., Gharabaghi, A., et al. (2011). Chronic stroke recovery after combined BCI training and physiotherapy: a case report. *Psychophysiology* 48, 578–582. doi: 10.1111/j.1469-8986.2010.01117.x

Cassim, F., Szurhaj, W., Sediri, H., Devos, D., Bourriez, J., Poirot, I., et al. (2000). Brief and sustained movements: differences in event-related (de)synchronization (ERD/ERS) patterns. *Clin. Neurophysiol.* 111, 2032–2039.

Chen, S., Shu, X., Jia, J., Wang, H., Ding, L., He, Z., et al. (2021). Relation Between Sensorimotor Rhythm During Motor Attempt/Imagery and Upper-Limb Motor Impairment in Stroke. *Clin. EEG Neurosci.* doi: 10.1177/15500594211019917 [Epub ahead of print].

Cicinelli, P., Traversa, R., and Rossini, P. M. (1997). Post-stroke reorganization of brain motor output to the hand: a 2–4 month follow-up with focal magnetic transcranial stimulation. *Electroencephalogr. Clin. Neurophysiol.* 105, 438–450.

Del Gratta, C., Della Penna, S., Tartaro, A., Ferretti, A., Torquati, K., Bonomo, L., et al. (2000). Topographic organization of the human primary and secondary somatosensory areas: an fMRI study. *Neuroreport* 11, 2035–2043. doi: 10.1097/00001756-200006260-00046

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Di Rienzo, F., Joassy, P., Kanthack, T., MacIntyre, T. E., Debarnot, U., Blache, Y., et al. (2019). Effects of Action Observation and Action Observation Combined with Motor Imagery on Maximal Isometric Strength. *Neuroscience* 418, 82–95. doi: 10.1016/j.neuroscience.2019.08.025

Frolov, A. A., Mokienko, O., Lyukmanov, R., Biryukova, E., Kotov, S., Turbina, L., et al. (2017). Post-stroke rehabilitation training with a motor-imagery-based brain-computer interface (bci)-controlled hand exoskeleton: a randomized controlled multicenter trial. *Front. Neurosci.* 11:400. doi: 10.3389/fnins.2017.00400

Hanakawa, T., Dimyan, M. A., and Hallett, M. (2008). Motor planning, imagery, and execution in the distributed motor network: a time-course study with functional MRI. *Cerebr. Cortex* 18, 2775–2788. doi: 10.1093/cercor/bhn036

Hotz-Boendermaker, S., Funk, M., Summers, P., Brugger, P., Hepp-Reymond, M. C., Curt, A., et al. (2008). Preservation of motor programs in paraplegics as demonstrated by attempted and imagined foot movements. *Neuroimage* 39, 383–394. doi: 10.1016/j.neuroimage.2007.07.065

Hummel, F., Andres, F., Altenmuller, E., Dichgans, J., and Gerloff, C. (2002). Inhibitory control of acquired motor programmes in the human brain. *Brain* 125(Pt 2), 404–420. doi: 10.1093/brain/awf030

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia* 33, 1419–1432.

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage* 14(1 Pt 2), S103–S109. doi: 10.1006/nimg.2001.0832

Johnson, N. N., Carey, J., Edelman, B. J., Doud, A., Grande, A., Lakshminarayan, K., et al. (2018). Combined rTMS and virtual reality brain–computer interface training for motor recovery after stroke. *J. Neural Eng.* 15:16009. doi: 10.1088/1741-2552/aa8ce3

Kaiser, V., Bauernfeind, G., Kreilinger, A., Kaufmann, T., Kübler, A., Neuper, C., et al. (2014). Cortical effects of user training in a motor imagery based brain–computer interface measured by fNIRS and EEG. *Neuroimage* 85, 432–444. doi: 10.1016/j.neuroimage.2013.04.097

Kaiser, V., Daly, I., Pichiorri, F., Mattia, D., Müller-Putz, G. R., and Neuper, C. (2012). Relationship Between Electrical Brain Responses to Motor Imagery and Motor Impairment in Stroke. *Stroke* 43, 2735–2740. doi: 10.1161/STROKEAHA.112.665489

Kaplan, A., Vasilyev, A., Liburkina, S., and Yakovlev, L. (2016). "Poor BCI Performers Still Could Benefit from Motor Imagery Training," in *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience. AC 2016. Lecture Notes in Computer Science*, Vol. 9743, eds D. Schmorrow and C. Fidopiastis (Cham: Springer). doi: 10.1007/978-3-319-39955-3_5

Klem, G. H., Lüders, H. O., Jasper, H. H., and Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 52, 3–6.

Kraeutner, S., Gionfriddo, A., Bardouille, T., and Boe, S. (2014). Motor imagery-based brain activity parallels that of motor execution: evidence from magnetic source imaging of cortical oscillations. *Brain Res.* 1588, 81–91. doi: 10.1016/j.brainres.2014.09.001

Li, M., Liu, Y., Wu, Y., Liu, S., Jia, J., and Zhang, L. (2014). Neurophysiological substrates of stroke patients with motor imagery-based brain-computer interface training. *Int. J. Neurosci.* 124, 403–415. doi: 10.3109/00207454.2013.850082

Li, W., Li, C., Xiang, Y., Ji, L., Hu, H., and Liu, Y. (2019). Study of the activation in sensorimotor cortex and topological properties of functional brain network following focal vibration on healthy subjects and subacute stroke patients: an EEG study. *Brain Res.* 1722:146338. doi: 10.1016/j.brainres.2019.146338

Lopez-Larraz, E., Antelis, J. M., Montesano, L., Gil-Agudo, A., and Minguez, J. (2012). "Continuous decoding of motor attempt and motor imagery from EEG activity in spinal cord injury patients," in *Proceedings of the 34th Annual International Conference of the IEEE EMBS*, (San Diego, CA: IEEE), 1798–1801.

Lopez-Larraz, E., Ray, A. M., Figueiredo, T. C., Bibian, C., Birbaumer, N., and Ramos-Murguialday, A. (2017). "Stroke lesion location influences the decoding of movement intention from EEG," in *Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, NY, 3065–3068.

Miller, K. J., Schalk, G., Fetz, E. E., den Nijs, M., Ojemann, J. G., and Rao, R. P. N. (2010). Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4430–4435. doi: 10.1073/pnas.0913697107

Mizuno, K., Abe, T., Ushiba, J., Kawakami, M., Ohwa, T., Hagimura, K., et al. (2018). Evaluating the effectiveness and safety of the electroencephalogram-based brain-machine interface rehabilitation system for patients with severe hemiparetic stroke: protocol for a randomized controlled trial (BEST-BRAIN Trial. *JMIR Res. Protoc.* 7:e12339.

Müller-Putz, G. R., Zimmermann, D., Graimann, B., Nestinger, K., Korisek, G., and Pfurtscheller, G. (2007). Event-related beta EEG-changes during passive and attempted foot movements in paraplegic patients. *Brain Res.* 1137, 84–91. doi: 10.1016/j.brainres.2006.12.052

Muralidharan, A., Chae, J., and Taylor, D. M. (2011). Extracting attempted hand movements from EEGs in people with complete hand paralysis following stroke. *Front. Neurosci.* 5:39. doi: 10.3389/fnins.2011.00039

Niazi, I. K., Mrachacz-Kersting, N., Jiang, N., Dremstrup, K., and Farina, D. (2012). Peripheral electrical stimulation triggered by self-paced detection of motor intention enhances motor evoked potentials. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20, 595–604. doi: 10.1109/TNSRE.2012.2194309

Nikulin, V. V., Hohlefeld, F. U., Jacobs, A. M., and Curio, G. (2008). Quasi-movements: a novel motor–cognitive phenomenon. *Neuropsychologia* 46, 727–742. doi: 10.1016/j.neuropsychologia.2007.10.008

Ono, T., Tomita, Y., Inose, M., Ota, T., Kimura, A., Liu, M., et al. (2015). Multimodal sensory feedback associated with motor attempts alters bold responses to paralyzed hand movement in chronic stroke patients. *Brain Topogr.* 28, 340–351.

Peng, W. W., Tang, Z. Y., Zhang, F. R., Li, H., Kong, Y. Z., Iannetti, G. D., et al. (2019). Neurobiological mechanisms of TENS-induced analgesia. *Neuroimage* 195, 396–408. doi: 10.1016/j.neuroimage.2019.03.077

Pfurtscheller, G., Lopes, and da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 11, 1842–1857.

Pichiorri, F., Morone, G., Petti, M., Toppi, J., Pisotta, I., Molinari, M., et al. (2015). Brain-computer interface boosts motor imagery practice during stroke recovery. *Ann. Neurol.* 77, 851–865. doi: 10.1002/ana.24390

Pivik, R. T., Broughton, R. J., Coppola, R., Davidson, R. J., Fox, N., and Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology* 30, 547–558.

Ramos-Murguialday, A., Broetz, D., Rea, M., Läer, L., Yilmaz, Ö., Brasil, F. L., et al. (2013). Brain-machine interface in chronic stroke rehabilitation: a controlled study. *Ann. Neurol.* 74, 100–108. doi: 10.1002/ana.23879

Ramos-Murguialday, A., Curado, M. R., Broetz, D., Yilmaz, Ö., Brasil, F. L., Liberati, G., et al. (2019). Brain-machine interface in chronic stroke:

randomized trial long-term follow-up. *Neurorehabil. Neural Repair* 33, 188–198. doi: 10.1177/1545968319827573

Rathee, D., Chowdhury, A., Meena, Y. K., Dutta, A., McDonough, S., and Prasad, G. (2019). Brain-machine interface-driven post-stroke upper-limb functional recovery correlates with beta-band mediated cortical networks. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1020–1031. doi: 10.1109/TNSRE.2019.2908125

Rimbert, S., Al-Chwa, R., Zaepffel, M., and Bougrain, L. (2018). Electroencephalographic modulations during an open- or closed-eyes motor task. *PeerJ* 6:e4492. doi: 10.7717/peerj.4492

Rimbert, S., Riff, P., Gayraud, N., Schmartz, D., and Bougrain, L. (2019a). Median nerve stimulation based bci: a new approach to detect intraoperative awareness during general Anesthesia. *Front. Neurosci.* 13:622. doi: 10.3389/fnins.2019.00622

Rimbert, S., Zaepffel, M., Riff, P., Adam, P., and Bougrain, L. (2019b). Hypnotic state modulates sensorimotor beta rhythms during real movement and motor imagery. *Front. Psychol.* 10:2341. doi: 10.3389/fpsyg.2019.02341

Rossini, P. M., Calautti, C., Pauri, F., and Baron, J. (2003). Post-stroke plastic reorganisation in the adult brain. *Lancet Neurol.* 2, 493–502. doi: 10.1016/S1474-4422(03)00485-X

Schomer, D. L., and Da Silva, F. L. (2012). *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Boston, MA: Lippincott Williams & Wilkins.

Shu, X., Chen, S., Meng, J., Yao, L., Sheng, X., Jia, J., et al. (2019). Tactile stimulation improves sensorimotor rhythm-based BCI performance in stroke patients. *IEEE Trans. Biomed. Eng.* 66, 1987–1995. doi: 10.1109/TBME.2018.2882075

Shu, X., Chen, S., Yao, L., Sheng, X., Zhang, D., Jiang, N., et al. (2018). Fast recognition of BCI-inefficient users using physiological features from EEG signals: a screening study of stroke patients. *Front. Neurosci.* 12:93. doi: 10.3389/fnins.2018.00093

Spüler, M., López-Larraz, E., and Ramos-Murguialday, A. (2018). On the design of EEG-based movement decoders for completely paralyzed stroke patients. *J. Neuroeng. Rehabil.* 15:110. doi: 10.1186/s12984-018-0438-z

Takemi, M., Masakado, Y., Liu, M., and Ushiba, J. (2013). Event-related desynchronization reflects downregulation of intracortical inhibition in human primary motor cortex. *J. Neurophysiol.* 110, 1158–1166. doi: 10.1152/jn.01092.2012

Tecchio, F., Rossini, P. M., Pizzella, V., Cassetta, E., and Romani, G. L. (1997). Spatial properties and interhemispheric differences of the sensory hand cortical representation: a neuromagnetic study. *Brain Res.* 767, 100–108.

Wang, L., Zhang, J., Zhang, Y., Yan, R., Liu, H., and Qiu, M. (2016). Conditional granger causality analysis of effective connectivity during motor imagery and motor execution in stroke patients. *Biomed Res. Int.* 2016:3870863. doi: 10.1155/2016/3870863

Wolpaw, D. J., Miner, L. A., Vaughan, T. M., and Wolpaw, J. R. (2000). Mu and beta rhythm topographies during motor imagery and actual movements. *Brain Topogr.* 12, 177–186.

Wong, W., Chan, S., Tang, K., Meng, F., and Tong, K. (2013). Neural correlates of motor impairment during motor imagery and motor execution in sub-cortical stroke. *Brain Inj.* 27, 651–663. doi: 10.3109/02699052.2013.771796

Yao, L., Meng, J., Zhang, D., Sheng, X., Zhu, X., and Zhan, W. (2013). Selective sensation based brain-computer interface via mechanical vibrotactile stimulation. *PLoS One* 8:e64784. doi: 10.1371/journal.pone.0064784

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 23 November 2021
doi: 10.3389/fnbot.2021.692183

Evaluating Convolutional Neural Networks as a Method of EEG–EMG Fusion

Jacob Tryon [1] ***and Ana Luisa Trejos*** [1,2]*

[1] School of Biomedical Engineering, Western University, London, ON, Canada, [2] Department of Electrical and Computer Engineering, Western University, London, ON, Canada

Wearable robotic exoskeletons have emerged as an exciting new treatment tool for disorders affecting mobility; however, the human–machine interface, used by the patient for device control, requires further improvement before robotic assistance and rehabilitation can be widely adopted. One method, made possible through advancements in machine learning technology, is the use of bioelectrical signals, such as electroencephalography (EEG) and electromyography (EMG), to classify the user's actions and intentions. While classification using these signals has been demonstrated for many relevant control tasks, such as motion intention detection and gesture recognition, challenges in decoding the bioelectrical signals have caused researchers to seek methods for improving the accuracy of these models. One such method is the use of EEG–EMG fusion, creating a classification model that decodes information from both EEG and EMG signals simultaneously to increase the amount of available information. So far, EEG–EMG fusion has been implemented using traditional machine learning methods that rely on manual feature extraction; however, new machine learning methods have emerged that can automatically extract relevant information from a dataset, which may prove beneficial during EEG–EMG fusion. In this study, Convolutional Neural Network (CNN) models were developed using combined EEG–EMG inputs to determine if they have potential as a method of EEG–EMG fusion that automatically extracts relevant information from both signals simultaneously. EEG and EMG signals were recorded during elbow flexion–extension and used to develop CNN models based on time–frequency (spectrogram) and time (filtered signal) domain image inputs. The results show a mean accuracy of 80.51 ± 8.07% for a three-class output (33.33% chance level), with an F-score of 80.74%, using time–frequency domain-based models. This work demonstrates the viability of CNNs as a new method of EEG–EMG fusion and evaluates different signal representations to determine the best implementation of a combined EEG–EMG CNN. It leverages modern machine learning methods to advance EEG–EMG fusion, which will ultimately lead to improvements in the usability of wearable robotic exoskeletons.

