



RESEARCH ARTICLE

# A deep learning method to predict ankle joint moment during walking at different speeds with ultrasound imaging: A framework for assistive devices control

Qiang Zhang<sup>1,2\*</sup> , Natalie Fragnito<sup>1,2</sup>, Xuefeng Bao<sup>3</sup>  and Nitin Sharma<sup>1,2</sup>

<sup>1</sup>Joint Department of Biomedical Engineering, North Carolina State University, Raleigh, NC, USA

<sup>2</sup>Joint Department of Biomedical Engineering, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>3</sup>Biomedical Engineering Department, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

\*Author for correspondence: Qiang Zhang, Email: qzhang25@ncsu.edu

Received: 05 April 2022; Revised: 14 July 2022; Accepted: 06 August 2022

**Key words:** convolutional neural network; deep learning; human intent; human walking; plantarflexion; ultrasound imaging

## Abstract

Robotic assistive or rehabilitative devices are promising aids for people with neurological disorders as they help regain normative functions for both upper and lower limbs. However, it remains challenging to accurately estimate human intent or residual efforts non-invasively when using these robotic devices. In this article, we propose a deep learning approach that uses a brightness mode, that is, B-mode, of ultrasound (US) imaging from skeletal muscles to predict the ankle joint net plantarflexion moment while walking. The designed structure of customized deep convolutional neural networks (CNNs) guarantees the convergence and robustness of the deep learning approach. We investigated the influence of the US imaging's region of interest (ROI) on the net plantarflexion moment prediction performance. We also compared the CNN-based moment prediction performance utilizing B-mode US and sEMG spectrum imaging with the same ROI size. Experimental results from eight young participants walking on a treadmill at multiple speeds verified an improved accuracy by using the proposed US imaging + deep learning approach for net joint moment prediction. With the same CNN structure, compared to the prediction performance by using sEMG spectrum imaging, US imaging significantly reduced the normalized prediction root mean square error by 37.55% ( $p < .001$ ) and increased the prediction coefficient of determination by 20.13% ( $p < .001$ ). The findings show that the US imaging + deep learning approach personalizes the assessment of human joint voluntary effort, which can be incorporated with assistive or rehabilitative devices to improve clinical performance based on the assist-as-needed control strategy.

## 1. Introduction

The human ankle plantarflexor muscles play an essential role in the lower limbs' activities of daily living. For example, they provide primary mechanical propulsion force in both forward and upward directions during walking, which is referred to as the "push-off" during the late stance phase of the gait cycle (Huang et al., 2015b; Zhang et al., 2022b). Weakened function or dysfunction of human ankle plantarflexor muscles due to neurological disorders and injuries such as stroke, multiple sclerosis, and spinal cord injury, can cause significant impairment of normal walking function, like decreased plantarflexion propulsion force, inefficient ankle mechanical power production, and reduced walking efficiency

(Nuckols et al., 2020). Traditional treatment for ankle plantarflexion weakness primarily depends on physical therapy, which is labor-intensive and limited by the availability of clinical professionals (Kirshblum et al., 2007; Langhorne et al., 2011; Deng et al., 2018).

Current technological advancements to combat loss of lower limb joint function include neurorehabilitative treatments such as powered exoskeletons (Ferris et al., 2005; Jamwal et al., 2012; Huo et al., 2014; Murray et al., 2014; Jackson and Collins, 2015; Meng et al., 2016; Young and Ferris, 2016; Esquenazi et al., 2017; Zhang et al., 2017), functional electrical stimulation (Bajd et al., 1999; Mushahwar et al., 2007; Sharma et al., 2009; Kirsch et al., 2017; Alibeji et al., 2018b), soft exosuits (Bae et al., 2015; Awad et al., 2017; Siviyy et al., 2020), and hybrid neuroprosthesis (Ha et al., 2016; Alibeji et al., 2018a; Kirsch et al., 2018; Tamburella et al., 2020; Molazadeh et al., 2021). Although these advancements have gained promising progress in helping people regain or enhance the lower limb's joint functionalities, most of these approaches did not consider the wearers' volitional effort or motion intent. However, a reliable estimation of volitional effort or motion intent is essential for more intuitive and transparent control method design, like assist-as-need (AAN) control (Pehlivan et al., 2016; Shahbazi et al., 2016). The AAN control-based rehabilitation methods often require the continuous estimation of human muscle force, joint moment, or human-machine interaction through measuring or combining modalities such as neural signals of the central nervous system (CNS), muscle contraction activities, joint kinematics, and joint kinetics (Lobo-Prat et al., 2014; Zhang et al., 2020a).

Existing continuous motion intent or volitional effort prediction approaches can roughly be divided into two categories: mechanical-type and biological-type. Mechanical-type approaches usually directly measure the physical interaction between the human and machine through force or torque sensors (Veneman et al., 2007; Huang et al., 2015a; Losey et al., 2018), or measure the motion intent based on IMU sensors (Baghdadi et al., 2018; Su et al., 2019). However, the susceptibility to undesired interaction is inevitable because of the misalignment between the human joint rotation center and the sensor's fixture rotation center (Zanotto et al., 2015; Gopura et al., 2016). In addition, mechanical-type approaches cannot reveal the physiological changes *in vivo*. In contrast, biological-type non-invasive approaches, like surface electromyography (sEMG) and ultrasound (US) signals, are capable of assessing the generated neuromuscular force or joint torque. Given the fact that 30–150 ms prior to the corresponding motion is generated, neuromuscular signals can reflect human intent or volitional effort without any information delay or loss (Huo et al., 2014; Ding et al., 2016; Zhang et al., 2021). The continuous motion intent or volitional effort prediction by using neuromuscular signals usually depends on nonlinear dynamic models, like Hill-type neuromuscular model (HNM) (Lloyd and Besier, 2003; Besier et al., 2009; Sartori et al., 2012; Sartori et al., 2014; Ao et al., 2017; Dick et al., 2017; Zhang et al., 2020a, b), that builds the correlated mapping between the input neuromuscular signals and the output muscle force, joint moment, or joint motion. However, the nonlinear model consists of many physiological parameters that need to be referred from literature or determined from complex system identification, which limits the prediction accuracy and efficiency when used for assistive or rehabilitative device control.

In recent years, machine learning methods have been widely applied to achieve promising performance in building the mapping between neuromuscular signals and human joint mechanical functions, including discrete motion pattern recognition based on sEMG signals (Bitzer and Van Der Smagt, 2006; Du et al., 2010; Khokhar et al., 2010; Huang et al., 2011; Duan et al., 2015; Atzori et al., 2016; Hu et al., 2018; Simao et al., 2019; Too et al., 2019; Chen et al., 2020; Coskun et al., 2021) and US signals (Sikdar et al., 2013; Akhlaghi et al., 2016; Huang et al., 2017; McIntosh et al., 2017; Dhawan et al., 2019; Yang et al., 2019; Rabe et al., 2020), as well as continuous kinematics or kinetics prediction (also known as regression problems) based on sEMG signals (Zhang et al., 2012; Zhou et al., 2019; Zhang et al., 2020a; Wang et al., 2021; Yu et al., 2021; Zhang et al., 2022a) and US signals (Guo et al., 2013; Cunningham and Loram, 2020; Zhang et al., 2020a; Jahanandish et al., 2021; Zhang et al., 2022a). Most of the conventional machine learning methods above relied on manually selected low-dimensional explicit features from either sEMG or US signals. For example, the most common time domain and frequency domain features of sEMG signals include mean absolute value, wavelength, root mean square, zero crossing, slope sign change, mean frequency, medial frequency, mean power, and frequency ratio (Phinyomark et al. 2012),

while most common architectural and functional features of US signals include pennation angle, fascicle length, muscle thickness, fascicle orientation, tissue displacement, tissue strain, and echogenicity (Hodges et al., 2003; Arampatzis et al., 2006; Shi et al., 2008; Guo et al., 2010; Damiano et al., 2013; Strasser et al., 2013; Hodson-Tole and Lai, 2019; Sheng et al., 2019; Zhang et al., 2020b).

However, the aforementioned low-dimensional explicit features are either too presumptuous or lack meaningful information, which usually causes non-comprehensive sampling of the state-function space on the targeted joint. In addition, as mentioned by Saxby et al. (2020), the processing of US imaging data to derive the aforementioned explicit features is relatively subjective, time-consuming, and tedious. Therefore, more advanced deep learning approaches have been developed to deal with the high-dimensional explicit and/or implicit features from either sEMG or US signals for continuous joint kinematics and/or kinetics estimation (indicating human motion intent or volitional effort) (Cunningham et al., 2017b; Liu et al., 2019; Cunningham and Loram, 2020; Ma et al., 2020; Xu et al., 2020; Yu et al., 2020; Pancholi et al., 2021). For example, in Xu et al. (2020), parallel convolutional neural networks (CNNs) were used to continuously predict an individual finger's force in real-time based on the temporal energy heatmap and frequency spectrum map from high-density sEMG signals, which outperformed the prediction performance by using a motor unit decomposition method and a conventional sEMG amplitude-based method. Yu et al. (2020) proposed a stacked autoencoder-based deep neural network to continuously estimate wrist kinetics from multiple degrees of freedom with the high-density sEMG signals, where the estimation performance was better than linear regression and support vector regression methods. In Cunningham and Loram (2020), the CNNs-based deep learning approach was implemented to map individual US image frames with a contextual reference frame (prior) to absolute (drift-free) muscle states, including ankle joint angle, joint moment, and sEMG signals from plantarflexor muscles.

The majority of aforementioned contributions by using deep learning methods focused on sEMG-based upper limb kinematics or kinetics estimation problem, and few studies investigated a similar topic for lower limb functional tasks. Although sEMG has been widely implemented to non-invasively measure muscle contraction and thus estimate the volitional effort or motion intent, some inherent challenges still exist, including high noise, cross-talk between adjacent muscles, the inability to measure deep muscle, and sensitivity to the electrode position. In contrast, US imaging provides the direct visualization of the targeted muscle with a high signal-to-noise ratio and can be used to measure muscles in different depths, which would be potentially promising when used for the closed-loop control of rehabilitative or assistive devices. Therefore, the motivation of this work is to investigate the continuous joint volitional effort prediction for lower limb joint functionalities by using high-dimensional features from US imaging and a deep learning method. To the best of our knowledge, only a few recent studies have investigated deep learning approaches for continuous ankle joint kinematics, kinetics, and muscle state estimation (Cunningham et al., 2017b; Cunningham and Loram, 2020). However, they only focused on the active and passive ankle joint movement tasks at the standing posture, and no functional dynamic locomotion tasks were discussed.

In this article, we investigate the feasibility of using high-dimensional US imaging signals and a deep CNN approach to predict the continuous ankle joint net plantarflexion moment during versatile walking tasks. The focus is to verify the robustness of the proposed approach when incorporating new data from various walking scenarios for the personalized prediction of net plantarflexion moment. Unlike features extraction methods in our previous studies (Zhang et al., 2020a,b; Zhang et al., 2021; Zhang et al., 2022a, b), this deep CNN approach does not require the extraction of conventional explicit features for establishing the mapping between skeletal muscle's US imaging signals and human joint mechanical functions. The implicit features would significantly increase the dimension of the neuromuscular features, which is beneficial for bridging the mapping relationship without losing a large quantity of necessary information.

This article is organized into five sections. **Section 2** presents the basic CNN-based deep learning background, the experimental setup for treadmill walking tasks, data collection, deep CNN construction, and statistical analysis. The results of the deep CNN approach-based human ankle joint net plantarflexion moment prediction and its comparison between sEMG and US imaging are presented in **Section 3**,

followed by an interpretation of results, discussion, and potential implementations in [Section 4](#). The conclusion is given in [Section 5](#).

## **2. Methods**

### ***2.1. Basic background of CNN-based deep learning***

In the past decade, CNN has had significant innovations and results in a multitude of fields, such as pattern recognition, classification, and regression problems, from image processing to voice recognition (O'Shea and Nash, 2015). One prominent property that CNN possesses is that it can simplify the input by utilizing various kernel filters for quick processing while maintaining or even amplifying crucial information necessary for recognition. Compared to the traditional artificial neural network (ANN), CNN specializes in learning significant patterns from massive data with fewer parameters. This encouraged researchers and developers to employ CNN to solve complex tasks with larger models that are unable to be properly addressed with classic ANNs.