Keywords: convolutional neural networks, EEG signals, EMG signals, human-machine interfaces, sensor fusion

Edited by:
Dingguo Zhang,
University of Bath, United Kingdom

Reviewed by:
Theerawit Wilaiprasitporn,
Vidyasirimedhi Institute of Science and Technology, Thailand
Yinfeng Fang,
Hangzhou Dianzi University, China

****Correspondence:***
Ana Luisa Trejos
atrejos@uwo.ca

Received: *07 April 2021*
Accepted: *28 October 2021*
Published: *23 November 2021*

Citation:
Tryon J and Trejos AL (2021) Evaluating Convolutional Neural Networks as a Method of EEG–EMG Fusion.
Front. Neurorobot. 15:692183.
doi: 10.3389/fnbot.2021.692183

1. INTRODUCTION

The field of assistive and rehabilitation robotics is rapidly growing, seeking to leverage modern technological advancements to help patients suffering from mobility issues to restore their quality of life. With musculoskeletal disorders being the largest contributor to worldwide disability (World Health Organization, 2019), there is a large market for such devices to help supplement the treatment provided by traditional therapy. Wearable upper-limb robotic exoskeletons, in particular, present a promising option for rehabilitation and assistance, since the patient can use the device during daily life to help assist with tasks, and they are not constrained to a single location during rehabilitation therapy. These devices, however, are still limited in their use, and one reason for this is that further development is required to advance the intelligence of the control methods used in these systems (Desplenter et al., 2020). The devices should be controlled in such a way that their use feels natural and comfortable for the user, regardless of the task being performed.

One popularly explored method to achieve this is the use of bioelectrical signals, produced by the body during motion, to directly control the wearable robotic exoskeletons by detecting the user's motion intention and movement activity based on the information encoded in these signals. Two popularly used bioelectrical signals are electroencephalography (EEG), recorded from brain activity, and electromyography (EMG), recorded from muscle activity (Sawangjai et al., 2020; Leelaarporn et al., 2021). These signals are measured using electrodes on the skin and can be decoded (often through the use of machine learning techniques) to facilitate the control of wearable robotic systems. Typically, devices will only make use of one bioelectrical signal type at a time (Desplenter et al., 2020); however, studies have emerged that have shown that the simultaneous use of EEG and EMG together can improve system performance (Leeb et al., 2011; Dulantha Lalitharatne et al., 2013; Xie et al., 2013; Novak and Riener, 2015; Li et al., 2017; Wöhrle et al., 2017; Loopez-Larraz et al., 2018; Sbargoud et al., 2019; Tryon et al., 2019; Gordleeva et al., 2020; Tortora et al., 2020; Tryon and Trejos, 2021). It has been shown that EEG–EMG fusion can improve classification accuracy as well as reliability, by leveraging the benefits of both signal types simultaneously. An example of this is the use of EEG–EMG fusion as a method to combat the effect of muscle fatigue on system performance. Studies have shown that EEG–EMG fusion models can maintain sufficient accuracy even during EMG signal attenuation brought on by muscle fatigue (Leeb et al., 2011; Tortora et al., 2020), demonstrating the increased reliability that can be obtained through the use of multiple signal types simultaneously. Typically, EEG–EMG fusion is used with machine learning to perform a classification task relevant to the control of a robotic exoskeleton device (for example, motion intention detection). A commonly used method to incorporate EEG–EMG fusion into machine-learning-based classification is to perform EEG–EMG fusion at the decision level, meaning that two classifiers are trained (one for EEG, one for EMG) and their outputs are combined using various techniques (Leeb et al., 2011; Wöhrle et al., 2017; Sbargoud et al., 2019; Tryon et al., 2019; Gordleeva et al., 2020; Tortora et al., 2020; Tryon and Trejos, 2021). Use of this method has been successfully demonstrated for tasks such as motion classification, for example, obtaining an accuracy of 92.0% while outperforming EEG and EMG only models (Leeb et al., 2011). Some examples exist of EEG–EMG fusion happening at the input level, meaning that EEG and EMG features are used simultaneously to train one classifier (Xie et al., 2013; Li et al., 2017; Loopez-Larraz et al., 2018; Tryon et al., 2019; Gordleeva et al., 2020; Tryon and Trejos, 2021). Studies that focus on this technique have been able to show accuracies similar to decision-level fusion studies, in one example obtaining an accuracy of 91.7% using a single classifier for gesture recognition (Li et al., 2017); however, when compared with decision-level fusion in the same study, input-level fusion is often found to yield poorer results (Gordleeva et al., 2020; Tryon and Trejos, 2021).

Despite promising results, further development is needed for EEG–EMG fusion techniques to improve their viability for use in wearable robotic systems. The vast majority of EEG–EMG fusion has been done using traditional machine learning methods that rely on manual feature extraction before training the classifier. Recently, new machine learning methods (often referred to as deep learning) have emerged that are capable of automatically extracting feature information from inputs. One of the most notable implementations of deep learning is the Convolutional Neural Network (CNN). These CNN models, originally developed for the image processing domain, work by using convolution layers that extract information from around an image before feeding it into traditional neural network layers (called Fully Connected layers). The model is not only able to learn patterns from within the data, like traditional machine learning, but also automatically learn what relevant information to extract from the input (instead of relying on the user to specify this manually through selection of appropriate features). The success of CNN classifiers have caused them to move beyond the image processing domain into other areas, with bioelectical signal classification being one of them. For both EEG (Roy et al., 2019) and EMG (Phinyomark and Scheme, 2018), CNNs are the most popularly used deep learning technique. Many studies have shown great results when using CNNs with EEG (Schirrmeister et al., 2017; Wang et al., 2018; Amin et al., 2019; Chaudhary et al., 2019; Dai et al., 2019; Ditthapron et al., 2019; Li et al., 2019; Tayeb et al., 2019; Zhang et al., 2019; Zhao et al., 2019; Tang et al., 2020; Wilaiprasitporn et al., 2020) and EMG (Atzori et al., 2016; Zhai et al., 2017; Ameri et al., 2018; Ding et al., 2018; Xia et al., 2018; Zia ur Rehman et al., 2018; Côté-Allard et al., 2019; Duan et al., 2019; Chen et al., 2020; Fang et al., 2021) signals and have been able to perform many tasks relevant to control, such as hand gesture recognition or the implementation of a Brain Computer Interface, based on Motor Imagery, to send device commands. A brief selection of CNN-based EEG/EMG literature with control-relevant tasks can be seen in **Table 1**. Despite the popularity of CNN models in EEG and EMG literature, the area of EEG–EMG fusion has yet to widely adopt the use of this technique. One study showed that CNNs can be used to fuse EEG and EMG (along with Electrooculography, known as EOG) for sleep stage classification (Banluesombatkul et al., 2021); however, it remains to be seen how an EEG–EMG CNN

TABLE 1 | A summary of select literature examples using CNN models with EEG or EMG signals.

Signal type	Application	Reference
EEG	2 Class motor imagery (e.g., left hand, right hand)	Wang et al., 2018; Chaudhary et al., 2019; Dai et al., 2019; Tayeb et al., 2019; Tang et al., 2020
	4 Class motor imagery (e.g., left hand, right hand, feet, tongue)	Schirrmeister et al., 2017; Amin et al., 2019; Li et al., 2019; Xu et al., 2019; Zhang et al., 2019
	6 Class motor imagery (i.e., elbow flexion/extension, forearm supination/pronation, hand open/close)	Zhao et al., 2019
	Person identification	Wilaiprasitporn et al., 2020
	P300 Classification	Ditthapron et al., 2019
EMG	Hand gesture classification	Zhai et al., 2017; Ding et al., 2018; Zia ur Rehman et al., 2018; Côté-Allard et al., 2019; Duan et al., 2019; Chen et al., 2020; Fang et al., 2021
	Wrist movement classification	Ameri et al., 2018
	Hand movement/Gesture classification	Atzori et al., 2016; Zhai et al., 2017
	Hand position estimation	Xia et al., 2018

The references are grouped by signal type and application.

classifier would perform if used during motion tasks that are relevant for control of assistive and rehabilitation robots. It is possible that the CNN model may extract information about the relationship between the two signals, recorded while the user is moving, that is not currently captured using manually selected features that have been combined for input-level fusion. There is further evidence of CNNs being able to extract information from both EEG and EMG, since a study was done where transfer learning (initially training a classifier for one type of data, then using that classifier with a different set of data) was performed between EEG and EMG datasets with CNNs. The study found that transfer learning was possible between the two signal types to classify concentration levels (EEG) and hand gestures (EMG) (Bird et al., 2020). This may indicate that there is a relationship between the bioelectrical signals that a CNN can detect; therefore, more experimentation is needed to further evaluate CNNs as a method of input level EEG–EMG fusion.

The objective of this work was to evaluate CNNs as a method of EEG–EMG fusion, and to perform an analysis of the feasibility of this technique when used for a classification task relevant to the control of assistive and rehabilitation robots. Multiple methods of representing and combing the EEG/EMG signals at the input level were investigated to see which method of EEG–EMG fusion would provide the best performance within the CNN classifier. This work provides an example of EEG–EMG fusion happening within the CNN classifier, and highlights the most promising methods to use for further development. To facilitate this evaluation, it was decided to train models to classify task weight during dynamic elbow flexion–extension motion. Task weight is the weight a user is holding during movement. This is relevant to the control of wearable robotic exoskeletons during assistance and rehabilitation because the presence of an external weight can affect the stability of a bioelectrical-signal-based control system (Desplenter and Trejos, 2018; Desplenter et al., 2020), as well as the accuracy of control-relevant classification tasks, such as hand gesture recognition (Teh and Hargrove, 2020). These control systems are often tuned for specific movement conditions; hence, being able to detect what the user is holding, will allow the control system to dynamically adapt to the new disturbance and provide more robust performance as the user changes tasks during their daily life. Measuring task weight during dynamic elbow flexion–extension motion provides a more realistic evaluation of the models (as opposed to isometric muscle contraction), since the end goal of EEG–EMG fusion is to use it within a wearable robotic exoskeleton during different motions. Dynamic movement, as well as the more indirect force measurement of task weight, can greatly increase the challenge of performing classification tasks with EEG and EMG signals; hence, EEG–EMG fusion provides a good opportunity to investigate potential improvements to address these limitations. The authors' previous work evaluated EEG–EMG fusion methods for task weight classification, and obtained accuracies of 83.01% using decision-level fusion and 80.75% using input-level fusion; however, this was done using fusion methods based on traditional machine learning classifiers with manual feature extraction (Tryon and Trejos, 2021). This paper focuses on evaluating CNN-based EEG–EMG fusion on the same classification task as a means of comparison.

2. METHODS

2.1. Data Collection and Signal Processing

To develop EEG–EMG-fusion-based CNN models, a dataset of EEG and EMG signals were collected during elbow flexion–extension motion from 32 healthy subjects (mean age 24.9 ± 5.4 years) who were voluntarily recruited following approval from the Human Research Ethics Board at Western University (Project ID: 112023). The data obtained from these subjects were also used in previous studies by the authors (Tryon et al., 2019; Tryon and Trejos, 2021). The subjects were instructed to perform the motion at two speeds level (approximately 10°/s and 150°/s) and three weight levels (0 lbs, 3 lbs, 5 lbs), implementing a 2 × 3 full factorial repeated measures study design. This resulted in six combinations of weight and speed being recorded (each pairing referred to as a trial). The order in which the trials were performed was randomized for each subject to limit any potential biasing effects caused by the ordering of the speed/weight pairings. Within each trial, elbow flexion–extension motion was performed for three repetitions using the subject's dominant arm (30 right handed, 2 left handed), with a 3 s pause in-between repetitions. Between each trial, subjects were given a 1-min rest period. While performing the elbow flexion–extension motion, the subject would self-regulate their motion speed to achieve an approximation of the targeted speed. Subjects were instructed by the experimenter to count seconds while performing each elbow flexion–extension repetition such that a 30 s motion duration was obtained for the slow speed (10°/s) repetitions and a 2 s duration was obtained for the fast speed (150°/s) repetitions. Assuming a 150° range of motion, this resulted in approximately the desired speed for each targeted speed level, while still allowing the subject to move dynamically in an unrestricted manner.

During data collection, the EEG and EMG signals were recorded using an Intronix 2024F Physiological Amplifier System (Intronix Technologies, Bolton, Canada). Both EEG and EMG were sampled at 4,000 Hz and a ground electrode was placed over the elbow bone of the subject's non-dominant arm to act as the system ground for the differential amplifier. The sampling rate of the measurement system was fixed for all channels and could not be altered, which is why it was higher than necessary, particularly for the EEG signals. In an actual wearable robotic device, this sampling rate would be lower to reduce hardware demands.

To record EEG signals, wired gold-cup electrodes, filled with electrically conductive paste, were placed on the subject's scalp above the C3, C4, and Cz locations, as specified by the 10–20 International System. These locations were chosen for this study since they correspond with the motor cortex of the brain, and should provide relevant signal information during movement. Prior to placing the electrodes, the subject's scalp was cleaned at the location of electrode placement with an abrasive gel to ensure that a proper electrical connection was established. Signals were recorded using bipolar channels, configured for a reference montage, with the reference point being an ear-clip electrode attached to the subject's ear lobe. During recording, the EEG signals were filtered with a 0.5–100 Hz band pass filter built into the Intronix 2024F system. After recording, the EEG signals were filtered again in software using a 0.5–40 Hz band pass filter (3^{rd} order Butterworth) (Vaid et al., 2015).

To record EMG signals, bipolar electrodes were placed over the biceps and triceps of the subject's dominant arm, as specified by the SENIAM Project guidelines. These muscles were chosen for this study since they are two of the main muscles that contribute to elbow flexion–extension motion. Prior to electrode placement, the subject's skin at the location of electrode placement was cleaned using an alcohol swab. During recording, the EMG signals were filtered with the measurement system's built-in 20–500 Hz band pass filter. Following recording, the EMG signals had the DC offset removed and were filtered again with another 20–500 Hz band pass filter (4^{th} order Butterworth) (De Luca, 2002).

After filtering, the signals were segmented to remove the portions of the recording where the subject was not moving. This was done using markers that were placed at the beginning and end of the subject's movement. The markers were placed manually by the experimenter during data recording using an external trigger system. Synchronized video recordings of the subject moving were also recorded for verification.

All signal processing and image generation was done offline using MATLAB 2019b with the Signal Processing Toolbox. An overview of the full data processing pipeline can be seen in **Figure 1**.

2.2. Image Generation

Once the EEG and EMG signals were processed and segmented, the next step was to convert the dataset into images that can act as suitable inputs to a CNN classifier. Since CNNs were developed initially as a tool for image recognition problems, their architecture relies on images as inputs; however, since an image is simply an array with a numerical value at each pixel location, it is possible to represent bioelectrical signals in such a way. In previous works that have used EEG and EMG signals as inputs to CNN models, there are two commonly used methods for representing the signals as images: calculating a time–frequency domain representation of the signal to generate spectrogram images (Zhai et al., 2017; Wang et al., 2018; Xia et al., 2018; Chaudhary et al., 2019; Côté-Allard et al., 2019; Dai et al., 2019; Duan et al., 2019; Tayeb et al., 2019) or organizing the processed signals in the time domain to create signal images (Atzori et al., 2016; Schirrmeister et al., 2017; Ameri et al., 2018; Ding et al., 2018; Zia ur Rehman et al., 2018; Amin et al., 2019; Côté-Allard et al., 2019; Li et al., 2019; Tayeb et al., 2019; Zhang et al., 2019; Zhao et al., 2019; Chen et al., 2020; Tang et al., 2020; Fang et al., 2021). Note that the term image here refers merely to a CNN input and does not require the use of an image in the colloquial sense (such as a picture). For example, signal images are just the time series data reshaped into a proper CNN input (discussed further in section 2.2.2) and the convolution is actually being done on the time series signal data directly. Both methods show prominent use with EEG and EMG-based models, with neither method demonstrating an obvious supremacy when it comes to model performance. Also, since EEG and EMG have never been used simultaneously as inputs to a CNN model to classify task weight, it is unclear which image type will allow for the best fusion of EEG and EMG within the classifier. Since both image types use a different domain representation, there is a chance they may target different responses of the signal, offering different information to the CNN classifier. Spectrogram images (time–frequency domain) may trend toward representing the oscillatory behavior of the signals, while the signal images (time domain representation) may trend toward representing time-varying behavior, such as changes in amplitude. However, this is not a given, as the CNN model is free to extract information it deems relevant from the inputs, and it remains to be seen which input method will provide the best performance when classifying task weight. For these reasons, both image types (spectrogram images and signal images) will be evaluated to determine which is the most suitable method to use for EEG–EMG fusion.

To increase the number of images to use for classifier training, during image generation the signals were windowed using a 250 ms window with 50% overlap. This windowing was used for both image types, with both a spectrogram and signal image being generated for each window. A window length of 250 ms was chosen, since studies have shown that 300 ms is the maximum amount of delay a system can experience before the user becomes unable to control the device (Tang et al., 2014). Even though this study was performed offline, limiting the window length to a time that fits within the real-time delay target allows for a more realistic evaluation of the EEG–EMG-fusion-based

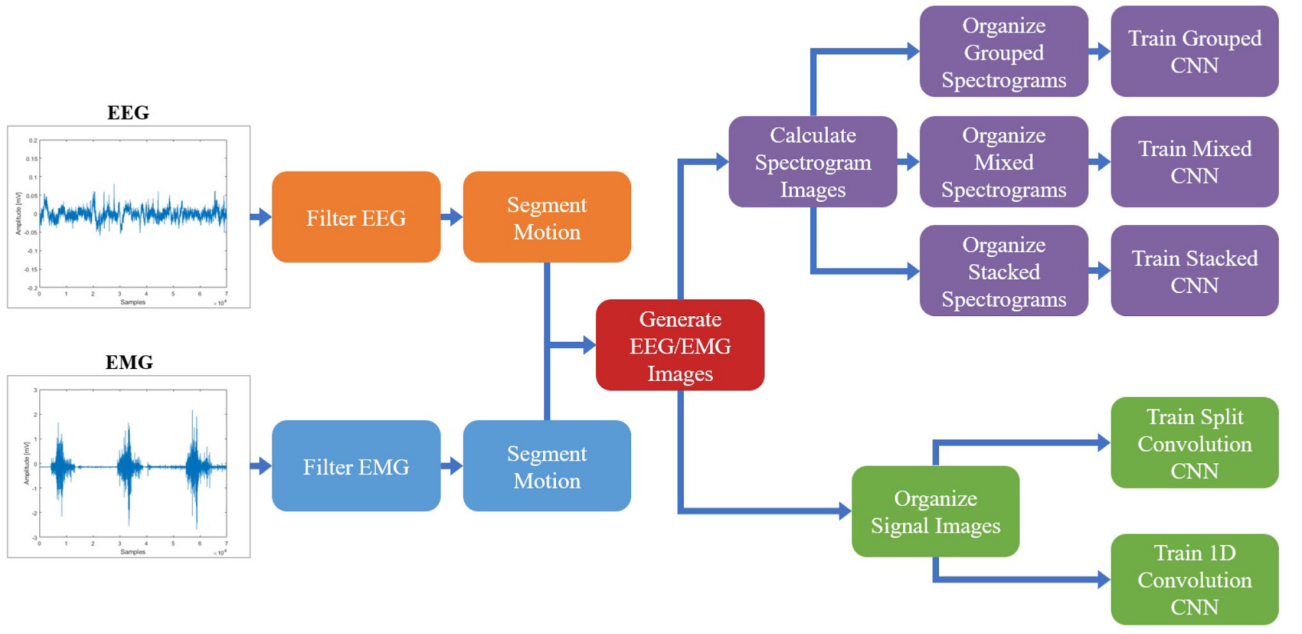


FIGURE 1 | The protocol followed to process the EEG/EMG signals, generate the spectrogram and signal images, and train the CNN models using different EEG–EMG fusion methods. The top path (purple) shows the steps used to develop the CNN models based on spectrogram image inputs, while the bottom path (green) shows the steps used to develop the CNN models based on signal image inputs. For all EEG–EMG-fusion-based CNN model types (represented by the final step of all paths), an EEG and EMG only version was also trained, to provide a baseline comparison for evaluating EEG–EMG Fusion.

CNN models as a potential method of control for assistive and rehabilitation robots.

2.2.1. Spectrogram Images

To generate the spectrogram images, a Short-Time Fourier Transform (STFT) was calculated for each window of the EEG and EMG signals, providing a time–frequency domain representation of the signals. The time and frequency resolution of the STFT was chosen so the resulting images would be of a suitable size for use as an input to a CNN model: large enough to have an appropriate time/frequency resolution, but not so large as to require an infeasible amount of memory and computational power. Using trial and error, a spectrogram image size of 68 × 32, for each signal channel, was chosen. For the time resolution, the STFT was calculated using a Hann window with a length of 56 samples and 75% overlap, which resulted in an image width of 68 pixels (for the 4,000 Hz sampling rate of the measurement system). The frequency resolution of the STFT was chosen so that an image height of 32 pixels would be obtained for the frequency range of interest for both EEG (0.5–40 Hz) and EMG (20–500 Hz). Due to the differences in bandwidth, this meant that EEG and EMG had different STFT frequency resolutions, but their image height was kept the same to simplify their combination into a single image during fusion. The STFT was calculated across the entire frequency range of 0–4,000 Hz using an FFT size of 3,200 for EEG and 256 for EMG. Then, the images were cropped to only include the portions of the image within the respective bandwidth of each signal type. This resulted in five spectrogram images (3 EEG channels and 2 EMG channels) of size 68 × 32 for each time window.

Following image generation, the pixel values of the spectrogram images were normalized to be between 0 and 1. Due to the highly variable nature of EEG and EMG signals between subjects, and the different scale in frequency magnitudes for EEG and EMG obtained from the STFT, the images were normalized for each subject and each signal type. After all spectrogram images were calculated for one subject, the max/min frequency magnitude value for EEG and the max/min frequency magnitude value for EMG were recorded and used to normalize all spectrogram images of that respective signal type for that subject. This ensured that both EEG and EMG spectrograms were given equal proportion within the image, regardless of the differences in signal amplitude present when recording both bioelectrical signals. This also ensured that differences observed in subject recordings did not cause certain images in the dataset to be improperly scaled based on an outlier subject.

Once the spectrogram images had been normalized, they were combined to facilitate the fusion of EEG and EMG at the input level. Multiple methods of combining the EEG and EMG spectrogram images were performed, to investigate which method of fusing the EEG and EMG spectrogram images would provide the best model performance. In the first method of fusion (referred to here as the grouped method), the EEG and EMG spectrograms were grouped by signal type and stacked vertically to create a single 68 × 160 image comprised of the five spectrograms. The three EEG spectrograms were placed at the top of the image (in the order of C3, C4, and Cz from top to bottom) and the two EMG spectrograms were placed on the bottom of the image (in the order of biceps, then triceps from top to bottom). This fusion method grouped spectrograms of the same signal type together within the image, causing the convolution of the image to initially happen within the same signal type and only fusing the signals initially along the single border between EEG and EMG. An example of this method can be seen in **Figure 2A**. The second fusion method (referred to here as the mixed method) stacked images vertically once again, but this time EEG and EMG spectrograms were alternated to provide a better mix between signal types. The order from top to bottom went C3, biceps, C4, and triceps, Cz. This method of fusion provides more areas within the image were EEG and EMG will be convolved together during the initial CNN layer, since there are more borders between the EEG and EMG portions of the image. An example of this method can be seen in **Figure 2B**. The final fusion method (referred to here as the stacked method) stacked the images depth-wise to create a multi-channel image, similar to how a color picture will have three values per pixel location to represent levels of red, green, and blue. In this case, every pixel location contained 5 values (one for each EEG and EMG spectrogram) to result in an image with a shape of 68 × 32 × 5. An example of this method can be seen in **Figure 2C**. To provide a baseline comparison for evaluating the fusion methods, spectrograms containing only EEG and only EMG signal information were also generated to see if fusion can outperform using one signal alone. Two spectrogram types were generated for both EEG and EMG: vertically stacked spectrograms (68 × 96 for EEG and 68 × 64 for EMG) to provide single-channel spectrograms to compare to the grouped/mixed methods, and depth-wise stacked spectrograms (68 × 32 × 3 for EEG and 68 × 32 × 2 for EMG) to provide multi-channel spectrograms to compare to the stacked method.

2.2.2. Signal Images

Conversely, generating the signal images only required the time series signals to be organized into an array to form the image, since the convolution is being performed on the time series data directly. After filtering, the five signal channels from each window were stacked vertically to create an image where the width was the number of time samples in that window, and the height was the number of signal channels. This resulted in a 1,000 × 5 image for each window, in which the pixels values of the image were the signal amplitude at that time point (in mV). The width of 1,000 resulted from the 250 ms window length with the 4,000 Hz sample rate used by the measurement system.

The signal images were normalized using the same method as the spectrogram images, by subject and by signal type. The max/min amplitude value of EEG and EMG for each subject was recorded and used to scale all signal values between 0 and 1. To account for magnitude differences between the two signal types, the EEG portion of the image was scaled using the EEG min/max and the EMG portion of the image was scaled using the EMG min/max, preventing the larger EMG values from dominating the image by diminishing the contribution of the smaller magnitude EEG signals. A graphical representation of the normalized signal image can be seen in **Figure 3**. Similar to the spectrogram images, signal images comprising of only EEG and only EMG were also generated to provide a comparison point for evaluating EEG–EMG fusion.

2.2.3. Qualitative Image Response

To help illustrate the response of the EEG/EMG signals during task weight changes, an example normalized spectrogram image along with a plot of the normalized signals for all three weight levels (0 lbs, 3 lbs, and 5 lbs) can be seen in **Figure 4**. Based on this qualitative assessment of the signal and spectrogram images, it can be seen that the images show different behavior in both the time domain and the time–frequency domain, depending on task weight. The distribution of frequency magnitudes across time/channels is different in the spectrogram images and the shape of the time domain signal varies in the signal images. This provides a qualitative demonstration that there are changing patterns within the images for different task weights, which may be able to be detected by the CNN models and used to train a classification model.

2.3. CNN Model Training

Once the dataset of images was developed, the CNN models based on fused EEG–EMG inputs were trained to classify task weight. Model training was done using TensorFlow 2.3.0 with Keras 2.4.3 (Chollet, 2015) in Python 3.8. The models trained were subject specific, meaning that each subject had a model trained using only their data. To accomplish this, each subject's data were split into three parts: training, validation, and testing. The first two repetitions of each speed–weight combination were dedicated as training data, while images generated from the third repetition were separated into two equally sized groups: validation and testing data. To ensure that no bias was induced by the split, the order of the windows within the third motion repetition was randomized and a stratified split was used to ensure a 50/50 division, while keeping the number of observations of each class balanced within the validation and testing set. The validation dataset was used during model optimization while the testing set was kept separate until the final model evaluation, in order to reduce potential for model bias and overfitting.

Model training had two stages. First, the base configuration of the model was determined (via trial and error) to determine design factors such as number of layers, batch size, and optimizer choice, among others. The base configuration used for each model type was the same for all subjects and is discussed further in sections 2.3.1 and 2.3.2. Once a base configuration for the model had been determined, the second stage of training was to tune the model further using hyperparameter optimization. This tuning focused on finding optimal parameter values for the setting of the layers within the set base model design. The structure of the model (e.g., number of layers, types of layers used, etc.) was not changed during this optimization, only select hyperparameter values were updated. Using Keras-Tuner 1.0.1 (O'Malley et al., 2019), the values of select hyperparameters were tuned using the Random Search optimization method to find the set that resulted in the best validation performance. The search

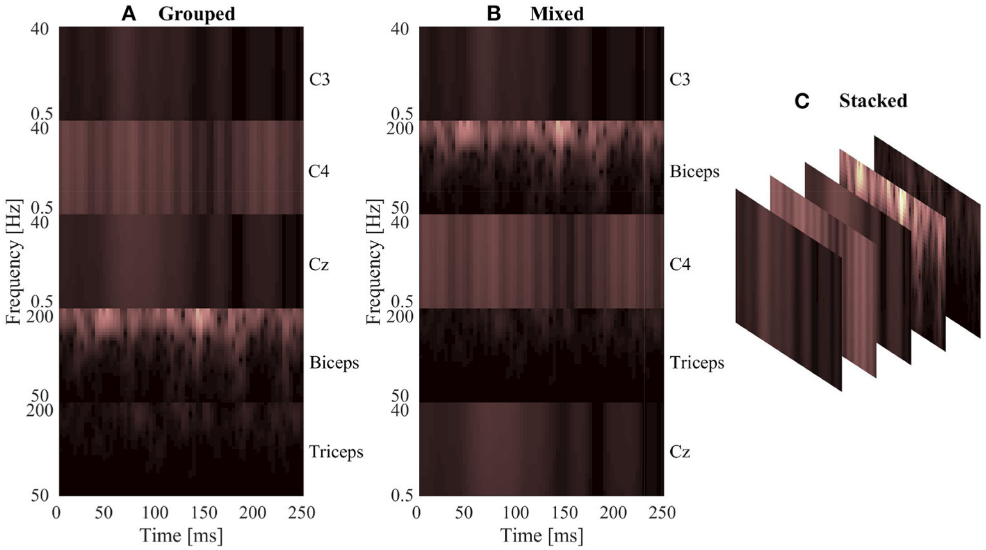


FIGURE 2 | A sample normalized spectrogram image to demonstrate the three EEG–EMG fusion methods used, where **(A,B)** show single-channel spectrograms and **(C)** visualizes a multi-channel spectrogram. **(A)** Shows the grouped method, where signal channels of the same type are grouped together within the image. **(B)** Shows the mixed method, where EEG and EMG channels are alternated to mix signal types. **(C)** Provides a visualization of the stacked method, where a multi-channel spectrogram is generated by combining the different EEG/EMG spectrograms in depth-wise manner.