The design of the training network is a crucial factor in determining CNN's prediction power. The training network takes input data, interprets it in the hidden layers, and outputs the prediction value through a fully-connected output layer. The CNN output will be compared with the ground truth, and the discrepancy would drive the CNN to adjust the layer parameters to minimize the discrepancy. The filters embedded in CNN can reduce the processing time but still ensure key features of the image are maintained for accurate prediction. The layer types in CNN are usually categorized into convolution, pooling, rectified linear activation (ReLU), batch normalization, dropout, and fully connected (FC) layers. Layer types are briefly discussed below.

The kernel functions, embedded in the convolution layer, filter the pixels in the original image to get various key features/patterns to obtain a tensor-shaped input (Yamashita et al., 2018). In this way, a correlation between the output (i.e., the joint moment) and features of the image is established through CNN. The kernel may move more than one pixel per step when it scans the original image, and the step size is called stride. The selection of the stride value should consider the trade-off between the fineness of the features and the size of the input tensor, which is proportional to the complexity of the designed CNN. A padding operation may be applied within the convolution layer's parameters to alter the size of the output image by adding rows and columns of zeros on each side of the input (Yamashita et al., 2018). In this way, the data point on the borders can also be taken into consideration. The pooling layer, similar to the convolution layer, extracts significant features but reduces the size and computational power required for data processing (O'Shea and Nash, 2015). The pooling layer can be implemented as an average pooling layer or max pooling layer. Average pooling, the method used in this article, returns the average of all the pixels within the kernel. Max pooling returns the maximum value of all the pixels within the kernel. With more layers, more features are extracted at the expense of extra computational time (O'Shea and Nash, 2015). The ReLU layer replaces negative values with zeros to remove non-linearities and allow for sparsity. Sparsity reduces the time required for training the model and ensures that the neurons are more specialized. Batch normalization reduces the time of the neural network initialization. This process functions by normalizing the mean and variance of the layers, reducing the dependence of gradients, and allowing higher learning rates. A higher learning rate correlates to reduced processing time (Ioffe and Szegedy, 2015). The dropout layer, which is placed after the final ReLU layer, sets input elements to zero at random with a designated probability. The purpose of this layer is to prevent the model from overfitting. The FC is the final layer, which takes the information from the prior layer and turns it into a vector for classification or regression problems.

### ***2.2. Subjects for experimental walking study***

The study was approved by the Institutional Review Board (IRB) at North Carolina State University (Approval number: 20602). Eight young participants (5 M/3 F) without any neurological disorders or orthopedic impairments, were recruited in this study. The details of anthropometric characteristics for

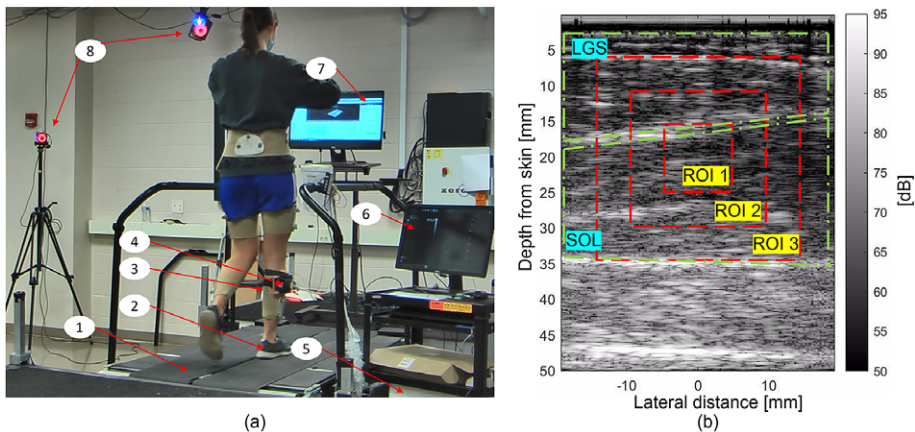
**Table 1.** Anthropometric characteristics (mean and one standard deviation [SD]) of eight young participants

Participant	Sub01	Sub02	Sub03	Sub04	Sub05	Sub06	Sub07	Sub08	Mean	SD
Height (m)	1.93	1.64	1.76	1.77	1.77	1.79	1.72	1.80	1.77	0.08
Mass (kg)	108.0	54.0	84.0	82.0	62.0	70.0	72.0	77.0	76.13	16.29
Age (Years)	24	24	30	27	22	25	26	29	25.87	2.69

each participant are shown in Table 1. Every participant was familiarized with the experimental procedures and signed an informed consent form before participating in the experiments.

### 2.3. Experimental protocol, data acquisition and pre-processing

Figure 1(a) shows the demonstration of the experimental setup. The treadmill walking data collection process involved the participants walking at speeds of 0.50, 0.75, 1.00, 1.25, and 1.50 m/s. The experiment consisted of an instrumented treadmill (Bertec Corp., Columbus, OH) with two separate belts, and each belt was mounted with a force plate (AMTI, Watertown, MA) to measure the ground reaction force (GRF) signals. Prior to the experiment, 39 retro-reflective markers were placed on the participant's lower extremities and pelvis to track the movement of each segment. The movements of retro-reflective markers were captured by 12 motion capture cameras (Vicon Motion Systems Ltd, Los Angeles, CA). To measure electrical and architectural muscle activities of plantarflexors, both sEMG sensors (SX230, Biometrics Ltd, Newport, UK) and a US transducer (L7.5SC Prodigy Probe, 6.4 MHz center frequency, S-Sharp, Taiwan) were placed on the surface of plantarflexors of the right leg. We used



**Figure 1.** Experimental setup of treadmill walking. (a) Illustration of treadmill walking experimental setup. The walking was performed at speeds of 0.50, 0.75, 1.00, 1.25, and 1.50 m/s. (1) Instrumented treadmill containing two split belts and in-ground force plates. (2) Participants' lower body with 39 retro-reflective markers attached for kinematics measurements. (3) Three sEMG channels to measure activities of LGS, MGS, and SOL muscles. (4) An ultrasound transducer for imaging of both the LGS and SOL muscles within the same plane. (5) The ultrasound imaging machine for collection of the ultra-fast radio frequency data. (6) A computer screen to show brightness mode (B-mode) US imaging. (7) Computer screen to show live markers and segment links of the participant. (8) 12 motion capture cameras to track markers' trajectories. (b) A representative B-mode US image with both LGS and SOL muscle in the same plane, as indicated within the upper and lower polygons. The lateral direction is the distance away from the US transducer longitudinal center, and the axial direction is the depth from the skin surface. Three red dashed square areas represent the three regions of interest with a size of  $100 \times 100$ ,  $200 \times 200$ , and  $300 \times 300$  pixels.

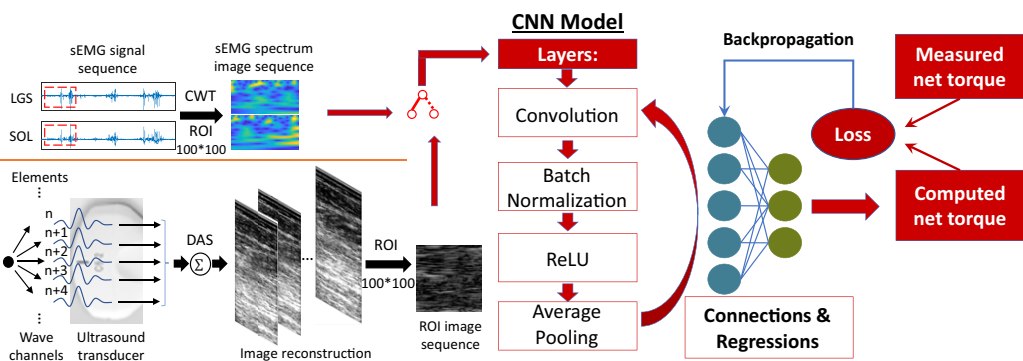


three channels to measure the sEMG signals from the shank muscles. One channel each was designated for the lateral gastrocnemius (LGS), medial gastrocnemius (MGS), and soleus (SOL) muscles. Close to the placement of sEMG sensors on LGS, the US transducer with a width of 38 mm was attached to the skin through a 3D-printed holder integrated with velcro straps. The transducer was positioned longitudinally and adjusted by observing the brightness mode (B-mode) muscle images on the US machine screen until a clear image that included both LGS and MGS muscles in the same plane was observed. An example of a representative US image in the longitudinal direction that contains both superficial LGS muscle and deep SOL muscle can be seen in [Figure 1\(b\)](#). The US imaging depth was set as 50 mm to include both muscles across all participants.

Under each walking speed, a 3-min walking trial was conducted on each participant, and data within the middle 20 s were collected for processing and analysis. The two channels GRF signals and three channels sEMG signals were sampled and collected at 1,000 Hz by using Nexus 2.9. Also, the 3D coordinates of reflective markers were synchronized and collected at 100 Hz by using Nexus 2.9. For US imaging data acquisition, we applied an ultra-fast frame rate to capture tiny deformation of the targeted muscles' architecture and facilitate the synchronization with other sensing channels. A trigger signal of 1,000 Hz was sent to the US machine to synchronize the data collection with GRF and sEMG signals, so the plane-wave raw radio frequency (RF) data of the targeted muscles were collected at 1,000 frames per second (FPS).

After data collection, pre-processing procedures were performed before the deep CNN model structure construction, as shown on the left side of [Figure 2](#). To process 3D markers' coordinates data, the gap-filling was executed in Nexus 2.9, which was followed by inputting both markers' coordinates and GRF signals to Visual3D (C-Motion, Rockville, MD) for inverse kinematics and dynamics calculation. Raw US RF data were beamformed by using the delay-and-sum (DAS) method (Lu et al., 1994; Thomenius, 1996) to generate the temporal B-mode US image sequence presented in [Figures 1\(b\)](#) and [2](#). Three regions of interest (ROIs) were selected with a size of  $100 \times 100$  (ROI1),  $200 \times 200$  (ROI2), and  $300 \times 300$  (ROI3) pixels and with the center position at the 0 mm laterally and 20 mm in depth to evaluate the net plantarflexion moment prediction performance by applying the deep CNN model.

In this study, we did not focus on the high-density sEMG signals since only three channels of sEMG signals were collected. First, sEMG signals were run through a band-pass filter between 20–450 Hz. To access the high-dimensional sEMG features, we applied a similar data processing method as mentioned by Chen et al. (2020) to create the sEMG spectrum image sequence for each channel that was



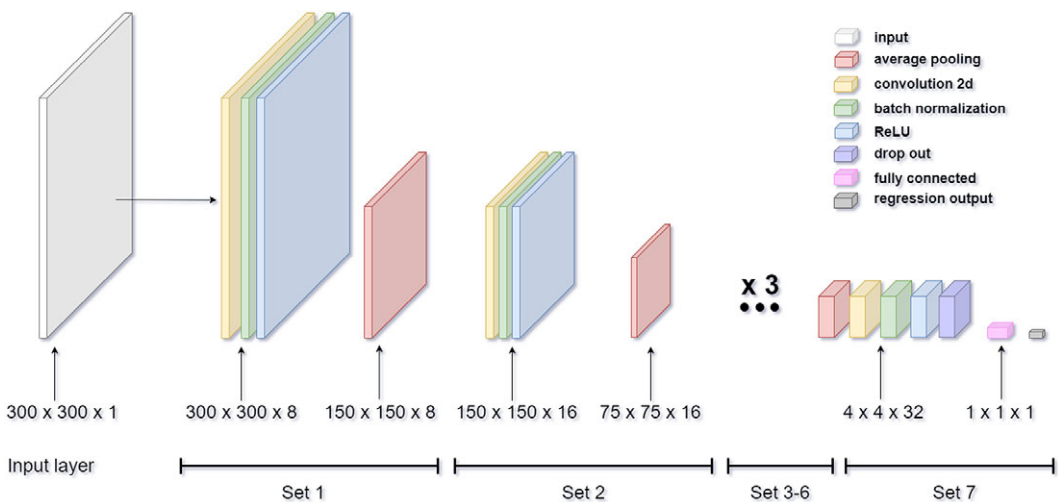
**Figure 2.** Data collection, pre-processing, and schematic illustration of the proposed deep CNN model calibration. The input to the CNN model includes either the time sequence data of cropped US imaging with different regions of interest or the time sequence data of sEMG spectrum imaging. Thirty-one layers were created in the designed CNN model, including one image input layer at the beginning, one fully connected layer, one regression output layer at the end, and seven sets of intermediate layers. Each intermediate set contained one convolution 2D layer, one batch normalization layer, one rectified linear unit (ReLU) layer, and one average pooling 2D layer.

synchronized with the US image sequence. To get the comparative ROI as the US image (take ROI1 as the example in this study), data points were segmented in series with a moving window length of 100 points ( $n-99, n-98, \dots, n$ ) and a moving step size of 1 point. The data points were then processed through a continuous wavelet transformation (CWT). After CWT, the signals were organized into scale components that contained the frequency-domain information of the time sequence sEMG data. The CWT was calculated with 50 scales obtaining a  $50 \times 100$  matrix at each time point of individual sEMG channels. To have a fair match with the US image, two  $50 \times 100$  matrices from LGS and SOL muscles sEMG channels were merged to get a  $100 \times 100$  matrix at each time instant. The compiled frames from all the data segmentation represent the sEMG spectrum image sequence.