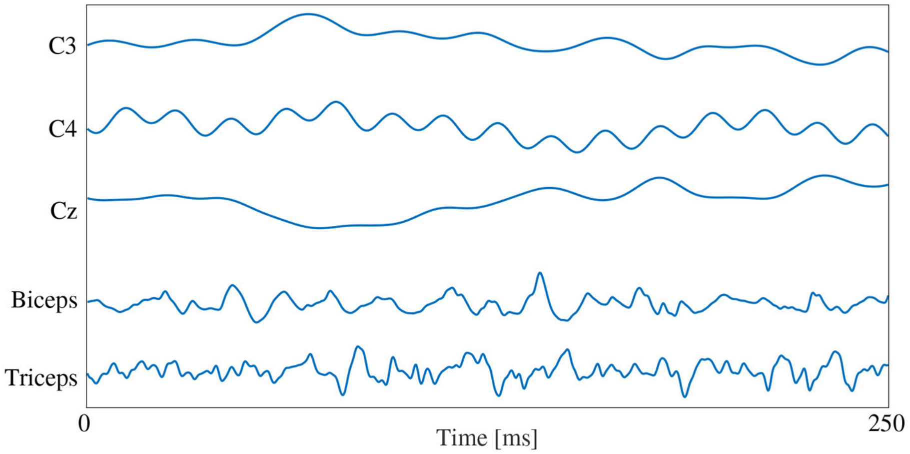


FIGURE 3 | A graphical representation of a sample normalized signal image. The image height contains 5 rows, one for each signal channel, and the image width is dictated by the number of samples in each 250 ms window (1,000 samples at the 4,000 Hz sampling rate).

space checked 50 random combinations of hyperparameters, and trained each combination twice to account for variance in model training. Using the validation dataset, the hyperparameters were evaluated and the set that resulted in the lowest validation loss was selected as the final hyperparameters to use for model training. Bayesian optimization was also tested as a potential method for hyperparameter tuning, but it was found to result in a slight reduction in performance compared to the Random Search method, so it was not used during training of the final models. Early Stopping (using a patience value of five and an epoch limit of 50) was also implemented into model training, using Keras, to stop classifier training once improvements were no longer seen in the validation loss of the model. This was done to prevent overfitting and to speed up training time. All models were optimized and trained using batch size of 32, which was found using trial and error. Categorical Cross-Entropy was used as the loss function with Adaptive Moment Estimation (ADAM) being used as the optimizer for all model types. A Stochastic Gradient Decent (SGD) optimizer was also tested, but it resulted in a reduction in accuracy and longer training times, so ADAM was chosen instead. The hyperparameters being tuned, and their range of possible values, were the same for all subjects; however, each subject had their own hyperparameter optimization performed to adjust the models better to the

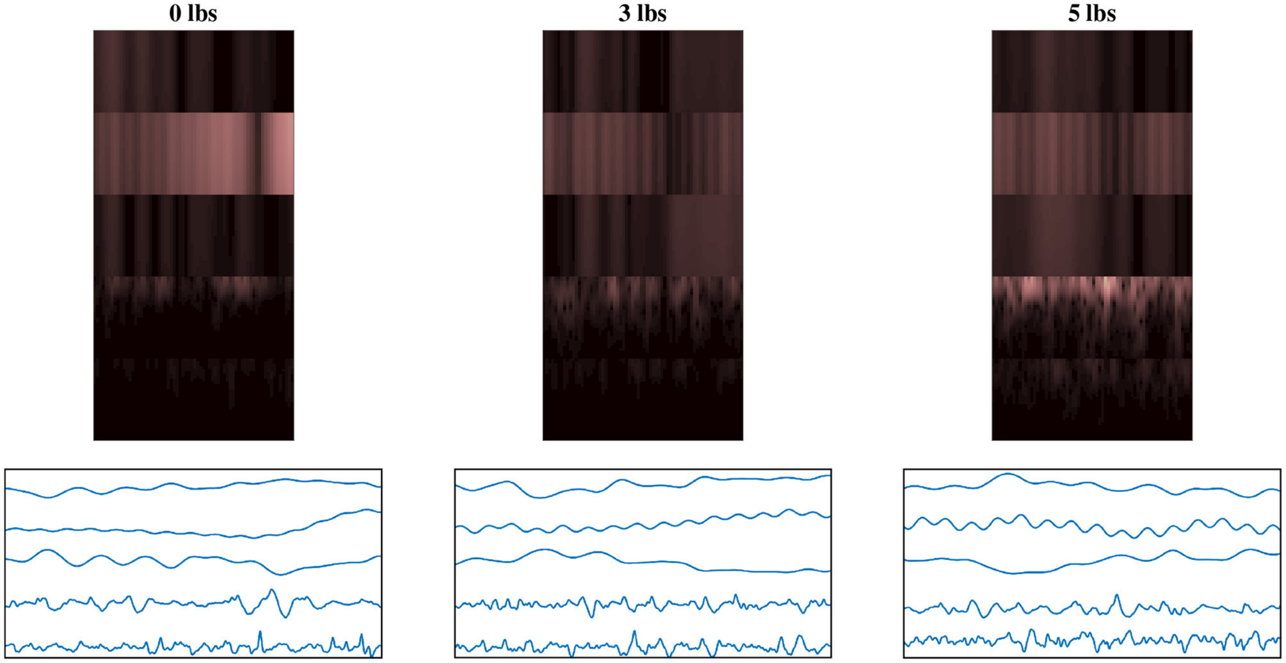


FIGURE 4 | Example normalized spectrogram images and graphical representations of sample normalized signal images for each of the three weight levels, showing the qualitative variations in the images as task weight changes. During different task weights, the distribution of frequency magnitudes across time/channels is different in the spectrogram images and the shape of the time domain signal varies in the signal images. The columns each represent a different task weight level (described by the label above), with the rows being a matched spectrogram and signal image taken from the same time window. The spectrograms shown use the grouped fusion method to arrange the channels. The images shown follow the same labeling convention as the sample images shown in **Figures 2**, **3**, excluded here to avoid clutter.

TABLE 2 | The hyperparameters tuned during optimization, with the range of possible values used by the Random Search algorithm.

Hyperparameter	Parameter values
Kernel size (Spectrogram)	3×3, 5×5, (third layer only) 7×7
Kernel width (Signal)	3–55 (step size of 2)
Filters	8, 16, 32, 64, 128, 256, 512, 1,024
Dropout %	0.0–0.5 (step size of 0.05)
Units (FC Layers)	20–500 (step size of 20)
ADAM learning rate	10^{-5}–10^{-2} (logarithmic sampling)

Unless specified (in brackets next to the hyperparameter name), all hyperparamters and value ranges shown were used for all model types. Two exceptions to this are the kernel size for the stacked models, which were limited to 3 × 3 to account for the smaller image size, and the split convolution filter, which did not include the 1,024 filter setting to prevent an out of memory error while training.

behavior seen in their specific EEG and EMG signals. The hyperparameters that were tuned for each model type can be seen in **Table 2** and are discussed further in sections 2.3.1 and 2.3.2.

2.3.1. Spectrogram CNN Models

A summary of the base model configuration for the spectrogram models can be seen in **Figure 5**. The base configuration for the spectrogram CNN models consisted of three convolution layers followed by two Fully Connected (FC) layers, with a third FC layer used to output the class probabilities. All convolution was done used valid padding, a stride of 1 × 1 and the Rectified Linear Unit (ReLu) for the activation function. Each convolution layer included three sub-layer steps: convolution, followed by a max pooling layer (with a size and stride of 2 × 2), and then a dropout layer to reduce overfitting. Both FC layers contained two sub-layers: the FC step, followed by a dropout layer. Batch Normalization was tested as an alternative to using dropout for the convolution layers, but it led to a reduction in accuracy so it was not used. The output FC layer used a softmax activation function to perform classification. This configuration was used for both the single-channel and multi-channel models (as well as their EEG and EMG only equivalents); the only difference between model types being the size of the inputted image. The hyperparameters chosen for tuning, and the range of values included in the search space, are shown in **Table 2**. Note that these are the same for both model types except for one deviation: the kernel size. For the multi-channel models, the kernel size was fixed at 3 × 3. This was to account for the smaller image size being fed into the model; with certain combinations of larger kernels, the tensor that was passed between convolution layers could be reduced below the minimum allowable size, causing an error in model training.

2.3.2. Signal CNN Models

For the signal CNN models, two base configurations were tested, shown in **Figure 6**. The first configuration employed a method

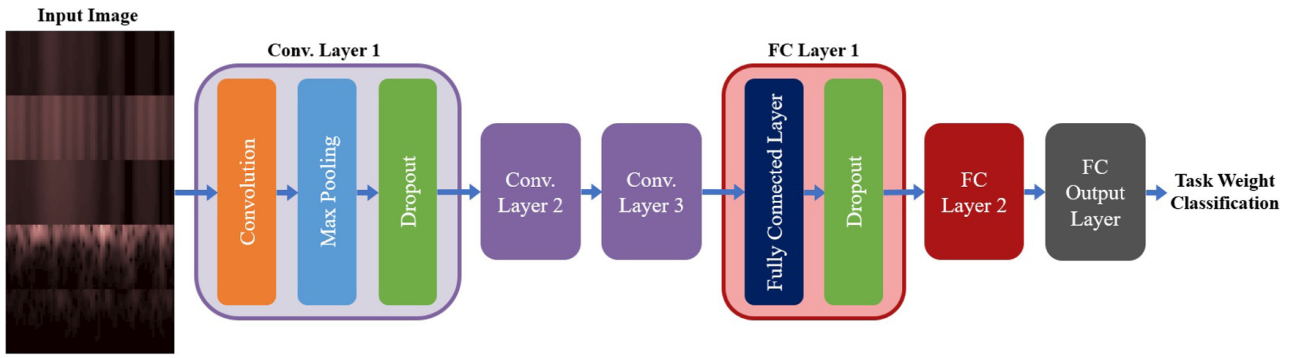


FIGURE 5 | The base model configuration used for all three spectrogram CNN model types. All spectrogram model types used three convolution layers, followed by two FC layers and an output FC layer to perform the final classification. Each convolution layer had three sub-layer steps (convolution, max pooling, and dropout) and each FC layer had two sub-layer steps (the FC step followed by dropout). Note, that repeated layers only show the sub-layers for the first layer, to reduce redundancy and condense the diagram.

commonly used when developing CNN models based on time domain signal inputs for EEG (Schirrmeister et al., 2017; Amin et al., 2019; Li et al., 2019; Zhao et al., 2019), referred to here as split convolution. The name arises from that fact that it takes the first convolution layer and splits it into two back-to-back convolution steps. This method sets the kernel size of the first two convolution layers such that convolution is only happening across one axis of the image at time, with Layer 1 having a kernel size of 1 × kernel width (to only convolve temporally across the time axis of the image) and Layer 2 having a kernel size of image height × 1 (to only convolve spatially across signal channels). The output of the temporal convolution layer is fed directly into the spatial convolution layer, with both layers using valid padding, stride of 1 × 1, and ReLu for the activation function. The output of the temporal convolution layer is fed into a max pooling layer (with a size and stride of 1 × 2), followed by a dropout layer. This is followed up by two FC layers (both using ReLu as the activation function and a dropout sub-layer), then a third output FC layer using a softmax activation function to perform the final classification. A summary of the base model configuration for the split convolution signal model can be seen in **Figure 6A**.

The second base configuration tested for the signal-image-based CNNs used regular one dimensional (1D) convolution layers to train the models. Unlike the split convolution, this layer type convolves across both the time and signal channel axis simultaneously as it moves across the time axis of the image (for this reason only a kernel width is specified, since all signal channels are always included in the convolution). This is a common method of using CNNs for time series signals, so it is useful to see how it compares to the split convolution method commonly seen in the EEG literature. This configuration was similar in makeup to the spectrogram base configuration (except using 1D convolution instead of 2D convolution), comprising of three convolution layers followed by two FC layers and a third FC layer for classification. All convolution layers used valid padding, a stride of 1 and ReLu for the activation function. Each convolution layer followed up the convolution step with a max pooling layer (with a size and stride of 2) then a dropout layer to reduce overfitting. Both FC layers used a dropout layer after the FC step. The output FC layer used a softmax activation function to perform the final classification. A summary of the base model configuration for the 1D convolution signal model can be seen in **Figure 6B**.

Both signal image model types used similar hyperparameter tuning settings; however, there were slight variations between them to account for the differences in the configurations. Due to an out-of-memory error while training, the split convolution models could not use a filter setting of 1,024 and was limited to 512 as the maximum number of filters for any one convolution layer. For both model types, the hyperparameters chosen for tuning, and the range of values included in the search space, are shown in **Table 2**.

2.4. Model Evaluation

Once the optimized models for each subject were trained, they were evaluated to assess the performance of CNN-based EEG–EMG fusion. To achieve this, the withheld test data for each subject were inputted to their final models to obtain predictions about what task weight was being held during each test image. Since three task weights were used during data collection (0 lbs, 3 lbs, and 5 lbs), each classifier was trained to output a three-class prediction, where each output label corresponded to one of three task weights. This output was compared with the actual class label to obtain an accuracy score for each model. This accuracy was then averaged across all subjects to obtain an overall accuracy score for each fusion method, which was then used to compare performance via statistical analysis (performed using IBM SPSS 27). First, the merits of each fusion method were evaluated by comparing EEG–EMG fusion to using EEG and EMG alone. The accuracy scores for each fusion method were compared to the accuracy scores of the EEG/EMG only models of the same model type to see if the increase in accuracy obtained via EEG–EMG fusion was