#### 2.4. CNN construction and training

This study focused on the stance phase of the gait cycle, where the GRF signals were used to distinguish the heel-strike and toe-off instances during each walking trial with a threshold of 5% body mass. As shown in Figure 2, the input of the CNN is either a US image sequence or an sEMG spectrum image sequence within the stance phase. In the data set used for deep CNN training, a 4:1 ratio was assigned for the training and validation networks. The designed deep CNN model with 31 layers in total was created utilizing Matlab (R2020a, MathWorks, MA). In this study, a customized mini-batch datastore approach was implemented to increase the CNN training efficiency due to the out-of-memory US image data. The mini-batch size was set as 128 for both training and validation procedures on each participant's data set. At the beginning of each training, the weight parameters of the CNN layers were randomly generated. Then, the parameters were optimized and updated iteratively based on the stochastic gradient descent with a momentum optimizer (Sutskever et al., 2013) and the randomly selected mini-batch size of data points (128) from the training database. Other parameters settings are given as: a maximum number of epochs as 30, initial learning rate as 0.001 and down to 0.0001 after 20 epochs, and validation frequency as training data length/mini-batch. In addition, the training data were shuffled before each training epoch, and the validation data were shuffled before each network validation.

The deep CNN model begins with an image input layer. Then, there are seven sets of layers that contain a convolution, batch normalization, ReLU, and average pooling layer (except for set 7) in respective order, as shown in Figure 3, where an example ROI with the size of  $300 \times 300 \times 1$  (width  $\times$  height  $\times$  depth) was input to the beginning layer. The detailed settings are given below.



**Figure 3.** The architecture of the designed deep CNN for the US image and sEMG spectrum image processing. The output size of each layer is based on the input US image's ROI size of  $300 \times 300$  pixels.

The convolution layers across seven sets have different filter numbers of size  $3 \times 3$ . The filter numbers in set 1 and set 2 are selected as 8 and 16, respectively, and the remaining sets have 32 filters. Each convolution layer has a stride of [1 1] and ‘same’ padding. Each average pooling filter has a filter of size  $2 \times 2$ , a stride of [2 2], and no padding work throughout the layer sets.

In set 1, the weight and bias of the convolution layer are  $3 \times 3 \times 1 \times 8$  and  $1 \times 1 \times 8$ , respectively. The batch normalization layer has an offset and scale of  $1 \times 1 \times 8$ . The size of the layers in set 1 remains  $300 \times 300 \times 8$  until the average pooling layer where the size shifts to  $150 \times 150 \times 8$ .

In set 2, the weight and bias of the convolution layer are  $3 \times 3 \times 8 \times 16$  and  $1 \times 1 \times 16$ , respectively. The batch normalization layer has an offset and scale of  $1 \times 1 \times 16$ . The size of the layers in set 2 remains  $150 \times 150 \times 16$  until the average pooling layer where the size shifts to  $75 \times 75 \times 16$ .

In set 3, the weight and bias of the convolution layer are  $3 \times 3 \times 16 \times 32$  and  $1 \times 1 \times 32$ , respectively. The batch normalization layer has an offset and scale of  $1 \times 1 \times 32$ . The size of the layers in set 3 remains  $75 \times 75 \times 32$  until the average pooling layer where the size shifts to  $75 \times 75 \times 32$ .

In set 4, the weight and bias of the convolution layer are  $3 \times 3 \times 32 \times 32$  and  $1 \times 1 \times 32$ , respectively. The batch normalization layer has an offset and scale of  $1 \times 1 \times 32$ . The size of the layer in set 4 remains  $37 \times 37 \times 32$  until the average pooling layer where the size shifts to  $18 \times 18 \times 32$ .

In sets 5, 6, and 7, the weight and bias of the convolution layer, the offset, and the scale of the batch normalization are all same as in set 4. Finally, the output size of the ReLU layer in set 7 remains  $4 \times 4 \times 32$ . However, instead of an average pooling layer, set 7 contains a dropout layer with a dropout probability of 20% to prevent overfitting, a fully connected layer with the weight and bias of size  $1 \times 512$  and  $1 \times 1$ , and a regression output layer in respective order. The output of each layer for an exemplar US image with an ROI of  $300 \times 300 \times 1$  can be found in the Supplementary Material.

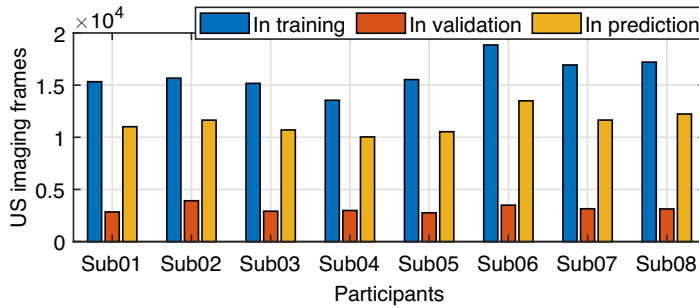
Under each walking speed, eight gait cycles were randomly selected from the 20 s of collected data. The first five stance cycles were segmented for the deep CNN model training and the last three stance cycles were for prediction analysis. Considering the high difference in plantarflexor muscles’ US imaging among different participants, we focused on the personalized CNN model instead of a generic CNN model across all participants in this study. Therefore, the trained CNN model was not intended to be universal across different participants and walking speeds, but universal across different walking speeds on the same participant. Furthermore, we applied a leave-one-speed-out cross-validation approach to avoid the overfitting issue in the training procedure. Therefore, there were five training models for each participant, and each model contained data in 20 stance cycles (5 out of 8 cycles each speed  $\times$  4 speeds) from randomly selected four speeds as the training set, data in five stance cycles (5 out of 8 cycles each speed  $\times$  1 speed) from the remaining speed as the validation set, and data in 15 stance cycles (other 3 out of 8 cycles each speed  $\times$  5 speeds) as the prediction set. The total US imaging frames that were used in training, validation, and prediction procedures varied for each participant. This is because each participant kept his/her preferred walking cadence on the treadmill at each speed, which is presented in Figure 4. Overall, around 62.5% of the selected data samples were used in the CNN model training and validation while 37.5% were used for the CNN-based prediction of each participant.

### 2.5. CNN performance evaluation and statistical analysis

To evaluate the performance of the CNN approach in terms of modeling the given US imaging data, the regression loss/quadratic loss/ $L_2$  loss, root mean square error (RMSE), and coefficient of determination ( $R^2$ ) metrics were calculated for both training and prediction procedures, which are formulated as

$$\text{Loss} = \frac{\sum_{i=1}^N (T_i - \hat{T}_i)^2}{2N}, \quad (1)$$





**Figure 4.** Individual US imaging frames in CNN training, validation, and prediction procedures across five walking speeds.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (T_i - \hat{T}_i)^2}{N}}, \tag{2}$$

$$R^2 = \frac{\left(\sum_{i=1}^N (T_i - \bar{T})(\hat{T}_i - \bar{\hat{T}})\right)^2}{\sum_{i=1}^N (T_i - \bar{T})^2 \sum_{i=1}^N (\hat{T}_i - \bar{\hat{T}})^2}, \tag{3}$$

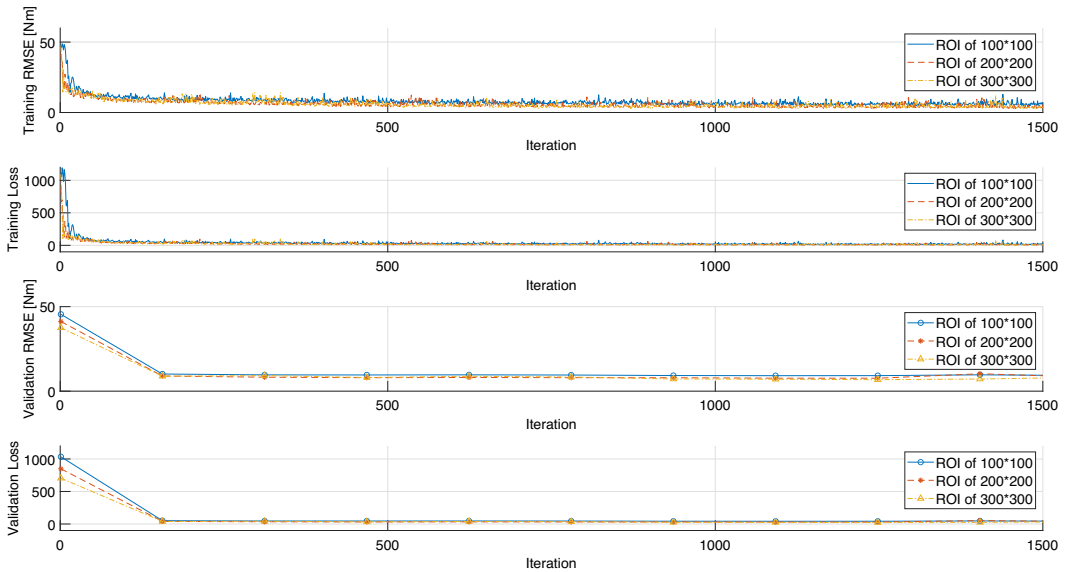
where  $N$  represents the number of training/prediction data samples,  $i$  represents the  $i$ th training/prediction sample in a data set,  $T_i$  and  $\hat{T}_i$  represent the ground truth label and CNN-based calculation for the  $i$ th training/prediction sample that is obtained from inverse dynamics, respectively.  $\bar{T}$  and  $\bar{\hat{T}}$  represent the average of the ground truth labels and the average of the CNN-based calculations, respectively. The regression loss is only concerned with the average magnitude of error irrespective of their direction, which has nice mathematical properties that make it easier to calculate gradients. However, due to squaring, calculations that are far away from actual values are penalized heavily in comparison to less deviated calculations. The RMSE serves to aggregate the magnitudes of the errors in calculations for various times into a single measure. RMSE depicts the CNN model’s accuracy and helps in comparing forecasting errors. In the prediction, RMSE values were normalized to individual peak net plantarflexion moment from ground truth labels, noted as N-RMSE.

The normality of the prediction N-RMSE,  $R^2$ , and time cost per image data sets across all participants was tested based on the Shapiro–Wilk parametric hypothesis test (SW test). According to the results from the SW test, a one-way repeated-measure analysis of variance (ANOVA for normal distribution) or Friedman’s tests (for not normal distribution) followed by Tukey’s honestly significant difference tests (Tukey’s HSD) was applied to evaluate the CNN-based prediction performance across all speeds but with different ROIs. Similarly, to evaluate if there was a significant difference between deep CNN-based predictions with US imaging data and sEMG spectrum imaging data with the same ROI size, a paired  $t$ -test (normal distribution) or a Wilcoxon signed rank test (not normal distribution) was applied. The significant difference level was chosen as  $p < .05$  for all statistical tests.