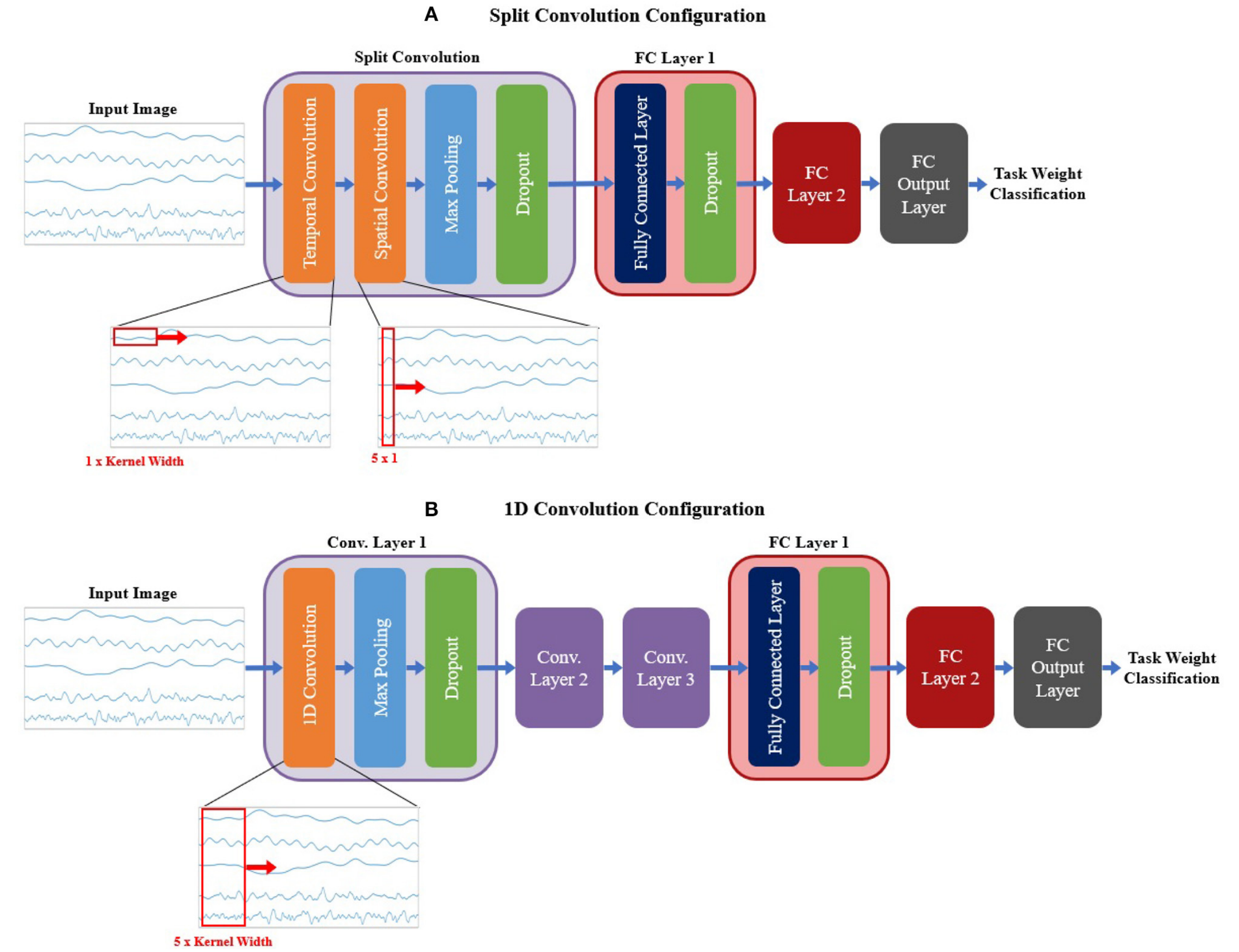


FIGURE 6 | The base model configurations used for the **(A)** split convolution and **(B)** 1D convolution models. Visual representations of the differences between both convolution types are shown in the expanded view below each diagram, detailing the changes in kernel size used to facilitate both types of convolution. Split convolution used one split convolution layer comprised of temporal and spatial convolution sub-layers, followed by a max pooling and dropout sub-layer. 1D convolution used three convolution layers, each with three sub-layer steps (convolution, max pooling, and dropout). All signal model types followed convolution with two FC layers (containing two sub-layer steps: the FC step followed by dropout) and an output FC layer to perform the final classification. Note, that repeated layers only show the sub-layers for the first layer, to reduce redundancy and condense the diagram.

statistically significant. A one-way Within-Subjects Analysis of Variance (ANOVA), followed by pairwise comparisons with the Bonferroni *post-hoc* test, was performed on the accuracy scores for the models of each type (four one-way ANOVAs in total). Separate ANOVAs were used for each model type to account for the different number of models present, depending on the type (the single-channel spectrogram model type contained 4 models, because of the use of both the grouped and mixed fusion methods, while the other model types only contained three models each). This prevents model type from being a factor for a two-way ANOVA, so separate one-way ANOVAs were used instead. Following this, the methods of EEG–EMG fusion were compared to each other using a one-way Within-Subjects ANOVA, (using the Bonferroni *post-hoc* test for pairwise comparisons) to determine if statistically significant differences exist between the accuracy obtained from each fusion method. The purpose of this was to see if any particular EEG–EMG fusion method provided a clear advantage in regard to classification accuracy.

To evaluate the robustness of each model further, the effect of movement speed on accuracy was also evaluated. The classifier output predictions were separated depending on the speed at which the movement was being performed, and accuracy was calculated separately for the fast and slow movement speed groups. Since changes in movement speed during dynamic motion can greatly affect bioelectrical signals, it is important to know how the CNN EEG–EMG fusion models will perform in the presence of such variability. To be useful in the control of robotic devices, the models need to be able to operate adequately during the different speeds required to perform

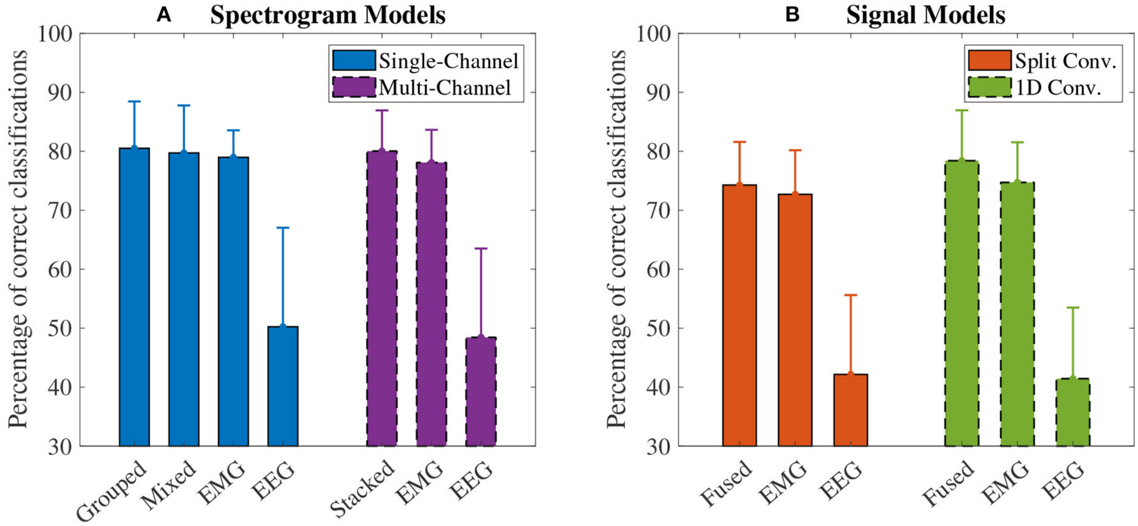


FIGURE 7 | The mean accuracy of all **(A)** spectrogram and **(B)** signal based CNN models, calculated across both speeds and all task weights. Error bars represent one standard deviation. Note that the *y* axis begins at 30% (chance level for these models is 33.33%).

various rehabilitation and assistance tasks. To see if the effect of speed was statistically significant, a two-way Within-Subjects ANOVA was performed on the speed-separated accuracies for each model type. Similar to the model accuracy one-way ANOVA, the two-way ANOVA was performed between models of the same type, resulting in four two-way ANOVAs in total. Note, for all statistical tests performed (on both the overall model accuracy and the speed specific accuracy), a significance threshold of $p < 0.05$ was used.

As a final analysis of model performance, the class predictions from every subject were combined and used to plot a confusion matrix for each CNN model. This was done to observe how the models performed for each task weight and to further verify that the classifiers were adequately trained. To evaluate the model fitting of each classifier further, the confusion matrices were used to calculate the class-wise precision (the likelihood that a class prediction is correct) and recall (the likelihood that all observations of a specific class are correctly classified) scores, to check the balance between both metrics.

3. RESULTS

3.1. Model Accuracy

The accuracy results for the spectrogram-based CNN models are summarized in **Figure 7A**. For all models, the mean accuracy was above chance level (33.33%). The highest accuracy was obtained by the grouped fusion method (80.51 ± 8.07%). This was higher than the other single-channel models, beating the EEG (50.24 ± 17.06%, $p < 0.001$) and mixed fusion method (79.72 ± 8.19%, $p = 0.025$) models, and trending toward a higher mean accuracy than EMG (78.98 ± 4.66%, $p = 1.000$), but the difference between these two was not statistically significant. The next highest performing spectrogram model was the stacked fusion method (80.03 ± 7.02%), which outperformed the multi-channel EEG model (48.44 ± 15.32%, $p < 0.001$), and trended toward a higher accuracy than the multi-channel EMG model (78.09 ± 5.65%, $p = 0.382$), but again this increase in accuracy was not statistically significant. The stacked fusion method also showed a higher mean accuracy than all other single-channel models (except for the grouped fusion method). When comparing the spectrogram fusion methods to their equivalent EEG/EMG model types, the increase in accuracy for all fusion models was statistically significant for EEG, but not EMG; however, a clear trend did emerge, where mean accuracy increased when using EEG–EMG fusion.

The accuracy results for the signal-based CNN models are summarized in **Figure 7B**. Again, all models showed a mean accuracy higher than chance level. The highest accuracy was observed from the 1D convolution EEG–EMG fusion model (78.40 ± 8.70%), which showed a statistically significant increase in accuracy over using EEG alone (41.44 ± 12.25%, $p < 0.001$), but not EMG alone (74.73 ± 6.90%, $p = 0.054$), even though the trend is toward an increase in accuracy. The split convolution EEG–EMG fusion model (74.28 ± 7.42%), while lower than 1D convolution fusion, also showed a statistically significant improvement over using only EEG (42.16 ± 13.67%, $p < 0.001$), but not EMG (72.70 ± 7.60%, $p = 0.401$); however, as with 1D convolution, the mean accuracy tends to increase between the split convolution fusion and EMG only models. When comparing the signal fusion methods to their equivalent EEG/EMG model types, the increase in accuracy for both fusion models was statistically significant for EEG, but not EMG; however, once again a trend did emerge where mean accuracy increased when using EEG–EMG fusion.

For comparing the EEG–EMG fusion methods of all model types together, the results of the pairwise comparisons can be seen in **Table 3**. The mean accuracy for split convolution was found to be statistically significantly lower than all other fusion

TABLE 3 | The *p* values obtained from the pairwise comparisons in the one-way ANOVA comparing the accuracy of the different CNN based EEG–EMG fusion methods.

Fusion method	Grouped	Mixed	Stacked	Split Conv.	1D Conv.
Grouped	-	**0.041**	1.000	**<0.001**	0.431
Mixed	**0.041**	-	1.000	**0.003**	1.000
Stacked	1.000	1.000	-	**<0.001**	1.000
Split Conv.	**<0.001**	**0.003**	**<0.001**	-	**0.018**
1D Conv.	0.431	1.000	1.000	**0.018**	-

Statistically significant values (p < 0.05) are shown in bold.

methods, indicating that it is the worst performing method of fusion. The difference in accuracy between grouped and mixed fusion was also found to be statistically significant, meaning that grouped fusion performed better than mixed within this sample group. Stacked, grouped, and 1D convolution fusion showed no statistical significance in their accuracy differences, meaning that these methods demonstrate similar performance within this sample group. In general, there was a trend of spectrogram-based methods having a higher mean accuracy than signal-based methods (which held true for both EEG–EMG fusion, as well as EEG and EMG alone).

3.2. Speed Specific Accuracy

The accuracy results, separated into the fast and slow speed groups, can be seen in **Figure 8**. For all four model types, the effect of speed was statistically significant ($p < 0.001$ for all). Looking at the plot, it can be seen that performance was significantly worse during the fast speed for all models. All models still remained above the chance level during the fast motion speed; however, EEG accuracy decreased almost to this point (with 1D convolution in particular being essentially at the chance level). It can also be seen that, even when accounting for speed, the trend of EEG–EMG fusion outperforming EEG and slightly increasing accuracy over EMG still remained; however, the increase was much less during fast motion (and in the case of 1D convolution, EMG alone was slightly higher than fusion during the fast speed).

3.3. Classifier Performance

The confusion matrices for all four model types can be seen in **Figures 9–12**, with each figure corresponding to one type of model. For each model type, a confusion matrix is presented for every model (fusion, EEG, and EMG), shown as sub-figures. Looking at the class outputs, it can be seen that all models successfully classified 0 lbs at a much higher rater rate when compared to 3 and 5 lbs (which were similar to each other in performance). An exception to this trend is the two signal-based EEG models (shown in **Figures 11B**, **12B** for split and 1D convolution, respectively), which had generally poor performance for all weight classes. The precision and recall scores for the spectrogram-based models are relatively similar between the two metrics, demonstrating that on average the fit of the models was balanced in its performance. The signal-based models show less balance between the two metrics comparatively, although not to a large degree.

4. DISCUSSION

The goal of this study was to evaluate if CNNs could be used as a new method of input level EEG–EMG fusion to classify task weight during dynamic elbow flexion–extension motion. The hope was that the CNN's ability to automatically learn relevant information from an inputted image may capture aspects of the EEG–EMG relationship not yet found when using manual feature extraction techniques. To this end, this study investigated several methods of representing the EEG–EMG signals as images (to convert the bioelectrical signals into a form suitable for input into a CNN), as well as ways to fuse EEG/EMG during convolution while in image form. This was done to act as a preliminary analysis of these methods, to see which CNN-based EEG–EMG fusion techniques show the most promise to justify their further development. This will ultimately benefit the field of rehabilitation and assistive robotics by providing a new method of EEG–EMG fusion that can be used by the control system of such devices to detect user tasks to adapt accordingly, resulting in devices that are safer and more comfortable to control.