### 3. Results

#### 3.1. Progress during CNN training

With the designed CNN layer number and structure of each layer, the training procedures with different US image’s ROIs converged to a small loss threshold for each participant. By taking Participant Sub08 as

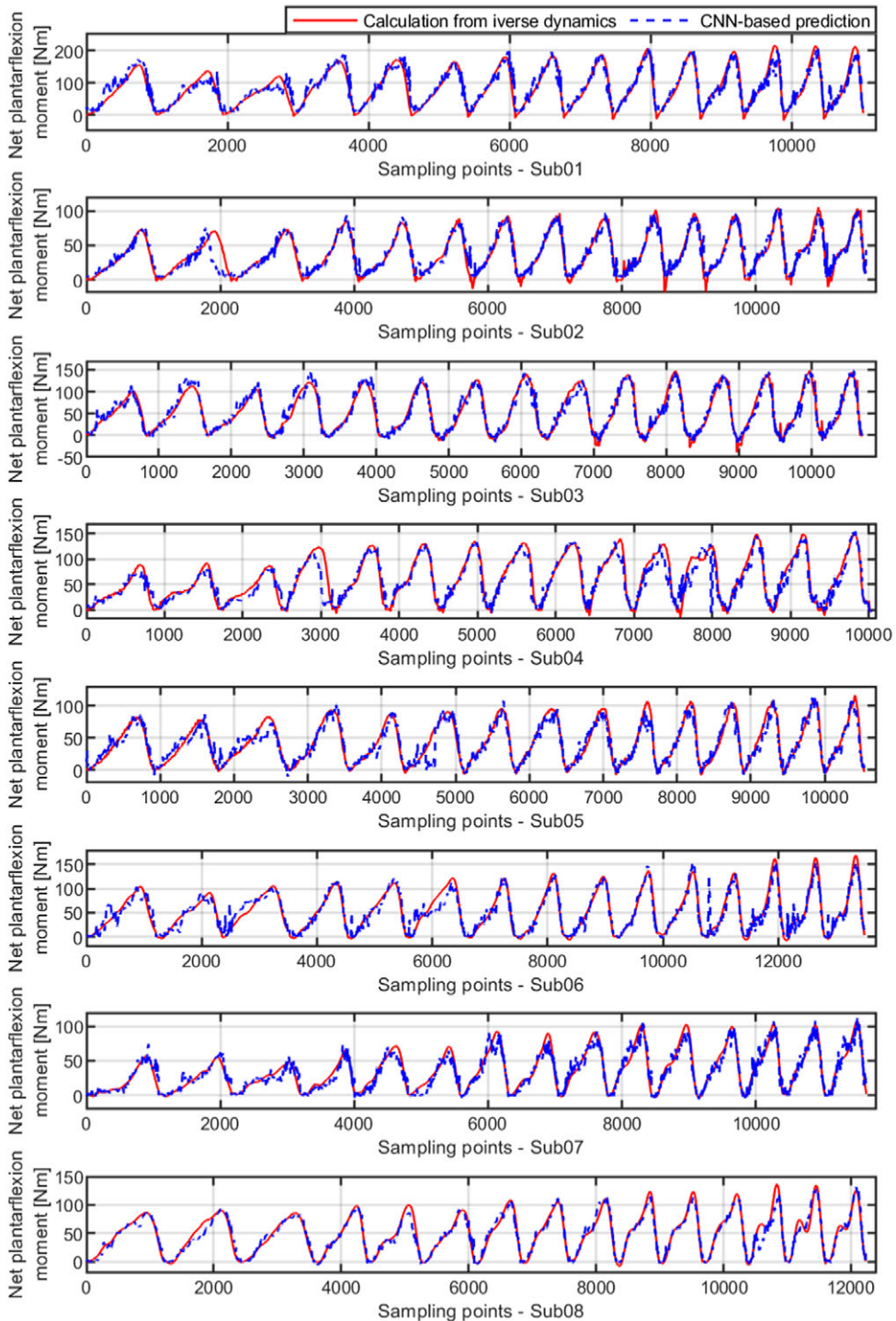


**Figure 5.** The convergence performance of US imaging-based net plantarflexion moment RMSE and loss with the increase of iteration number during the CNN training and validation procedures. The data set shown here is from Participant Sub08.

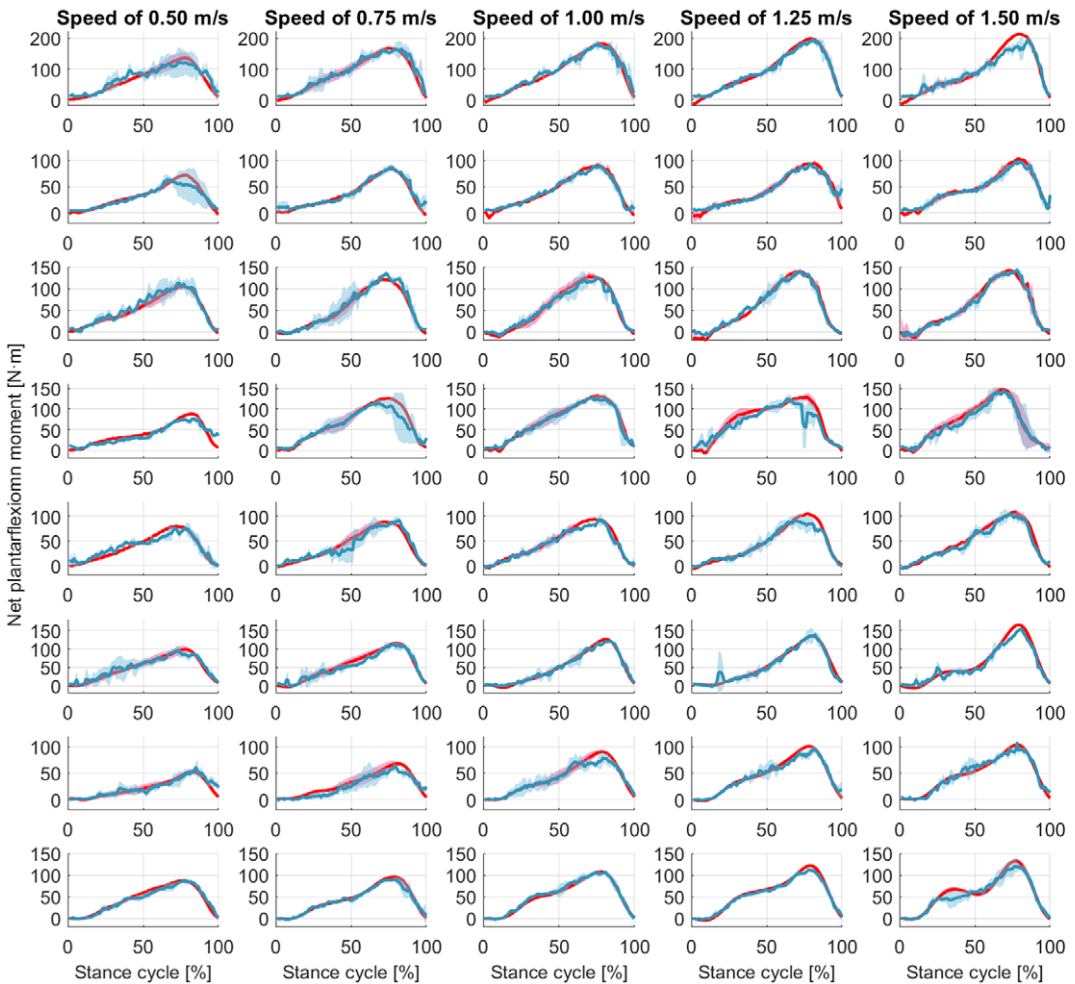
an example, the upper two plots and lower two plots in Figure 5 show the loss and RMSE convergence results during the CNN training and validation procedures. It can be observed that after around 300 iterations, the training loss reached a very small value and it did not decrease significantly even though the iteration number increased. It is also observed that a larger ROI resulted in faster convergence speed as well as a smaller steady threshold. Ultimately, the validation RMSE values converged to 13.75, 12.93, and 11.96 Nm, with the ROIs of  $100 \times 100$ ,  $200 \times 200$ , and  $300 \times 300$  pixels, respectively. These results indicate that the larger ROI would be beneficial for better net plantarflexion moment prediction performance. Correspondingly, with the same US imaging frame number, the time cost for the CNN training varies significantly due to the selection of different ROIs. With the aforementioned US imaging frame numbers, as shown in Figure 4, across eight participants, the CNN training times are  $3.82 \pm 1.30$  min,  $8.47 \pm 2.46$  min, and  $17.70 \pm 3.77$  min for ROIs of  $100 \times 100$ ,  $200 \times 200$ , and  $300 \times 300$  pixels, respectively.

### 3.2. Personalized CNN-based net plantarflexion moment prediction performance

Once the personalized CNN model was trained with data sets from all five walking speeds for each participant, new data sets from each walking speed not involved in the training procedure were used to verify the effectiveness and robustness of the personalized CNN model, known as the inter-speed prediction performance. Figure 6 shows the continuous CNN-based prediction with the ROI of  $100 \times 100$  pixels and the ground truth from inverse dynamics individually, represented by blue dashed and red solid curves, respectively. From left to right, every three ramp-up-down curves are from one walking speed out of the five. As the walking speed increased, the peak net plantarflexion moment increased, but the sampling points that were used in the prediction decreased. Therefore, the curves for both ground truth and CNN-based prediction become denser temporally. After separating the prediction performance under each walking speed and normalizing the sampling points throughout each walking stance cycle, the net plantarflexion moment prediction performance corresponding to Figure 6 is demonstrated in Figure 7, where the red and blue center curves and shadowed areas represent the mean and standard deviation values (three stance cycles for each curve) of the ground truth and CNN model-

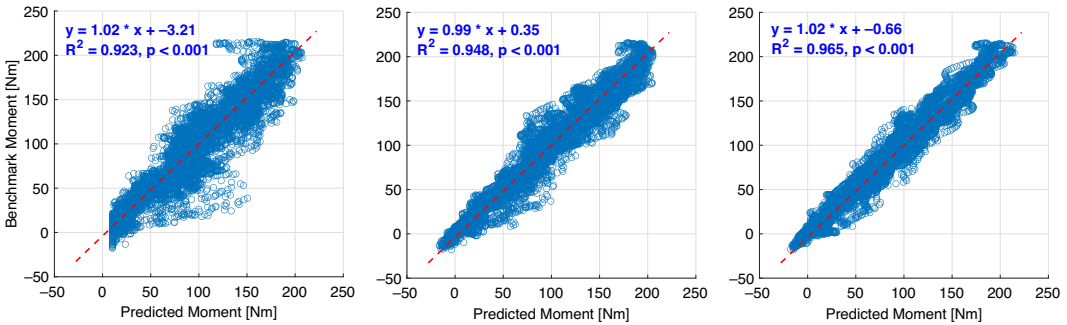


**Figure 6.** Ankle joint net plantarflexion moment prediction time sequence on each participant by using US images with ROI of  $100 \times 100$  pixels and the deep learning approach. The red solid and blue dashed curves represent the measurements from inverse dynamics and prediction from the CNN model. For each walking speed, three walking stance cycles are included for prediction, therefore, 15 periodic curves are shown for each participant (with the speed order of 0.50, 0.75, 1.00, 1.25, and 1.50 m/s).



**Figure 7.** Ankle joint net plantarflexion moment prediction as a percentage of the stance cycle (0% for heel-strike and 100% for toe-off) by using US images with ROI of  $100 \times 100$  pixels and the deep learning approach. The red and blue center curves and shadowed areas represent the mean and standard deviation values (three stance cycles for each curve) of the ground truth and CNN model-based prediction, respectively. Each row subplots represent data from individual participant while each column subplots represent individual walking speed out of five.

based prediction, respectively. Each row subplots represent data from individual participant while each column subplots represent individual walking speed out of five. Taking results from Sub01 as an example, with the three ROIs, the scatter plots between the ground truth and the CNN-based prediction are shown in Figure 8, where the red dashed line on each plot represents the 45-degree ( $y = x$ ) line. It is observed that all data points are distributed along the 45-degree line, indicating a highly linear relationship between the CNN-based prediction and ground truth. A linear regression model was conducted for each ROI, and the coefficients of slope and y-intercept are labeled on each plot. Promisingly, all slope values are very close to 1 ( $p < .001$ ) while all y-intercept values are very close to 0 ( $p < .01$ ). The  $R^2$  values between the prediction and ground truth are 0.923, 0.948, and 0.965, respectively. Similar results are also observed from data analysis on other participants across multiple walking speeds.



**Figure 8.** Scatter plots between the net plantarflexion moment benchmark and CNN-based prediction from Sub01 by using ROIs of 100 × 100 (left), 200 × 200 (middle), and 300 × 300 (right) pixels.