Looking at the model accuracy for each method, it can be seen that all models performed above the chance level (33.33%), and that the precision/recall scores were relatively balanced between the two metrics (albeit less so for the signal-based models than the spectrogram models). This shows that the CNN classifiers were successfully able to decode task weight information from the EEG/EMG signals, indicating that this classification method is feasible for this task. When comparing EEG–EMG fusion to using EEG or EMG alone, a clear trend is seen where using EEG–EMG fusion improves the performance of the models. For all model types, EEG–EMG fusion resulted in some level of accuracy improvement, as well generally higher precision and recall scores (and for the classes where the precision/recall scores were not higher, they were almost the same). Even though no statistically significant difference was found between EEG–EMG fusion and using EMG alone, this does not completely invalidate the use of this new method. Despite the current iteration of these models showing that the improvements gained from using EEG–EMG fusion compared EMG are small, the fact that improvements are consistently observed when using fusion demonstrates that the method shows potential as a tool to improve task weight classification and should be investigated further. By focusing future work on developing improvements to model performance, the accuracy gains of using EEG–EMG fusion may be increased, providing a stronger justification for its use over EMG alone. Based on the trend, it is highly likely that increasing study power through the recruitment of more subjects may result in the difference in accuracy becoming statistically significant. Also, improving the quality of the EEG signals may improve the EEG–EMG fusion models further. Looking at the EEG models, a clear drop in accuracy and classifier performance can be seen when compared to EMG and EEG–EMG fusion, which is likely due to the noisy nature of EEG signals. Due to

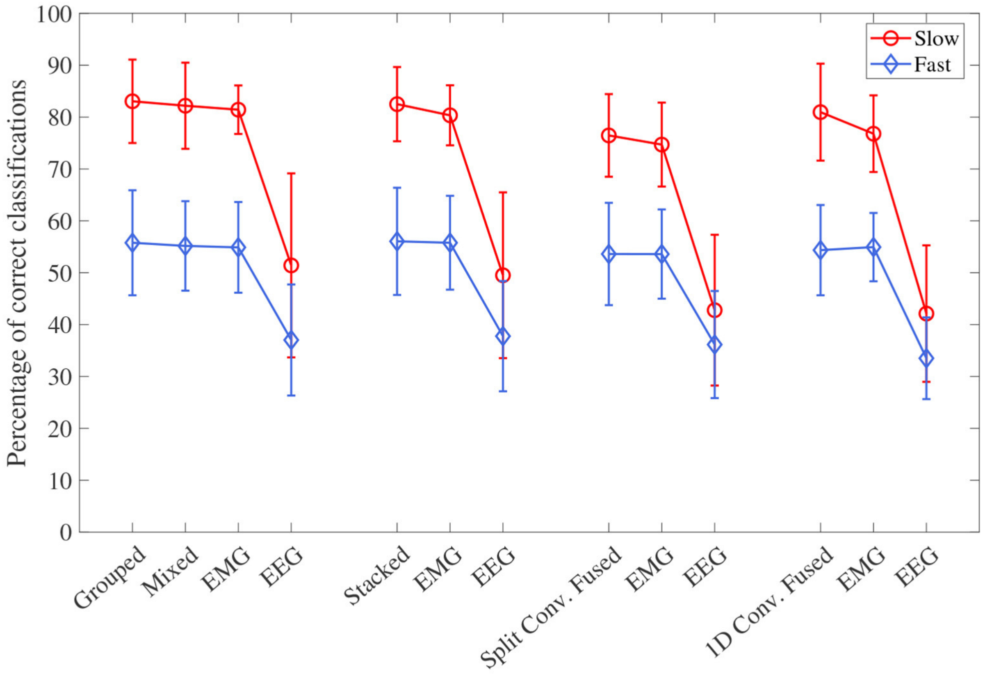
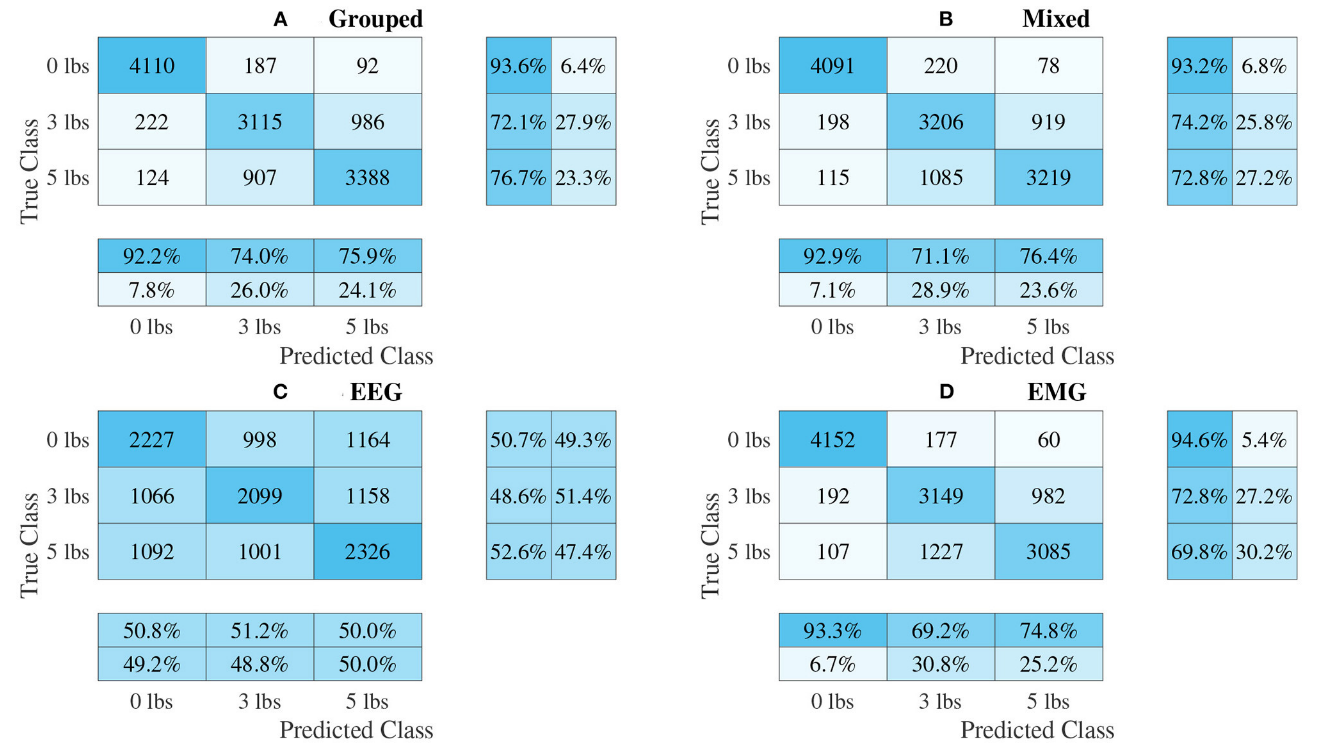


FIGURE 8 | The mean accuracy for all CNN models, separated by the two speed levels (fast and slow). Models of the same type are grouped together, with the order of the groups from left to right as follows: single-channel spectrogram models, multi-channel spectrogram models, split convolution signal models, and 1D convolution signal models. Error bars represent ± one standard deviation.

their significantly smaller signal amplitude, EEG is more prone to signal contamination from motion artifacts and magnetic interference when compared to EMG, which can make it harder to use for classification. The use of more advanced noise rejection techniques and better measurement hardware may improve EEG task weight classification, which should in turn improve the EEG–EMG fusion models. Increasing the amount of EEG channels being used may also help improve the EEG models, as well as EEG–EMG fusion, since it will allow the classifier to draw from more sources from different areas in the brain. However, this trade-off needs to be balanced when using this application for wearable robotics, as these devices are very limited in the hardware resources available. Even though EEG showed worse performance compared to EMG, it was still clearly able to be of some benefit to the EEG–EMG fusion models, since their mean accuracy always tended to be higher than the models based on EMG alone. As a preliminary analysis of EEG–EMG fusion, this work was able to demonstrate that there is a clear benefit to using CNN-based EEG–EMG fusion over just using EEG or EMG alone. It showed a trend of CNN-based EEG–EMG fusion resulting in an increase in mean accuracy, demonstrating the feasibility of these methods and providing a justification for their continued development. Future work should focus on improving these models further to increase the improvements that these techniques provide.

Another objective of this work was to see which methods of combining EEG/EMG would result in the best performance when using CNN models. Looking at the accuracy results of each fusion method, it is clear to see that the CNNs models did perform differently depending on the method used. Of all the fusion methods, split convolution using signal images as inputs performed the worst (and this difference was found to be statistically significant when compared to all other model types). Even though other studies have used this method successfully for classification when only using EEG signals (Schirrmeister et al., 2017; Amin et al., 2019; Li et al., 2019; Zhao et al., 2019), it is clear from this work that it is not suitable when used with EEG/EMG together for task weight classification. For signal-image-based models, using a traditional 1D convolution to perform CNN-based EEG–EMG fusion results in better performance. For the spectrogram-image-based models, it was less obvious which fusion type is superior. The grouped method had the highest mean accuracy, and the increase over the mixed method was statistically significant, which implies that of the two ways to mix EEG/EMG spectrograms, using the grouped method is better. Between grouped and stacked methods though, the difference in accuracy was not statistically significant, so it is less clear which method is best. It should be noted that the stacked spectrogram method is much more computationally efficient than the grouped method (CNNs can perform convolution on a smaller image with multiple channels faster than a larger image with only one channel), which may be a reason to use the stacked method. Since both methods have similar accuracy, the faster method is more ideal, as the end goal of these models is to be used in real time in wearable robotic exoskeletons. Regardless, both methods should be developed



A Grouped

	0 lbs	3 lbs	5 lbs		
0 lbs	4110	187	92	93.6%	6.4%
3 lbs	222	3115	986	72.1%	27.9%
5 lbs	124	907	3388	76.7%	23.3%
	92.2%	74.0%	75.9%		
	7.8%	26.0%	24.1%		

B Mixed

	0 lbs	3 lbs	5 lbs		
0 lbs	4091	220	78	93.2%	6.8%
3 lbs	198	3206	919	74.2%	25.8%
5 lbs	115	1085	3219	72.8%	27.2%
	92.9%	71.1%	76.4%		
	7.1%	28.9%	23.6%		

C EEG

	0 lbs	3 lbs	5 lbs		
0 lbs	2227	998	1164	50.7%	49.3%
3 lbs	1066	2099	1158	48.6%	51.4%
5 lbs	1092	1001	2326	52.6%	47.4%
	50.8%	51.2%	50.0%		
	49.2%	48.8%	50.0%		

D EMG

	0 lbs	3 lbs	5 lbs		
0 lbs	4152	177	60	94.6%	5.4%
3 lbs	192	3149	982	72.8%	27.2%
5 lbs	107	1227	3085	69.8%	30.2%
	93.3%	69.2%	74.8%		
	6.7%	30.8%	25.2%		

FIGURE 9 | Confusion matrices, using the combined classification results for all subjects, for the single-channel spectrogram-based CNN models. **(A)** Shows the matrix for the grouped fusion method while **(B)** shows the matrix for the mixed fusion method. **(C,D)** Show the matrices for the EEG and EMG only models, respectively. Each matrix contains a positive/negative precision score summary in the final two rows, and a positive/negative recall score summary in the final two columns.

further in future work to investigate which method is ultimately superior. Comparing between spectrogram-image-based models and signal-image-based models, it can be seen that, in general, the mean accuracy of spectrogram models was higher. This is also confirmed when looking at the confusion matrices, as the precision and recall scores are not as balanced for the signal models. This even held true for the EEG and EMG only models, in particular EEG, which showed a significant drop in accuracy (as well as precision and recall) for the signal models. This makes sense, since it is well known that much of the relevant information related to motor tasks is encoded in the frequency of the EEG signals (Vaid et al., 2015). It is likely that the time-domain-based representation of the signal images was not able to capture this information as well as the time–frequency-based representation used in the spectrogram images could. This, in turn, would also affect the EEG–EMG fusion methods, which are drawing information EEG, as well as EMG. Despite the lower mean accuracy, no statistically significant difference was found between the 1D convolution, grouped, and stacked methods. This means that even though the trend would make it seem like the 1D convolution method is worse, it should still be considered for future development. One potential benefit of the 1D convolution method is that it requires fewer processing steps to generate the images. Performing a calculation like a STFT can be comparatively time consuming, and computationally expensive, so the use of signal-image-based models may be justified when used in a real-time context for a wearable robotic system. The slight decline in model performance may be outweighed by the efficiency provided by the simpler method; however, further testing and development is needed to confirm this. Since the purpose of this experiment was to investigate the initial feasibility of the different CNN-based EEG–EMG fusion methods, an extensive evaluation of the computational complexity of each algorithm was not performed. The discussion here is based merely on qualitative observations; however, next steps should focus on additional quantitative evaluations of model complexity, which will become essential for moving the models toward a real-time application when integrating them into a wearable device. Ultimately, all three fusion methods (grouped, stacked, and 1D convolution) should continue to be improved and investigated, since there was not one method shown to definitely have better performance and all three methods have clear benefits.

The models can be evaluated further by looking at the speed separated results, as well as the confusion matrices, to examine how robust the classifiers are to changes in task weight and motion speed. Looking at the confusion matrices in **Figures 9–12**, it can be seen that task weight affected classification accuracy. All models were able to recognize the 0 lbs class at a much higher rate than the 3 lbs and 5 lbs classes. While both of these

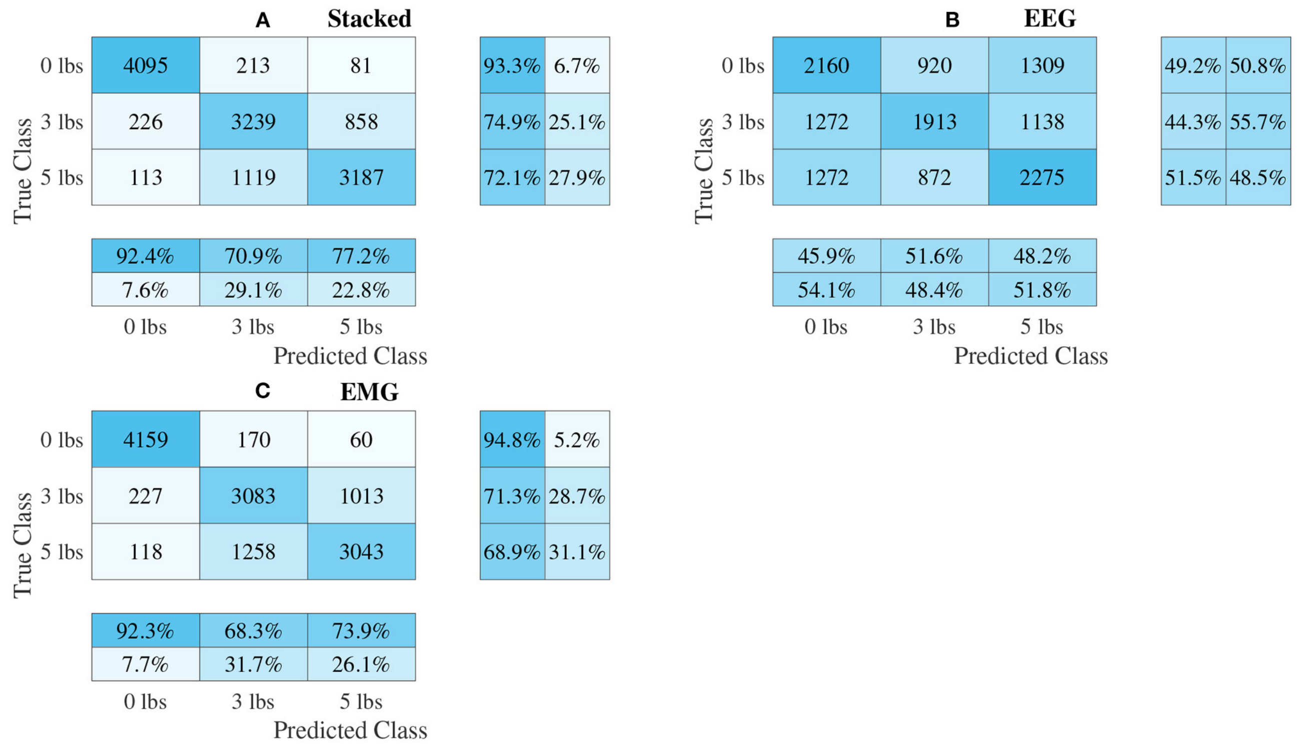
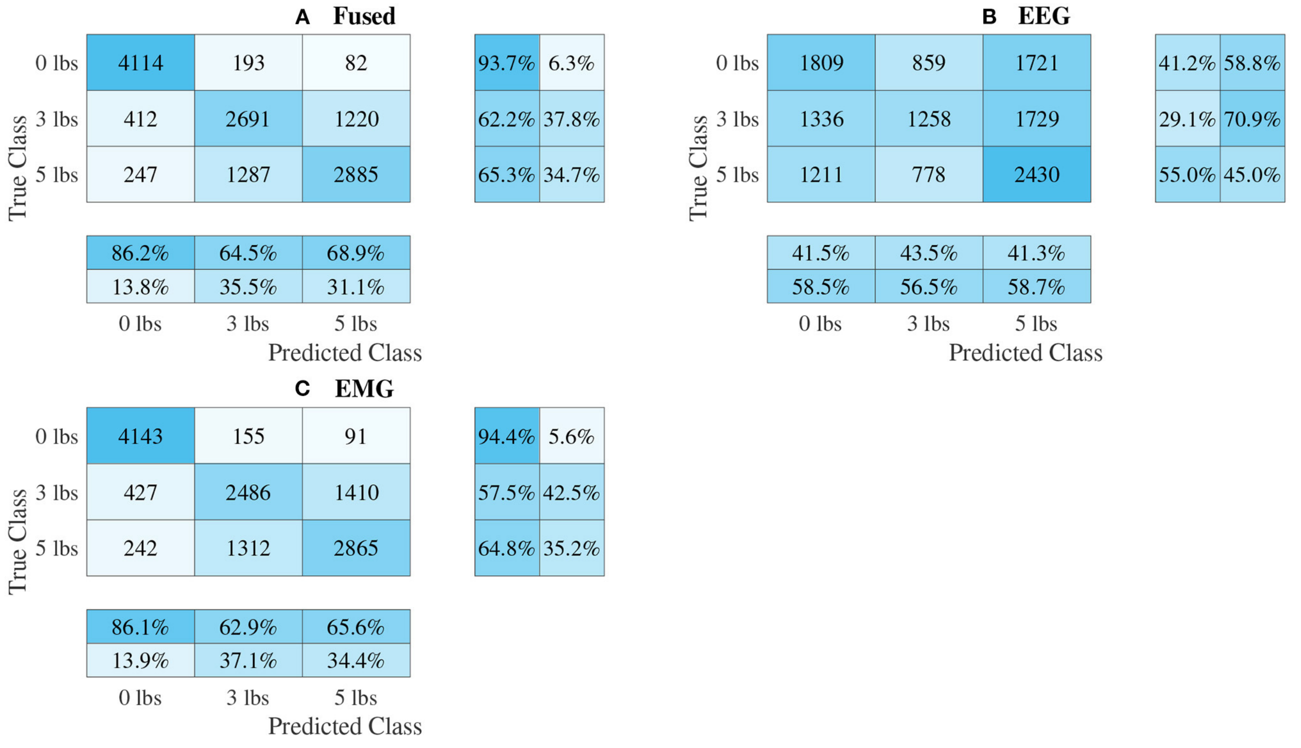


FIGURE 10 | Confusion matrices, using the combined classification results for all subjects, for the multi-channel spectrogram-based CNN models. **(A)** Shows the matrix for the stacked fusion method, while **(B,C)** how the matrices for the EEG and EMG only models, respectively. Each matrix contains a positive/negative precision score summary in the final two rows, and a positive/negative recall score summary in the final two columns.

classes still had relatively good precision and recall scores, 3 lbs and 5 lbs were often misclassified as each other, but not 0 lbs, which implies that the models had a harder time distinguishing the smaller difference in weight. This still does present some level of benefit to a wearable robotic exoskeleton, since even knowing that the user is holding something or not, could be useful for allowing the control system to adapt; however, future work should look at improving the model results further to make them more consistent across different task weights. It is clear from **Figure 8** that speed also has a great effect on performance for all models, with the fast speed having a significantly lower accuracy than the slow speed. The EEG–EMG fusion models were still above chance level when moving at the fast speed, which means that they are still able to recognize relevant patterns in the EEG/EMG signals, just not as effectively. It also should be noted that the trend of EEG–EMG fusion having higher accuracy than using EEG or EMG alone continued, even when separated by speed; however, the increase was very small during fast speed (and the 1D convolution model was actually slightly less accurate than EMG during fast motion). There are multiple things that may be causing this phenomenon. First, faster movements are more likely to cause the EEG and EMG signals to be corrupted by motion artifacts. The more aggressive movements performed by the subject during the fast motion speed may be causing more motion artifacts, which in turn makes the signals harder to use for classification. To alleviate this, more advanced filtering techniques should be used during signal processing to remove this noise. The second reason why the fast motion may be harder to classify is due the nature of task weight classification itself. Despite being related to muscle force (a heavy weight needs more muscle force to move), the task weight itself is not actually a direct measurement of muscle force. The muscle force required to perform an elbow flexion–extension repetition will be a combination of the speed at which the subject was moving and the weight they are holding. It is possible that this is causing smaller weights, moving at a faster speed, to have the appearance of a larger weights at a slower speed, causing the misclassification. EMG in particular may be prone to showing this pattern, since EMG is a measurement of muscle activation. This theory is supported by the authors' previous work, which classified task weight using EEG–EMG-fusion based on traditional machine learning techniques that rely on manual feature extraction. In this study, it was found that all EEG–EMG fusion models showed a statistically significant improvement in accuracy when adding a feature for speed information (in this case a categorical label for fast and slow), seeing improvements of 1.2% for the best performing fusion method (Tryon and Trejos, 2021). Basic knowledge about the speed of the motion given to the classifier was enough to help improve accuracy, so it stands to reason this could be possible for the CNN models as well. Future work should investigate ways to include speed information into the input of the CNN, and evaluate the effect that this has on classifier



A Fused

True Class	0 lbs	3 lbs	5 lbs		
0 lbs	4114	193	82	93.7%	6.3%
3 lbs	412	2691	1220	62.2%	37.8%
5 lbs	247	1287	2885	65.3%	34.7%
	86.2%	64.5%	68.9%		
	13.8%	35.5%	31.1%		

Predicted Class

B EEG

True Class	0 lbs	3 lbs	5 lbs		
0 lbs	1809	859	1721	41.2%	58.8%
3 lbs	1336	1258	1729	29.1%	70.9%
5 lbs	1211	778	2430	55.0%	45.0%
	41.5%	43.5%	41.3%		
	58.5%	56.5%	58.7%		

Predicted Class

C EMG

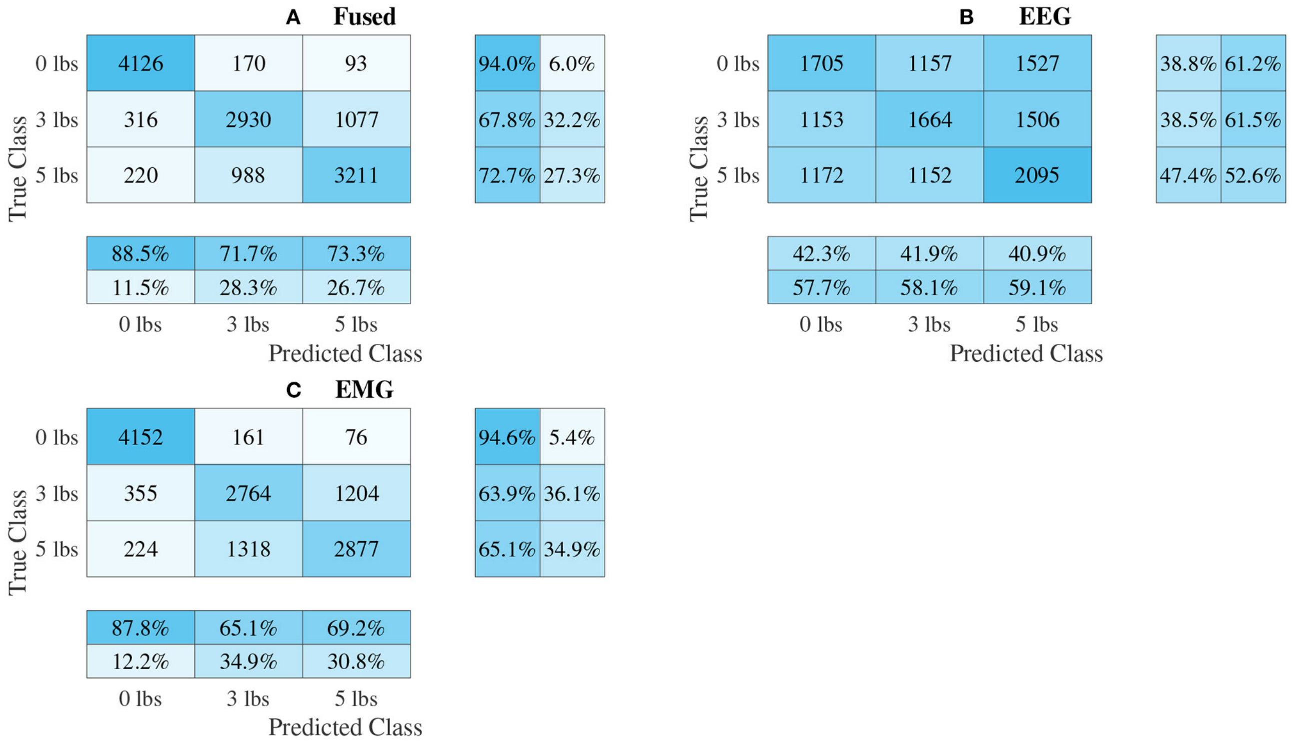
True Class	0 lbs	3 lbs	5 lbs		
0 lbs	4143	155	91	94.4%	5.6%
3 lbs	427	2486	1410	57.5%	42.5%
5 lbs	242	1312	2865	64.8%	35.2%
	86.1%	62.9%	65.6%		
	13.9%	37.1%	34.4%		

Predicted Class

FIGURE 11 | Confusion matrices, using the combined classification results for all subjects, for the split convolution signal-image-based CNN models. **(A)** Shows the matrix for the EEG–EMG fusion model, while **(B,C)** how the matrices for the EEG and EMG only models, respectively. Each matrix contains a positive/negative precision score summary in the final two rows, and a positive/negative recall score summary in the final two columns.

performance. Finally, the reduction in accuracy seen during the fast motion trials could be due to the way the CNN models fit to the data. The nature of how the EEG/EMG signals were windowed mean that there are more observations of movement during the slow speed than the fast speed (since for slow motion it took longer to complete an elbow flexion–extension repetition, and there were the same number of repetitions for both speeds). It is possible that the models became fitted more heavily toward the slow speed data points, causing poorer performance for the fast speed. To account for this, future work should look at collecting more repetitions for the fast motion speed to balance out the classifier training.

Based on the results of this work, CNN-based EEG–EMG fusion has shown to be a feasible method for classification of task weight, and warrants further development to increase the improvements provided by this technique. One potential area for improvement is in the dataset used to train the models. As previously discussed, increasing the number of subjects may improve study power and allow for more statistically significant results; however, this can also allow for the development of generalized models that do not need to be subject specific. Ideally, to allow for ease of use, a wearable robotic exoskeleton should be able to function for any user with minimal training/calibration required. With a large enough sample of the population, general classification models can be pre-trained so that new users can skip the time consuming step of classifier training. An improved dataset can also benefit subject specific models by collecting more elbow flexion–extension repetitions, as well as more combinations of speed and weight. One aspect of CNN models is that their performance can be reduced for smaller training datasets (Luo et al., 2018), so collecting more data per subject should improve performance. More speed/weight combinations will help to provide a more in-depth analysis of the robustness of the classifiers, and will improve their functionality. Since this was the first analysis of CNN-based EEG–EMG fusion, only a small range of weights (0lbs to 5 lbs) and two speeds (approximately 10°/s and 150°/s) were evaluated. It is possible that the inclusion of more task weights, and a larger range of allowable dynamic motion speeds, will affect the classifier performance further, so this effect should be investigated in future work. The current task weight resolution of the classifiers (three weight levels) may limit their use for assistance with daily-living tasks, where the user is unpredictability lifting many objects of varied weights; however, this resolution could still be relevant for more controlled tasks, such as rehabilitation. During rehabilitation exercises, the movement patterns and weight changes performed by the user will be more predictable than activities of daily living, making the use of these classifiers more feasible. The models developed for this work could be used to help the control system of a wearable robotic rehabilitation device automatically adapt changing weights as the user performs different exercises, and will not require the



True Class	0 lbs	3 lbs	5 lbs		
0 lbs	4126	170	93	94.0%	6.0%
3 lbs	316	2930	1077	67.8%	32.2%
5 lbs	220	988	3211	72.7%	27.3%
	88.5%	71.7%	73.3%		
	11.5%	28.3%	26.7%		

True Class	0 lbs	3 lbs	5 lbs		
0 lbs	1705	1157	1527	38.8%	61.2%
3 lbs	1153	1664	1506	38.5%	61.5%
5 lbs	1172	1152	2095	47.4%	52.6%
	42.3%	41.9%	40.9%		
	57.7%	58.1%	59.1%		

True Class	0 lbs	3 lbs	5 lbs		
0 lbs	4152	161	76	94.6%	5.4%
3 lbs	355	2764	1204	63.9%	36.1%
5 lbs	224	1318	2877	65.1%	34.9%
	87.8%	65.1%	69.2%		
	12.2%	34.9%	30.8%		

FIGURE 12 | Confusion matrices, using the combined classification results for all subjects, for the 1D convolution signal-image-based CNN models. **(A)** Shows the matrix for the EEG–EMG fusion model, while **(B,C)** how the matrices for the EEG and EMG only models, respectively. Each matrix contains a positive/negative precision score summary in the final two rows, and a positive/negative recall score summary in the final two columns.

user/therapist to enter the weight change manually, via some external input method, which may feel cumbersome for the user (for example a smartphone app). The ultimate goal, however, is to keep improving the CNN-based EEG–EMG fusion models to increase their resolution, making them a viable tool for use in many different applications, such as assistance with daily tasks.

One method that may improve CNN-based EEG–EMG fusion is to increase the complexity of the models via the inclusion of other deep learning architectures into the model configurations. One popular example of this is the development of models that combine CNNs with Long Short-Term Memory (LSTM) classifiers. LSTM models are beneficial for the classification of information that changes over time, by retaining a memory of inputs (Greff et al., 2017). Since the behavior of EEG and EMG signals will change depending on what stage of elbow flexion–extension motion is currently being evaluated (for example the biceps muscle should be more dominant during flexion), LSTMs may benefit the model by being able to incorporate this information better than using only a CNN. Other studies have shown that CNNs, combined with LSTMs, can be used for EEG (Ditthapron et al., 2019; Zhang et al., 2019; Wilaiprasitporn et al., 2020) and EMG (Xia et al., 2018) classification, and LSTMs alone have been used during decision-level EEG–EMG fusion (Tortora et al., 2020), so there is evidence to suggest that this can be a beneficial technique for improving EEG/EMG models. Future work should evaluate the use of combined CNN–LSTM models for input-level EEG–EMG fusion. Another potential way of improving CNN-based EEG–EMG fusion is to explore other methods of calculating time–frequency signal images. While the STFT is a popular time–frequency representation method, it is far from the only one. Other studies have shown that Wavelet-Transform-based images can also work for EEG (Chaudhary et al., 2019; Xu et al., 2019) and EMG (Côté-Allard et al., 2019) CNN models, so future work should investigate these methods as an alternative to using STFT spectrograms for CNN-based EEG–EMG fusion. Improving these models will move them closer to being practically implemented within a wearable robotic exoskeleton, where they can improve the usability of these devices during rehabilitation and assistive tasks.