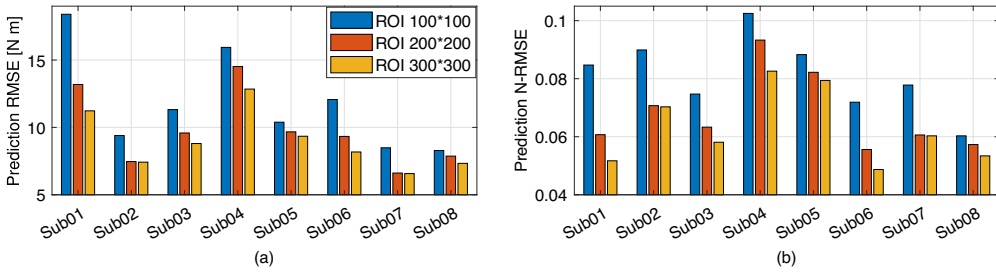
**Table 2.** Results of linear regression analysis between net plantarflexion moment CNN-based prediction and ground truth with different ROIs, including mean, standard error (SE), and *p*-value of slope and *y*-intercept coefficients, as well as *R*<sup>2</sup> values

Participants	ROI	Slope			<i>y</i> -intercept			<i>R</i> <sup>2</sup>
		Estimate	SE	<i>p</i> -value	Estimate	SE	<i>p</i> -value	
Sub01	1	1.021	0.003	<0.001	-3.214	0.316	<0.001	0.923
	2	0.994	0.002	<0.001	0.347	0.224	0.240	0.948
	3	1.022	0.002	<0.001	-0.659	0.147	<0.001	0.965
Sub02	1	1.011	0.003	<0.001	0.321	0.132	0.043	0.913
	2	1.021	0.003	<0.001	0.530	0.129	0.051	0.934
	3	1.016	0.003	<0.001	0.158	0.159	0.628	0.947
Sub03	1	0.992	0.002	<0.001	0.409	0.182	0.227	0.926
	2	0.987	0.002	<0.001	1.225	0.120	<0.001	0.949
	3	0.993	0.002	<0.001	1.198	0.139	<0.001	0.957
Sub04	1	1.001	0.004	<0.001	3.524	0.290	<0.001	0.835
	2	0.993	0.003	<0.001	5.627	0.211	<0.001	0.868
	3	0.988	0.003	<0.001	3.223	0.284	<0.001	0.896
Sub05	1	1.021	0.003	<0.001	0.269	0.142	0.111	0.855
	2	1.012	0.003	<0.001	1.153	0.165	<0.001	0.908
	3	1.036	0.003	<0.001	1.182	0.175	<0.001	0.921
Sub06	1	1.009	0.003	<0.001	-0.395	0.132	0.571	0.870
	2	1.012	0.003	<0.001	2.647	0.144	<0.001	0.936
	3	0.998	0.002	<0.001	2.157	0.171	<0.001	0.943
Sub07	1	0.986	0.003	<0.001	1.797	0.155	<0.001	0.883
	2	0.964	0.002	<0.001	2.216	0.069	<0.001	0.911
	3	0.993	0.002	<0.001	2.783	0.076	<0.001	0.931
Sub08	1	1.008	0.002	<0.001	0.469	0.120	0.005	0.928
	2	1.019	0.002	<0.001	0.674	0.112	<0.001	0.936
	3	0.993	0.002	<0.001	0.643	0.176	0.002	0.942

Table 2 summarizes the linear regression results between the CNN-based net plantarflexion moment prediction and the ground truth from each participant with different ROIs. All standard errors of the slope coefficients are less than .005 with the *p*-value from the *F*-statistic less than 0.001. All standard errors of the *y*-intercept coefficients are less than 0.32 but with a higher variation of *p*-value from the *F*-statistic. The goodness of the linear regression is verified with pretty high *R*<sup>2</sup> values (0.835–0.965) for all participants. Except for Sub04 with all three ROIs, Sub05, Sub06, and Sub07 with ROI1, all *R*<sup>2</sup> values between the CNN-based prediction and ground truth are higher than 0.9.

For each ROI, one prediction RMSE value was calculated across 15 stance cycles used in the prediction procedure on each participant to quantitatively assess the accuracy and robustness of the CNN-based net plantarflexion moment prediction by using B-mode US imaging regardless of walking speeds, and the results are presented in Figure 9(a). It is observed that, for the same participant, the increase of the ROI size



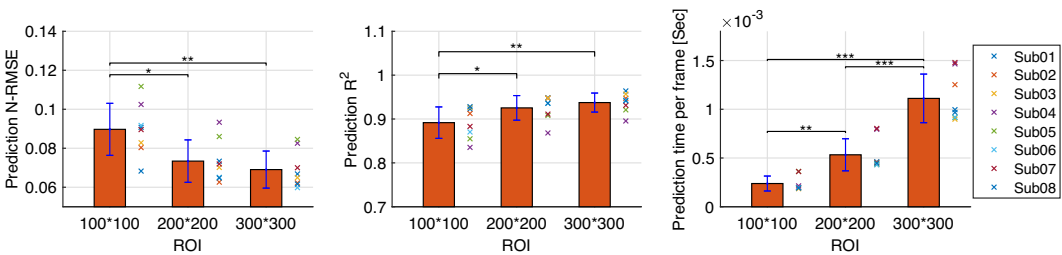


**Figure 9.** The individual net plantarflexion moment prediction RMSE and N-RMSE values of 15 stance cycles across five walking speeds by using the trained personalized CNN model with different ROIs.

can moderately decrease the prediction RMSE, although the decrease is minor between ROI  $200 \times 200$  and ROI  $300 \times 300$  on Sub02 and Sub07. However, due to the high variance of subjective peak net plantarflexion moment, using the CNN-based prediction RMSE values to directly evaluate the inter-subject performance is not feasible. Therefore, the relative prediction error, defined as the ratio between the prediction RMSE value and individual peak moment (N-RMSE), was calculated and shown in Figure 9(b). Notably, among all inter-speed predictions, 95.8% (23 out of 24) cases exhibited less than 10% N-RMSE values regardless of the ROI size.

**3.3. Influence of US imaging ROI size on CNN-based prediction**

The inter-subject analysis of the US imaging ROI size’s effect on the CNN-based net plantarflexion moment prediction is detailed in this subsection, including the prediction N-RMSE, prediction  $R^2$ , and the time cost per US image during prediction. Figure 10 depicts the inter-subject results of the above evaluation metrics, where each point on the right side of each bar represents a data sample from each participant. Results from the SW test showed that every data group followed a normal distribution with different ROIs. All prediction N-RMSE values are less than 11.20% with these three ROIs, indicating a successful prediction performance by using the deep learning approach. Compared to ROI1, ROI2 and ROI3 significantly reduced the N-RMSE values by 18.17% ( $p = .024$ ) and 22.97% ( $p = .004$ ), respectively. However, there was no significant difference between ROI2 and ROI3 ( $p = .727$ ). As mentioned before, all  $R^2$  values are higher than 0.9 except for Sub04 with all three ROIs, Sub05, Sub06, and Sub07 with ROI1. Compared to ROI1, ROI2 and ROI3 significantly increased the  $R^2$  values by 3.78% ( $p = .042$ ) and 5.14% ( $p = .009$ ), respectively. However, no significant difference was observed between ROI2 and ROI3 ( $p = .687$ ). Except for the prediction accuracy and robustness, the prediction time cost per US

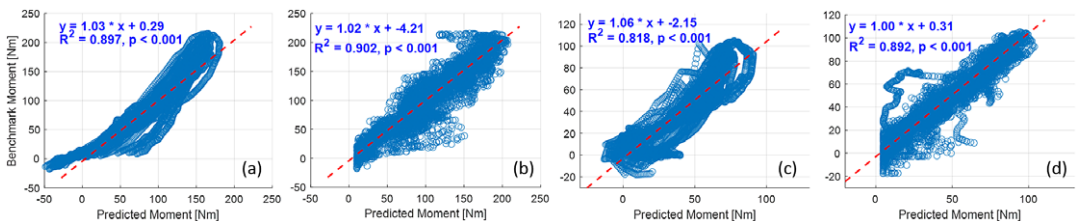


**Figure 10.** CNN-based net plantarflexion moment prediction results summary across eight participants. Left – Prediction RMSE values normalized to corresponding peak plantarflexion moment, Middle –  $R^2$  values between net plantarflexion moment prediction and ground truth observation from inverse dynamics, Right – Prediction time cost for each US image frame. Asterisks \*, \*\*, and \*\*\* represent the statistically significant difference levels at  $p < .05$ ,  $p < .01$ , and  $p < .001$ , respectively.

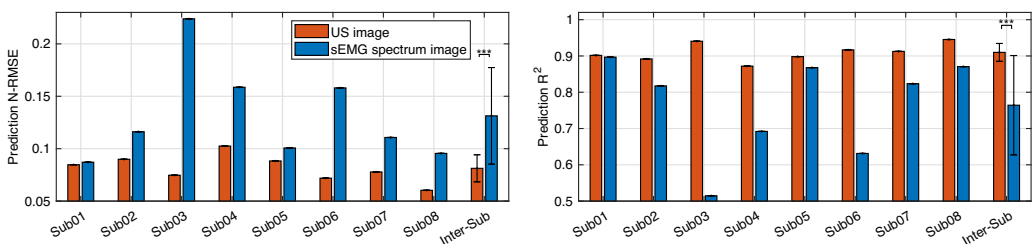
imaging frame is essential for implementing the real-time control of robotic rehabilitative or assistive devices. Results in Figure 10 show that the time cost per frame is less than 1.5 ms regardless of the ROI size, which is potentially fast enough for closed-loop control of robotic devices. Furthermore, compared to ROI1, ROI2, and ROI3 significantly increased the prediction time cost per image frame by 223.80% ( $p = .009$ ) and 461.97% ( $p < .001$ ), respectively. In addition, ROI3 significantly increased the time cost by 206.42% ( $p < .001$ ) when compared to ROI2. The results showed that the prediction efficiency was more sensitive to the ROI size than the prediction accuracy.

**3.4. Prediction performance comparison between US imaging and sEMG**

The deep CNN model training and prediction procedures concerning the sEMG spectrum image sequence were similar to the US image sequence with the ROI of  $100 \times 100$  pixels. Similar to the results presented in Figure 8, the comparative scatter plots between the benchmark and predicted net plantarflexion moment values from two representative participants are shown in Figure 11. We found that with the same constructed CNN architecture, the net plantarflexion moment prediction performance is superior to using the US image than the sEMG spectrum image with the same ROI. The higher  $R^2$  values on these two participants when using the US image indicate a stronger linear relationship between the net moment benchmark and CNN-based net moment prediction. The individual and inter-subject prediction N-RMSE and  $R^2$  values are presented in Figure 12. Except for Sub01, results from all other participants consistently support the hypothesis that US image + CNN significantly outperforms sEMG + CNN regarding net plantarflexion moment prediction across versatile walking speeds. Across eight participants, statistical analysis shows that with the same ROI and the same CNN architecture, US images significantly reduced the prediction N-RMSE values by 37.55% ( $p < .001$ ) and increased the prediction  $R^2$  values by 20.13% ( $p < .001$ ) when compared to sEMG spectrum images.



**Figure 11.** Comparative results of scatter plots between the net plantarflexion moment benchmark and CNN-based prediction: (a) with sEMG spectrum image on Sub01; (b) with US image on Sub01; (c) with sEMG spectrum image on Sub02; (d) with US image on Sub02.



**Figure 12.** Comparative results of the net plantarflexion moment prediction N-RMSE and  $R^2$  values by using the proposed deep CNN architecture and the same ROI size US image and sEMG spectrum image.

## 4. Discussions

During the past decade, deep learning and CNNs have revolutionized many fields of machine learning, including speech/language recognition, computer vision, biological disease diagnosis, and robotics. Thus, there is no doubt that they may improve the analysis of medical imaging and contribute to incorporating medical imaging into the closed-loop control of biomedical robotic devices. From the perspective of rehabilitation engineering, deep learning and CNNs will be beneficial to bridge the gap between the rehabilitative or assistive device market (that requires intuitive and transparent control methods based on human motion intent or volitional efforts prediction) and recent scientific research results in exoskeletons and prostheses (that show that sEMG-based or US imaging-based proportional control is possible).

### 4.1. Results interpretation

In this article, the investigation tested the hypothesis that the joint net moment is encoded instantaneously with the 3D muscle collagenous structure and is observable using 2D B-mode US imaging. We used the ultra-fast US imaging data collection to generate 26,570–35,838 US images from each of eight participants (in a total of 244,753 images) containing dynamic architectural changes of LGS and SOL muscle during walking at five different speeds. We introduced a framework for the application of CNN-based deep learning to the continuous prediction of ankle joint net moment during treadmill walking at multiple speeds by using those recorded US imaging, and we evaluated the performance of using different sizes of ROI and compared the results with sEMG spectrum imaging with the same size of ROI. When the proposed US imaging + deep learning approach was used, ROI1, ROI2, and ROI3 displayed a prediction RMSE range of 8.48–19.71 Nm ( $13.04 \pm 3.85$  Nm), 6.61–15.94 Nm ( $10.68 \pm 3.17$  Nm), and 6.57–14.52 ( $9.98 \pm 2.63$  Nm), respectively, a prediction N-RMSE range of 0.068–0.112 ( $0.089 \pm 0.013$ ), 0.063–0.093 ( $0.073 \pm 0.011$ ), and 0.059–0.084 ( $0.069 \pm 0.009$ ), respectively, and a prediction  $R^2$  range of 0.835–0.928 ( $0.892 \pm 0.035$ ), 0.868–0.949 ( $0.925 \pm 0.028$ ), and 0.896–0.965 ( $0.938 \pm 0.022$ ), respectively. The results present that the joint net moment of human ankle plantarflexion can be well predicted by the B-mode US imaging signals when people do treadmill walking at different speeds. The joint moment prediction by using the deep CNN model illustrates a good performance across different sizes of US imaging's ROI, which implies the size of ROI does not play an essential role in the net plantarflexion moment prediction performance, and the deep CNN model is relatively robust among different ROIs, thus we believe that it is a good avenue to explore.

A close look at the net plantarflexion moment prediction performance in [Figure 6](#) presents the subjective smoothness of the deep CNN-based prediction curves. Since an ultra-fast frame rate was employed to collect raw US imaging RF data, tiny muscle deformation was captured across frames, but in the meanwhile, the resolution of each image would be reduced (more noise of each frame) given the center frequency was fixed at 6.4 MHz. One advancement of using deep CNN is that no pre-defined explicit features from US images are required, which usually depends on the high resolution of US images. Therefore, moderate noise in US images due to the high collection frame rate did not impose a noticeable influence on the prediction performance, including timings and peak values. The smoothness issue shown in [Figure 6](#) could be easily addressed by adding a low-pass filter for real-time implementations. One point we would like to make is that the high gait-to-gait and person-to-person variations make it very challenging to build a generalized deep CNN model for all participants. Therefore, in this study, a personalized deep CNN model was trained for each participant and the individual CNN model presented robustness across different walking speeds. The inter-subject analysis in the previous section was based on the individual deep CNN models.

In deep CNN model training, the training accuracy was not sensitive to the increase of the ROI, but the training time was sensitive to the increase of the ROI. When the size of ROI increased from  $100 \times 100$  to  $300 \times 300$  pixels, the average training time increased around 4.6 times (from 3.82 to 17.70 min), although the increase is not linearly related. Impressively, with the deep learning architecture and B-mode US imaging signals, even a fairly small ROI could generate relatively high net moment prediction accuracy

**Table 3.** The comparison results of the normalized prediction N-RMSE values among different studies

Approaches	HNM (Zhang et al., 2022b)	SVR (Zhang et al., 2022a)	FFNN (Zhang et al., 2022a)
N-RMSE [%]	9.48 ± 1.51	8.49 ± 2.74	10.13 ± 3.44
R <sup>2</sup>	0.91 ± 0.04	0.93 ± 0.05	0.89 ± 0.08
Approaches	CNN-ROI1	CNN-ROI2	CNN-ROI3
N-RMSE [%]	8.97 ± 1.33	7.34 ± 1.09	6.91 ± 0.95
R <sup>2</sup>	0.89 ± 0.03	0.93 ± 0.03	0.94 ± 0.02

Note. Reported data are the mean ± standard deviation across conditions and participants, where previous studies applied data fusion between sEMG and US imaging signals.

Abbreviation: HNM, hill-type neuromuscular mode; SVR, support vector regression; FFNN, feedforward neural network.

(with an average N-RMSE of 8.97%). Regarding the overall accuracy, in this article, the comparison results between US images and sEMG spectrum images with the same size ROI and same deep CNN structure demonstrated the superior net moment prediction performance when B-mode US images were applied. In comparison to existing literature, it is fundamental to note that the results should be compared only with analyses considering the continuous prediction/estimation of joint moment/torque or joint angle by using neuromuscular signals, and the error should be normalized to the corresponding peak value. By using machine learning methods, it is common to see in the literature estimation/prediction error of down to 7–15% (Shi et al., 2008; Dick et al., 2017; Huang et al., 2017; Jahanandish et al., 2019; Zhou et al., 2019; Ma et al., 2020; Jahanandish et al., 2021; Wang et al., 2021; Yu et al., 2021; Zhang et al., 2022a,2022b). However, most of these studies consider using either sEMG or US signals for the upper limb motion or moment estimation/prediction, while a few of them consider the lower limb applications but with relatively simple tasks. Thus, an investigation of complex lower limb tasks, like walking at different speeds, would be reasonable and necessary to justify the proposed B-mode US image + deep CNN approach.

In terms of the ankle joint net plantarflexion moment prediction accuracy, we compared the N-RMSE values between the current study and these established methods in our previous studies (as shown in Table 3), which indicates the superior net plantarflexion moment prediction accuracy by using the proposed US image + deep CNN approach. Compared to the traditional machine learning methods, the deep CNN method achieved many improvements, such as estimation/prediction accuracy, dynamic characteristic, robustness, less pre-processing effort, and more richness of features. In addition, as shown in Figure 10, the time spent per frame during prediction was less than 1.5 ms, which would easily satisfy the real-time application requirement. Compared to the deep CNN-based prediction time, the processing time to obtain B-mode US images online is more significant. It takes around 20–30 ms to generate a B-mode US image based on the build-in online beamforming algorithm, which theoretically indicates the current prediction time is nearly negligible in real-time. As an ongoing project, the real-time implementation of the deep CNN-based net moment prediction is under development.

#### 4.2. Scientific and clinical significance

The ultimate objective of using the deep CNN-based joint net moment prediction is to design more intuitive and efficient control approaches for wearable assistive robotic devices when considering human volitional contraction efforts. One typical control approach for lower-limb robotic devices to assist human subjects is to measure interaction dynamics or estimate the joint moment from inverse dynamics. However, the nature of interaction dynamics and inverse dynamics may not be appropriate in some cases, as the user must produce a certain torque on the joints to initiate the motion before the devices can generate assistance. If the users are not able to produce sufficient torques on their joints, the robotic devices may not be successfully controlled. Fortunately, this disadvantage could be avoided by using human intent-based control of robotic devices according to biological signals, like sEMG, that are sent from the CNS to the functional motor units. One of the main advantages of using sEMG is the small time difference (between 30–150 ms in Zhou et al. (1995), Lloyd and Besier (2003), and Blackburn et al.

(2009)) between signal generation and motion execution, which enables the human intent-based control of wearable robotic devices (the control command generation of the robotic devices advances the generation of the human joint moment or limb motion). Therefore, sEMG signals have been successfully applied in robotic devices control in the past, like examples in Kawamoto and Sankai (2002), Fleischer and Hommel (2008), Lenzi et al. (2012), and Peternel et al. (2016). The crucial perspective of sEMG-based control approaches is that even if the user is not capable of producing sufficient joint motion or joint moment, the motion intent of the human user can still be detected and then the wearable robotic devices can be controlled.

Considering the potential limitations of using sEMG as mentioned in the introduction section, we proposed to use the alternative modality, that is, US imaging, to estimate/predict human volitional/residual joint torque. In our previous study Zhang et al. (2021), we assessed that US imaging-derived signals of the tibialis anterior muscle preceded the ankle joint motion around 45–85 ms. Similarly, the experimental observations from Begovic et al. (2014) showed that the time delay between the onset of muscle fiber motion and the muscle contraction force generation was  $49.7 \pm 7.0$  ms of the quadriceps femoris muscle. Therefore, the time latency between the onset of US imaging-derived signals and the onset of joint torque/motion indicates the comparative joint torque/motion prediction potentials as using sEMG signals. While the CNNs have been applied to many fields as mentioned at the beginning of this section, their application to B-mode US image data from skeletal muscles for a purpose of biomechanics analysis is relatively novel. From the perspective of gait kinematics or kinetics analysis, the standard way to measure the net joint moment depends on the measurement of 3D coordinates of reflective markers via a motion capture system, GRF signals, and inverse dynamics calculations. However, this traditional way encounters several shortcomings: 1) the limitation of instrumental setup in a lab environment and not applicable for field testing, 2) the effort-consuming data collection and post-processing, 3) the challenge of online data acquisition after inverse dynamics, 4) no accessibility of muscle activities during functional tasks. The proposed US image + deep CNN approach provides a forward dynamics approach to estimate/predict net joint moment that has more flexibility in terms of the experimental environment, data collection and analysis, and real-time implementation. Furthermore, the US image + deep CNN approach opens a new gate for sensing volitional user intent and extracting proportional control signals for rehabilitative or assistive robotic devices, including powered ankle-foot prosthesis or ankle exoskeleton, in response to muscle deformations in the shank.

In terms of the control of the powered ankle-foot prosthesis, the net ankle joint moment predicted by using the US image + deep CNN approach could be used to modulate the impedance parameters for the prosthesis, mainly including stiffness and damping coefficients, to imitate the biological ankle joint impedance during the walking stance phase as mentioned in Rouse et al. (2014) and Lee et al. (2016). This has the potential to benefit people with below-knee amputation walking on a treadmill or level ground, such as the increase of gait symmetry, the realization of natural gait patterns, improvement of comfort, and reduction of metabolic cost. In terms of the control of the powered ankle exoskeleton, the net ankle joint moment prediction could be used to adjust the personalized assistance torque profile parameters from the exoskeleton during the walking stance phase, including timings and amplitudes, known as biological mechanism-based control (Nuckols et al., 2021) or proportional sonomyographic control (Dhawan et al., 2019). This manipulation aims to reproduce an individualized assistance torque on the ankle joint that is proportional to the biological torque prediction from the US image + deep CNN approach, as so to enhance the walking function for able-bodied persons, reduce human energy expenditure, or restore normative gait pattern for persons with weakened ankle plantarflexion functions cause by stroke, spinal cord injury, or multiple sclerosis.

Since we focus on the most common and relative complex locomotion tasks, that is, treadmill walking with multiple speeds, and the ankle joint plantarflexion function is essential to provide “push-off” effort during walking, both LGS and SOL muscles were investigated to understand the mapping between the B-mode US images and net ankle joint moment. However, the significance of the current investigation lies in the potential application to estimate/predict other joints’ mechanical functions, as



well as to estimate/predict states generally from individual muscles, given the fact that the information is well encoded in skeletal muscles' collagen structure and are observable by using US imaging (Cunningham and Loram, 2020). Recent research studies have also applied US imaging + deep (machine) learning to estimate skeletal muscles' activation levels (Cunningham et al., 2017b; Cunningham and Loram, 2020; Feigin et al., 2020), fascicle length (Rosa et al., 2021), fascicle orientation (Cunningham et al., 2017a), and muscle segmentation (Carneiro and Nascimento, 2013; Zhou et al., 2020).

### 4.3. Limitation and future work

Although the proposed US image + deep CNN approach showed promising performance in terms of ankle joint net moment prediction accuracy and time efficiency, some limitations still exist in the current study. First of all, raw US imaging RF data beamforming to B-mode US images, net architecture, and hyperparameter settings seem to be fundamental for the analysis of US image data with deep CNNs, since they can strongly change the final prediction accuracy in the testing image set, and time to converge. However, the aforementioned aspects were pre-selected and no comparison studies between different beamforming algorithms, different net architectures, or different hyperparameter settings were conducted since the focus of the current study was to investigate the feasibility and effectiveness of the US image + deep CNN approach instead of a systematic analysis of various deep CNN models. Secondly, the ultra-fast frame rate was applied to collect the raw US imaging RF data, but B-mode US images at 1,000 FPS were generated offline by using the DAS beamforming method for both deep CNN model training and prediction. However, this ultra-fast frame rate is likely to be down-sampled to 1/30–1/20 for real-time implementation according to the previous discussion. Thirdly, only the plantarflexor muscles were involved in the net plantarflexion moment prediction without considering the effects of antagonistic muscle groups. Lastly, only two sEMG channels were used to collect signals from LGS and SOL muscles and generate spectrum images for comparison with US images. It is hard to specify if the inferior prediction performance of using sEMG spectrum images came from the small number of channels and the noisy nature of sEMG signals.

Our future work will focus on 1) the effects of various hyperparameters in the deep CNN model on the US images-based ankle joint net moment prediction performance; 2) the real-time implementation of the proposed US image + deep CNN approach for ankle joint net moment prediction during treadmill walking at multiple speeds on both neurologically-unimpaired persons and incomplete spinal cord injury/stroke patients; 3) the incorporation of the real-time ankle joint net moment with closed-loop assistance-as-needed control for a cable-driven ankle exoskeleton as mentioned in Zhang (2021) on both subject groups.

## 5. Conclusion

This article proposed the first comprehensive analysis of a novel computational method for estimating human walking volitional effort on the ankle joint using US images of skeleton muscles. Specifically, we investigated the effectiveness and robustness of the B-mode US imaging + deep CNN approach for the ankle joint net moment prediction during treadmill walking at multiple speeds. We evaluated the effects of the US image's ROI size on the net moment prediction performance and compared the deep CNN-based prediction performance by using B-mode US image data and sEMG spectrum image data with the same ROI size. Results from eight participants without any neurological disorders tested the hypothesis that the joint net moment was encoded instantaneously with the 3D muscle collagenous structure and was observable using 2D B-mode US images. The proposed framework achieved a relatively high net moment prediction accuracy (with an average N-RMSE of 8.97%) even by using a small ROI of  $100 \times 100$  pixels. The prediction time per image was less than 1.5 ms, which could easily satisfy the real-time implementations when incorporated with the closed-loop control of robotic devices. With the same deep CNN structure, the comparison results showed that US images significantly reduced the prediction N-RMSE

values by 37.55% ( $p < .001$ ) and increased the prediction  $R^2$  values by 20.13% ( $p < .001$ ), when compared to sEMG spectrum images. Impressively, the proposed US image + deep CNN approach produced accurate results comparable or superior to the average classical machine learning methods, which suggests that further studies may lead to the improvement of the overall field of US imaging-driven control of the ankle-foot prosthesis or ankle exoskeleton.

**Acknowledgments.** We would like to thank Alison Myers and Ashwin Iyer for their help in the experiments' conduction and data collection. We would also like to thank all participants for their involvement in the study.

**Supplementary Material.** To view supplementary material for this article, please visit <http://doi.org/10.1017/wtc.2022.18>.

**Data Availability Statement.** Data availability is not applicable to this article as no new data were created or analyzed in this study.

**Funding Statement.** This research was supported by NSF Career Award (grant number 2002261).

**Ethical Standards.** The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The study was approved by the Institutional Review Board of North Carolina State University (protocol number 20602 and date of approval 02/10/2020).

**Competing Interests.** The authors declare no competing interests exist.

**Authorship Contributions.** Q.Z. and N.S. contributed to the conception and design of the study. Q.Z. and N.F. collected the data and performed data analysis. Q.Z. abstracted and interpreted the results. Q.Z. and N.F. wrote the first draft, and approved the final version of the manuscript. X.B. and N.S. revised the manuscript and approved the final version.

## References

- Akhlaghi N, Baker CA, Lahlou M, Zafar H, Murthy KG, Rangwala HS, Kosecka J, Joiner WM, Pancrazio JJ and Sikdar S (2016) Real-time classification of hand motions using ultrasound imaging of forearm muscles. *IEEE Transactions on Biomedical Engineering* **63**(8), 1687–1698.
- Alibeji NA, Molazadeh V, Dicianno BE and Sharma N (2018a) A control scheme that uses dynamic postural synergies to coordinate a hybrid walking neuroprosthesis: Theory and experiments. *Frontiers in Neuroscience* **12**, 159.
- Alibeji NA, Molazadeh V, Moore-Clingenpeel F and Sharma N (2018b) A muscle synergy-inspired control design to coordinate functional electrical stimulation and a powered exoskeleton: Artificial generation of synergies to reduce input dimensionality. *IEEE Control Systems Magazine* **38**(6), 35–60.
- 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 Transactions on Neural Systems and Rehabilitation Engineering* **25**(8), 1125–1134.
- Arampatzis A, Karamanidis K, Stafiliadis S, Morey-Klapsing G, DeMonte G and Brüggemann G-P (2006) Effect of different ankle-and knee-joint positions on gastrocnemius medialis fascicle length and emg activity during isometric plantar flexion. *Journal of Biomechanics* **39**(10), 1891–1902.
- 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. *Frontiers in Neurobotics* **10**, 9.
- Awad LN, Bae J, O'donnell K, De Rossi SM, Hendron K, Sloot LH, Kudzia P, Allen S, Holt KG, Ellis TD and Walsh, CJ. (2017) A soft robotic exosuit improves walking in patients after stroke. *Science Translational Medicine* **9**(400), eaai9084.
- Bae J, De Rossi SMM, O'Donnell K, Hendron KL, Awad LN, Dos Santos TRT, De Araujo VL, Ding Y, Holt KG, Ellis TD, et al. (2015). A soft exosuit for patients with stroke: Feasibility study with a mobile off-board actuation unit. In *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*. New York: IEEE, pp. 131–138.
- Baghdadi A, Cavuoto LA and Crassidis JL (2018) Hip and trunk kinematics estimation in gait through kalman filter using imu data at the ankle. *IEEE Sensors Journal* **18**(10), 4253–4260.
- Bajd T, Kralj A, Štefančič M and Lavrač N (1999) Use of functional electrical stimulation in the lower extremities of incomplete spinal cord injured patients. *Artificial Organs* **23**(5), 403–409.
- Begovic H, Zhou G-Q, Li T, Wang Y and Zheng Y-P (2014) Detection of the electromechanical delay and its components during voluntary isometric contraction of the quadriceps femoris muscle. *Frontiers in Physiology* **5**, 494.
- Besier TF, Fredericson M, Gold GE, Beaupré GS and Delp SL (2009) Knee muscle forces during walking and running in patellofemoral pain patients and pain-free controls. *Journal of Biomechanics* **42**(7), 898–905.
- Bitzer S and Van Der Smagt P (2006) Learning emg control of a robotic hand: Towards active prostheses. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*. New York: IEEE, pp. 2819–2823.
- Blackburn JT, Bell DR, Norcross MF, Hudson JD and Engstrom LA (2009) Comparison of hamstring neuromechanical properties between healthy males and females and the influence of musculotendinous stiffness. *Journal of Electromyography and Kinesiology* **19**(5), e362–e369.

- Carneiro G and Nascimento JC** (2013) Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(11), 2592–2607.
- Chen L, Fu J, Wu Y, Li H and Zheng B** (2020) Hand gesture recognition using compact cnn via surface electromyography signals. *Sensors* **20**(3), 672.
- Coskun M, Yildirim O, Demir Y and Acharya UR** (2022) Efficient deep neural network model for classification of grasp types using semg signals. *Journal of Ambient Intelligence and Humanized Computing* **13**, 4437–4450.
- Cunningham R, Harding P and Loram I** (2017a) Deep residual networks for quantification of muscle fiber orientation and curvature from ultrasound images. In *Annual Conference on Medical Image Understanding and Analysis*. New York: Springer, pp. 63–73.
- Cunningham RJ, Harding PJ and Loram ID** (2017b) The application of deep convolutional neural networks to ultrasound for modelling of dynamic states within human skeletal muscle. Preprint, [arXiv:1706.09450](https://arxiv.org/abs/1706.09450).
- Cunningham RJ and Loram ID** (2020) Estimation of absolute states of human skeletal muscle via standard b-mode ultrasound imaging and deep convolutional neural networks. *Journal of the Royal Society Interface* **17**(162), 20190715.
- Damiano DL, Prosser LA, Curatalo LA and Alter KE** (2013) Muscle plasticity and ankle control after repetitive use of a functional electrical stimulation device for foot drop in cerebral palsy. *Neurorehabilitation and Neural Repair* **27**(3), 200–207.
- Deng W, Papavasileiou I, Qiao Z, Zhang W, Lam K-Y and Han S** (2018) Advances in automation technologies for lower extremity neurorehabilitation: A review and future challenges. *IEEE Reviews in Biomedical Engineering* **11**, 289–305.
- Dhawan AS, Mukherjee B, Patwardhan S, Akhlaghi N, Diao G, Levay G, Holley R, Joiner WM, Harris-Love M and Sikdar S** (2019) Proprioceptive sonomyographic control: A novel method for intuitive and proportional control of multiple degrees-of-freedom for individuals with upper extremity limb loss. *Scientific Reports* **9**(1), 1–15.
- Dick TJ, Biewener AA and Wakeling JM** (2017) Comparison of human gastrocnemius forces predicted by hill-type muscle models and estimated from ultrasound images. *The Journal of Experimental Biology* **220**(9), 1643–1653.
- Ding Q, Xiong A, Zhao X and Han J** (2016) A review on researches and applications of semg-based motion intent recognition methods. *Acta Automatica Sinica* **42**(1), 13–25.
- Du Y-C, Lin C-H, Shyu L-Y and Chen T** (2010) Portable hand motion classifier for multi-channel surface electromyography recognition using grey relational analysis. *Expert Systems with Applications* **37**(6), 4283–4291.
- Duan F, Dai L, Chang W, Chen Z, Zhu C and Li W** (2015) Semg-based identification of hand motion commands using wavelet neural network combined with discrete wavelet transform. *IEEE Transactions on Industrial Electronics* **63**(3), 1923–1934.
- Esquenazi A, Talaty M and Jayaraman A** (2017) Powered exoskeletons for walking assistance in persons with central nervous system injuries: A narrative review. *Physical Medicine and Rehabilitation* **9**(1), 46–62.
- Feigin M, Zwecker M, Freedman D and Anthony BW** (2020) Detecting muscle activation using ultrasound speed of sound inversion with deep learning. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. New York: IEEE, pp. 2092–2095.
- Ferris D, Sawicki G and Domingo A** (2005) Powered lower limb orthoses for gait rehabilitation. *Topics in Spinal Cord Injury Rehabilitation* **11**(2), 34–49.
- Fleischer C and Hommel G** (2008) A human–exoskeleton interface utilizing electromyography. *IEEE Transactions on Robotics* **24**(4), 872–882.
- Gopura R, Bandara D, Kiguchi K and Mann G** (2016) Developments in hardware systems of active upper-limb exoskeleton robots: A review. *Robotics and Autonomous Systems* **75**, 203–220.
- Guo J-Y, Zheng Y-P, Xie H-B and Chen X** (2010) Continuous monitoring of electromyography (EMG), mechanomyography (MMG), sonomyography (SMG) and torque output during ramp and step isometric contractions. *Medical Engineering & Physics* **32**(9), 1032–1042.
- Guo J-Y, Zheng Y-P, Xie H-B and Koo TK** (2013) Towards the application of one-dimensional sonomyography for powered upper-limb prosthetic control using machine learning models. *Prosthetics and Orthotics International* **37**(1), 43–49.
- Ha KH, Murray SA and Goldfarb M** (2016) An approach for the cooperative control of fes with a powered exoskeleton during level walking for persons with paraplegia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **24**(4), 455–466.
- Hodges P, Pengel L, Herbert R and Gandevia S** (2003) Measurement of muscle contraction with ultrasound imaging. *Muscle & Nerve* **27**(6), 682–692.
- Hodson-Tole E and Lai A** (2019) Ultrasound-derived changes in thickness of human ankle plantar flexor muscles during walking and running are not homogeneous along the muscle mid-belly region. *Scientific Reports* **9**(1), 1–11.
- Hu Y, Wong Y, Wei W, Du Y, Kankanhalli M and Geng W** (2018) A novel attention-based hybrid cnn-rnn architecture for semg-based gesture recognition. *PLoS One* **13**, e0206049.
- Huang J, Huo W, Xu W, Mohammed S and Amirat Y** (2015a) Control of upper-limb power-assist exoskeleton using a human-robot interface based on motion intention recognition. *IEEE Transactions on Automation Science and Engineering* **12**(4), 1257–1270.
- Huang, Shorter KA, Adamczyk PG and Kuo AD** (2015b) Mechanical and energetic consequences of reduced ankle plantar-flexion in human walking. *The Journal of Experimental Biology* **218**(22), 3541–3550.
- Huang Y, Yang X, Li Y, Zhou D, He K and Liu H** (2017) Ultrasound-based sensing models for finger motion classification. *IEEE Journal of Biomedical and Health Informatics* **22**(5), 1395–1405.
- Huang H, Zhang F, Hargrove LJ, Dou Z, Rogers DR and Englehart KB** (2011) Continuous locomotion-mode identification for prosthetic legs based on neuromuscular–mechanical fusion. *IEEE Transactions on Biomedical Engineering* **58**(10), 2867–2875.

- Huo W, Mohammed S, Moreno JC and Amirat Y (2014) Lower limb wearable robots for assistance and rehabilitation: A state of the art. *IEEE Systems Journal* **10**(3), 1068–1081.
- Ioffe S and Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*. Lille, France: PMLR, pp. 448–456.
- Jackson RW and Collins SH (2015) An experimental comparison of the relative benefits of work and torque assistance in ankle exoskeletons. *Journal of Applied Physiology* **119**(5), 541–557.
- Jahanandish MH, Fey NP and Hoyt K (2019) Lower-limb motion estimation using ultrasound imaging: A framework for assistive device control. *IEEE Journal of Biomedical and Health Informatics* **23**(6), 2505–2514.
- Jahanandish MH, Rabe KG, Fey NP and Hoyt K (2021) Ultrasound features of skeletal muscle can predict kinematics of upcoming lower-limb motion. *Annals of Biomedical Engineering* **49**(2), 822–833.
- Jamwal PK, Xie SQ, Hussain S and Parsons JG (2012) An adaptive wearable parallel robot for the treatment of ankle injuries. *IEEE/ASME Transactions on Mechatronics* **19**(1), 64–75.
- Kawamoto H and Sankai Y (2002) Comfortable power assist control method for walking aid by hal-3. In *IEEE International Conference on Systems, Man and Cybernetics*, Vol. 4. New York: IEEE, p. 6.
- Khokhar ZO, Xiao ZG and Menon C (2010) Surface emg pattern recognition for real-time control of a wrist exoskeleton. *Biomedical Engineering Online* **9**(1), 1–17.
- Kirsch N, Alibeji N and Sharma N (2017) Nonlinear model predictive control of functional electrical stimulation. *Control Engineering Practice* **58**, 319–331.
- Kirsch N, Bao X, Alibeji N, Dicianno B and Sharma N (2018) Model-based dynamic control allocation in a hybrid neuroprosthesis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26**(1), 224–232.
- Kirshblum SC, Priebe MM, Ho CH, Scelza WM, Chiodo AE and Wuermsler L-A (2007) Spinal cord injury medicine. 3. Rehabilitation phase after acute spinal cord injury. *Archives of Physical Medicine and Rehabilitation* **88**(3), S62–S70.
- Langhorne P, Bernhardt J and Kwakkel G (2011) Stroke rehabilitation. *The Lancet* **377**(9778), 1693–1702.
- Lee H, Rouse EJ and Krebs HI (2016) Summary of human ankle mechanical impedance during walking. *IEEE Journal of Translational Engineering in Health and Medicine* **4**, 1–7.
- Lenzi T, De Rossi SMM, Vitiello N and Carrozza MC (2012) Intention-based emg control for powered exoskeletons. *IEEE Transactions on Biomedical Engineering* **59**(8), 2180–2190.
- Liu G, Zhang L, Han, B, Zhang T, Wang Z and Wei P (2019) Semg-based continuous estimation of knee joint angle using deep learning with convolutional neural network. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. New York: IEEE, pp. 140–145.
- Lloyd DG and Besier TF (2003) An EMG-driven musculoskeletal model to estimate muscle forces and knee joint moments in vivo. *Journal of Biomechanics* **36**(6), 765–776.
- Lobo-Prat J, Kooren PN, Stienen AH, Herder JL, Koopman BF and Veltink PH (2014) Non-invasive control interfaces for intention detection in active movement-assistive devices. *Journal of Neuroengineering and Rehabilitation* **11**(1), 1–22.
- Loosey DP, McDonald CG, Battaglia E and O'Malley MK (2018) A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction. *Applied Mechanics Reviews* **70**(1), 010804.
- Lu J-y, Zou H and Greenleaf JF (1994) Biomedical ultrasound beam forming. *Ultrasound in Medicine & Biology* **20**(5), 403–428.
- Ma C, Lin C, Samuel OW, Xu L and Li G (2020) Continuous estimation of upper limb joint angle from semg signals based on sca-ml deep learning approach. *Biomedical Signal Processing and Control*, **61**, 102024.
- McIntosh J, Marzo A, Fraser M and Phillips C (2017) Echoflex: Hand gesture recognition using ultrasound imaging. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery New York, NY, USA pp. 1923–1934.
- Meng W, Xie SQ, Liu Q, Lu CZ and Ai Q (2016) Robust iterative feedback tuning control of a compliant rehabilitation robot for repetitive ankle training. *IEEE/ASME Transactions on Mechatronics* **22**(1), 173–184.
- Molazadeh V, Zhang Q, Bao X and Sharma N (2021) An iterative learning controller for a switched cooperative allocation strategy during sit-to-stand tasks with a hybrid exoskeleton. *IEEE Transactions on Control Systems Technology*, **30**(3): 1021–1036.
- Murray SA, Ha KH, Hartigan C and Goldfarb M (2014) An assistive control approach for a lower-limb exoskeleton to facilitate recovery of walking following stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **23**(3), 441–449.
- Mushahwar VK, Jacobs PL, Normann RA, Triolo RJ and Kleitman N (2007) New functional electrical stimulation approaches to standing and walking. *Journal of Neural Engineering* **4**(3), S181.
- Nuckols RW, Dick TJ, Beck ON and Sawicki GS (2020) Ultrasound imaging links soleus muscle neuromechanics and energetics during human walking with elastic ankle exoskeletons. *Scientific Reports* **10**(1), 1–15.
- Nuckols R, Lee S, Swaminathan K, Orzel D, Howe R and Walsh C (2021) Individualization of exosuit assistance based on measured muscle dynamics during versatile walking. *Science Robotics* **6**(60), eabj1362.
- O'Shea, K and Nash R (2015) An introduction to convolutional neural networks. Preprint, [arXiv:1511.08458](https://arxiv.org/abs/1511.08458).
- Pancholi, S., Joshi, A. M., and Joshi, D. (2021). A robust and accurate deep learning based pattern recognition framework for upper limb prosthesis using semg. Preprint [arXiv:2106.02463](https://arxiv.org/abs/2106.02463).
- Pehlivan AU, Losey DP and O'Malley MK (2016) Minimal assist-as-needed controller for upper limb robotic rehabilitation. *IEEE Transactions on Robotics* **32**(1), 113–124.



- Peternel L, Noda T, Petrič T, Ude A, Morimoto J and Babič J** (2016) Adaptive control of exoskeleton robots for periodic assistive behaviours based on emg feedback minimisation. *PLoS One* **11**(2), e0148942.
- Phinyomark A, Phukpattaranont P and Limsakul C** (2012) Feature reduction and selection for emg signal classification. *Expert Systems with Applications* **39**(8), 7420–7431.
- Rabe KG, Jahanandish MH, Boehm JR, Fey AM, Hoyt K and Fey NP** (2020) Ultrasound sensing can improve continuous classification of discrete ambulation modes compared to surface electromyography. *IEEE Transactions on Biomedical Engineering* **68**(4), 1379–1388.
- Rosa LG, Zia JS, Inan OT and Sawicki GS** (2021) Machine learning to extract muscle fascicle length changes from dynamic ultrasound images in real-time. *PLoS One* **16**(5), e0246611.
- Rouse EJ, Hargrove LJ, Perreault EJ and Kuiken TA** (2014) Estimation of human ankle impedance during the stance phase of walking. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **22**(4), 870–878.
- Sartori M, Farina D and Lloyd DG** (2014) Hybrid neuromusculoskeletal modeling to best track joint moments using a balance between muscle excitations derived from electromyograms and optimization. *Journal of Biomechanics* **47**(15), 3613–3621.
- Sartori M, Reggiani M, Pagello E and Lloyd DG** (2012) Modeling the human knee for assistive technologies. *IEEE Transactions on Biomedical Engineering* **59**(9), 2642–2649.
- Saxby DJ, Killen BA, Pizzolato C, Carty CP, Diamond LE, Modenese L, Fernandez J, Davico G, Barzan M, Lenton G, and da Luz, SB.** (2020) Machine learning methods to support personalized neuromusculoskeletal modelling. *Biomechanics and Modeling in Mechanobiology* **19**(4), 1169–1185.
- Shahbazi M, Atashzar SF, Tavakoli M and Patel RV** (2016) Robotics-assisted mirror rehabilitation therapy: A therapist-in-the-loop assist-as-needed architecture. *IEEE/ASME Transactions on Mechatronics* **21**(4), 1954–1965.
- Sharma N, Stegath K, Gregory CM and Dixon WE** (2009) Nonlinear neuromuscular electrical stimulation tracking control of a human limb. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **17**(6), 576–584.
- Sheng Z, Sharma N and Kim K** (2019) Quantitative assessment of changes in muscle contractility due to fatigue during NMES: An ultrasound imaging approach. *IEEE Transactions on Biomedical Engineering* **67**(3), 832–841.
- Shi J, Zheng YP, Huang QH and Chen X** (2008) Continuous monitoring of sonomyography, electromyography and torque generated by normal upper arm muscles during isometric contraction: Sonomyography assessment for arm muscles. *IEEE Transactions on Biomedical Engineering* **55**(3), 1191–1198.
- Sikdar S, Rangwala H, Eastlake EB, Hunt IA, Nelson AJ, Devanathan J, Shin A and Pancrazio JJ** (2013) Novel method for predicting dexterous individual finger movements by imaging muscle activity using a wearable ultrasonic system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **22**(1), 69–76.
- Simao M, Mendes N, Gibaru O and Neto P** (2019) A review on electromyography decoding and pattern recognition for human-machine interaction. *IEEE Access* **7**, 39564–39582.
- Siviy C, Bae J, Baker L, Porciuncula F, Baker T, Ellis TD, Awad LN and Walsh CJ** (2020) Offline assistance optimization of a soft exosuit for augmenting ankle power of stroke survivors during walking. *IEEE Robotics and Automation Letters* **5**(2), 828–835.
- Strasser EM, Draskovits T, Prashchak M, Quittan M and Graf A** (2013) Association between ultrasound measurements of muscle thickness, pennation angle, echogenicity and skeletal muscle strength in the elderly. *Age* **35**(6), 2377–2388.
- Su B-Y, Wang J, Liu S-Q, Sheng M, Jiang J and Xiang K** (2019) A cnn-based method for intent recognition using inertial measurement units and intelligent lower limb prosthesis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27**(5), 1032–1042.
- Sutskever I, Martens J, Dahl G and Hinton G** (2013) On the importance of initialization and momentum in deep learning. In *International Conference on Machine Learning*. Atlanta, USA: PMLR, pp. 1139–1147.
- Tamburella F, Tagliamonte NL, Pisotta I, Masciullo M, Arquilla M, Van Asseldonk EHF, Van der Kooij H, Wu AR, Dzeladini F, Ijspeert AJ, and Molinari, M** (2020) Neuromuscular controller embedded in a powered ankle exoskeleton: Effects on gait, clinical features and subjective perspective of incomplete spinal cord injured subjects. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28**(5), 1157–1167.
- Thomenius KE** (1996) Evolution of ultrasound beamformers. In *1996 IEEE Ultrasonics Symposium. Proceedings*, Vol. 2. New York: IEEE, pp. 1615–1622..
- Too J, Abdullah AR and Saad NM** (2019) Classification of hand movements based on discrete wavelet transform and enhanced feature extraction. *International Journal of Advanced Computer Science and Applications* **10**(6), 83–89.
- Veneman JF, Kruidhof R, Hekman EE, Ekkelenkamp R, Van Asseldonk EH and Van Der Kooij H** (2007) Design and evaluation of the Lopes exoskeleton robot for interactive gait rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **15**(3), 379–386.
- Wang X, Dong D, Chi X, Wang S, Miao Y