5. CONCLUSION

This work demonstrated the feasibility of using CNNs as a method of input level EEG–EMG fusion for task weight classification during dynamic elbow flexion–extension. It presents a new EEG–EMG fusion method that can be used to improve the performance of bioelectrical signal controlled robotic devices for assistance and rehabilitation. During the experiment performed, it was shown that a trend exists where EEG–EMG fusion resulted in a higher mean accuracy compared to using EEG and EMG alone. Different methods of representing the EEG/EMG signals for use in the CNNs were also evaluated, and it was found that time–frequency-image-based models (spectrograms) tended to outperform time domain (signal) models; however, signal models using 1D convolution may still have the potential to match spectrogram model performance. Future work should expand upon the results shown here, and focus on improving performance by increasing model complexity through the inclusion of other deep learning architectures (such as Long Short-Term Memory networks), as well as, investigating other time–frequency image representation methods (such as Wavelet Transforms). It should also focus on improving the training dataset used by collecting EEG/EMG signals during more speed/weight combinations, collecting more motion repetitions from each subject, and collecting data from a larger population of subjects, to allow for a more in-depth analysis of model robustness, as well as better trained models. Using CNNs to facilitate EEG–EMG fusion presents a new way to utilize these bioelectrical signals for the control of wearable robotic devices, and implementing EEG–EMG fusion for task weight classification will allow such devices to adapt to changes in system dynamics so that they can perform assistive and rehabilitation tasks in a more stable and robust way. This will ultimately improve the user experience, leading to safer devices that can be more widely adopted as a new treatment and assistance solution for musculoskeletal disorders.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the Human Research Ethics Board at Western University has not given approval to share the data collected for this study. Requests to access the datasets should be directed to atrejos@uwo.ca.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Research Ethics Board at Western University (Project ID: 112023). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JT and ALT contributed to conception and design of the study. JT performed the data collection, model development, results analysis, and wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This research was funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada under grant RGPIN-2020-05648, by the Canadian Foundation for Innovation (CFI) and the Ontario Research Fund (ORF) under grant 35152, and by the Ontario Ministry of Economic Development, Trade and Employment, and the Ontario Ministry of Research and Innovation through the Early Researcher Award (ER14-10-159).

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Evan Friedman from Intronix Technologies for providing the measurement equipment used for this study.

REFERENCES

Ameri, A., Akhaee, M. A., Scheme, E., and Englehart, K. (2018). Real-time, simultaneous myoelectric control using a convolutional neural network. *PLoS ONE* 13:e0203835. doi: 10.1371/journal.pone.0203835

Amin, S. U., Alsulaiman, M., Muhammad, G., Mekhtiche, M. A., and Shamim Hossain, M. (2019). Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Future Generat. Comput. Syst*. 101, 542–554. doi: 10.1016/j.future.2019.06.027

Atzori, M., Cognolato, M., and Müller, H. (2016). Deep learning with convolutional neural networks applied to electromyography data: a resource for the classification of movements for prosthetic hands. *Front. Neurorob*. 10:9. doi: 10.3389/fnbot.2016.00009

Banluesombatkul, N., Ouppaphan, P., Leelaarporn, P., Lakhan, P., Chaitusaney, B., Jaimchariyatam, N., et al. (2021). MetaSleepLearner: A pilot study on fast adaptation of bio-signals-based sleep stage classifier to new individual subject using meta-learning. *IEEE J. Biomed. Health Inform*. 25, 1949–1963. doi: 10.1109/JBHI.2020.3037693

Bird, J. J., Kobylarz, J., Faria, D. R., Ekart, A., and Ribeiro, E. P. (2020). Cross-domain MLP and CNN transfer learning for biological signal processing: EEG and EMG. *IEEE Access* 8, 54789–54801. doi: 10.1109/ACCESS.2020.2979074

Chaudhary, S., Taran, S., Bajaj, V., and Sengur, A. (2019). Convolutional neural network based approach towards motor imagery tasks EEG signals classification. *IEEE Sens. J*. 19, 4494–4500. doi: 10.1109/JSEN.2019.2899645

Chen, H., Zhang, Y., Li, G., Fang, Y., and Liu, H. (2020). Surface electromyography feature extraction via convolutional neural network. *Int. J. Mach. Learn. Cybern*. 11, 185–196. doi: 10.1007/s13042-019-00966-x

Chollet, F. (2015). *Keras*. Available online at: https://keras.io

Côté-Allard, U., Fall, C. L., Drouin, A., Campeau-Lecours, A., Gosselin, C., Glette, K., et al. (2019). Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE Trans. Neural Syst. Rehabil. Eng*. 27, 760–771. doi: 10.1109/TNSRE.2019.2896269

Dai, M., Zheng, D., Na, R., Wang, S., and Zhang, S. (2019). EEG classification of motor imagery using a novel deep learning framework. *Sensors* 19:551. doi: 10.3390/s19030551

De Luca, C. J. (2002). *Surface electromyography: detection and recording*. DelSys Incorporated, Technical Report.

Desplenter, T., and Trejos, A. L. (2018). Evaluating muscle activation models for elbow motion estimation. *Sensors* 18:1004. doi: 10.3390/s18041004

Desplenter, T., Zhou, Y., Edmonds, B. P., Lidka, M., Goldman, A., and Trejos, A. L. (2020). Rehabilitative and assistive wearable mechatronic upper-limb devices: a review. *J. Rehabil. Assist. Technol. Eng*. 7:2055668320917870. doi: 10.1177/2055668320917870

Ding, Z., Yang, C., Tian, Z., Yi, C., Fu, Y., and Jiang, F. (2018). sEMG-based gesture recognition with convolution neural networks. *Sustainability* 10:1865. doi: 10.3390/su10061865

Ditthapron, A., Banluesombatkul, N., Ketrat, S., Chuangsuwanich, E., and Wilaiprasitporn, T. (2019). Universal joint feature extraction for P300 EEG classification using multi-task autoencoder. *IEEE Access* 7, 68415–68428. doi: 10.1109/ACCESS.2019.2919143

Duan, N., Liu, L. Z., Yu, X. J., Li, Q., and Yeh, S. C. (2019). Classification of multichannel surface-electromyography signals based on convolutional neural networks. *J. Indust. Inf. Integr*. 15, 201–206. doi: 10.1016/j.jii.2018.09.001

Dulantha Lalitharatne, T., Teramoto, K., Hayashi, Y., and Kiguchi, K. (2013). Towards hybrid EEG-EMG-based control approaches to be used in biorobotics applications: current status, challenges and future directions. *PALADYN J. Behav. Rob*. 4, 147–154. doi: 10.2478/pjbr-2013-0009

Fang, Y., Zhang, X., Zhou, D., and Liu, H. (2021). Improve inter-day hand gesture recognition via convolutional neural network based feature fusion. *Int. J. Humanoid Rob*. 18:2050025. doi: 10.1142/S0219843620500255

Gordleeva, S. Y., Lobov, S. A., Grigorev, N. A., Savosenkov, A. O., Shamshin, M. O., Lukoyanov, M. V., et al. (2020). Real-time EEG-EMG human-machine interface-based control system for a lower-limb exoskeleton. *IEEE Access*. 8, 84070–84081. doi: 10.1109/ACCESS.2020.2991812

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., and Schmidhuber, J. (2017). LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst*. 28, 2222–2232. doi: 10.1109/TNNLS.2016.2582924

Leeb, R., Sagha, H., Chavarriaga, R., and Millán, J. D. R. (2011). A hybrid brain-computer interface based on the fusion of electroencephalographic and electromyographic activities. *J. Neural Eng*. 8, 1–5. doi: 10.1088/1741-2560/8/2/025011

Leelaarporn, P., Wachiraphan, P., Kaewlee, T., Udsa, T., Chaisaen, R., Choksatchawathi, T., et al. (2021). Sensor-driven achieving of smart living: a review. *IEEE Sens. J*. 21, 10369–10391. doi: 10.1109/JSEN.2021.3059304

Li, X., Samuel, O. W., Zhang, X., Wang, H., Fang, P., and Li, G. (2017). A motion-classification strategy based on sEMG-EEG signal combination for upper-limb amputees. *J. Neuroeng. Rehabil*. 14, 1–13. doi: 10.1186/s12984-016-0212-z

Li, Y., Zhang, X. R., Zhang, B., Lei, M. Y., Cui, W. G., and Guo, Y. Z. (2019). A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding. *IEEE Trans. Neural Syst. Rehabil. Eng*. 27, 1170–1180. doi: 10.1109/TNSRE.2019.2915621

Loopez-Larraz, E., Birbaumer, N., and Ramos-Murguialday, A. (2018). "A hybrid EEG-EMG BMI improves the detection of movement intention in cortical stroke patients with complete hand paralysis," in *International Conference of the IEEE Engineering in Medicine and Biology Society* (Honolulu: Institute of Electrical and Electronics Engineers Inc.), 2000–2003.

Luo, C., Li, X., Wang, L., He, J., Li, D., and Zhou, J. (2018). "How does the data set affect CNN-based image classification performance?" in *International Conference on Systems and Informatics* (Nanjing), 361–366.

Novak, D., and Riener, R. (2015). A survey of sensor fusion methods in wearable robotics. *Rob. Auton. Syst*. 73, 155–170. doi: 10.1016/j.robot.2014.08.012

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. (2019). *Keras Tuner*. Available online at: https://github.com/keras-team/keras-tuner

Phinyomark, A., and Scheme, E. (2018). EMG pattern recognition in the era of big data and deep learning. *Big Data Cogn. Computi*. 2:21. doi: 10.3390/bdcc2030021

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., and Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *J. Neural Eng*. 16:051001. doi: 10.1088/1741-2552/ab260c

Sawangjai, P., Hompoonsup, S., Leelaarporn, P., Kongwudhikunakorn, S., and Wilaiprasitporn, T. (2020). Consumer grade EEG measuring sensors as research tools: a review. *IEEE Sens. J*. 20, 3996–4024. doi: 10.1109/JSEN.2019.2962874

Sbargoud, F., Djeha, M., Guiatni, M., and Ababou, N. (2019). WPT-ANN and belief theory based EEG/EMG data fusion for movement identification. *Traitement du Signal* 36, 383–391. doi: 10.18280/ts.360502

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., et al. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain. Mapp*. 38, 5391–5420. doi: 10.1002/hbm.23730

Tang, X., Li, W., Li, X., Ma, W., and Dang, X. (2020). Motor imagery EEG recognition based on conditional optimization empirical mode decomposition and multi-scale convolutional neural network. *Expert Syst. Appl*. 149:113285. doi: 10.1016/j.eswa.2020.113285

Tang, Z., Zhang, K., Sun, S., Gao, Z., Zhang, L., and Yang, Z. (2014). An upper-limb power-assist exoskeleton using proportional myoelectric control. *Sensors* 14, 6677–6694. doi: 10.3390/s140406677

Tayeb, Z., Fedjaev, J., Ghaboosi, N., Richter, C., Everding, L., Qu, X., et al. (2019). Validating deep neural networks for online decoding of motor imagery movements from eeg signals. *Sensors* 19:210. doi: 10.3390/s19010210

Teh, Y., and Hargrove, L. J. (2020). "The effects of limb position and external load on offline myoelectric pattern recognition control," in *IEEE International Conference on Biomedical Robotics and Biomechatronics* (New York, NY), 654–659.

Tortora, S., Tonin, L., Chisari, C., Micera, S., Menegatti, E., and Artoni, F. (2020). Hybrid human-machine interface for gait decoding through Bayesian fusion of EEG and EMG classifiers. *Front. Neurorob*. 14:582728. doi: 10.3389/fnbot.2020.582728

Tryon, J., Friedman, E., and Trejos, A. L. (2019). "Performance evaluation of EEG/EMG fusion methods for motion classification," in *IEEE International Conference on Rehabilitation Robotics* (Toronto, ON), 971–976.

Tryon, J., and Trejos, A. L. (2021). Classification of task weight during dynamic motion using EEG-EMG fusion. *IEEE Sens. J*. 21, 5012–5021. doi: 10.1109/JSEN.2020.3033256

Vaid, S., Singh, P., and Kaur, C. (2015). "EEG signal analysis for BCI interface: a review," in *International Conference on Advanced Computing and Communication Technologies* (Haryana), 143–147.

Wang, Z., Cao, L., Zhang, Z., Gong, X., Sun, Y., and Wang, H. (2018). Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition. *Concurrency Comput. Pract. Exp*. 30:e4413. doi: 10.1002/cpe.4413

Wilaiprasitporn, T., Ditthapron, A., Matchaparn, K., Tongbuasirilai, T., Banluesombatkul, N., and Chuangsuwanich, E. (2020). Affective EEG-based person identification using the deep learning approach. *IEEE Trans. Cogn. Dev. Syst*. 12, 486–496. doi: 10.1109/TCDS.2019.2924648

Wöhrle, H., Tabie, M., Kim, S. K., Kirchner, F., and Kirchner, E. A. (2017). A hybrid FPGA-based system for EEG- and EMG-based online movement prediction. *Sensors* 17:1552. doi: 10.3390/s17071552

World Health Organization (2019). *Musculoskeletal Conditions*. Available online at: https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions (accessed Mar 19, 2021).

Xia, P., Hu, J., and Peng, Y. (2018). EMG-based estimation of limb movement using deep learning with recurrent convolutional neural networks. *Artif. Organs*. 42, E67–E77. doi: 10.1111/aor.13004

Xie, P., Chen, X., Ma, P., Li, X., and Su, Y. (2013). "Identification method of human movement intention based on the fusion feature of EEG and EMG," in *World Congress on Engineering* (London), 1–5.

Xu, B., Zhang, L., Song, A., Wu, C., Li, W., Zhang, D., et al. (2019). Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification. *IEEE Access*. 7, 6084–6093. doi: 10.1109/ACCESS.2018.2889093

Zhai, X., Jelfs, B., Chan, R. H., and Tin, C. (2017). Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network. *Front. Neurosci*. 11:379. doi: 10.3389/fnins.2017.00379

Zhang, R., Zong, Q., Dou, L., and Zhao, X. (2019). A novel hybrid deep learning scheme for four-class motor imagery classification. *J. Neural Eng*. 16:066004. doi: 10.1088/1741-2552/ab3471

Zhao, D., Tang, F., Si, B., and Feng, X. (2019). Learning joint space "time" frequency features for EEG decoding on small labeled data. *Neural Netw*. 114, 67–77. doi: 10.1016/j.neunet.2019.02.009

Zia ur Rehman, M. Z., Waris, A., Gilani, S. O., Jochumsen, M., Niazi, I. K., Jamil, M., et al. (2018). Multiday EMG-Based classification of hand motions with deep learning techniques. *Sensors* 18:2497. doi: 10.3390/s18082497

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership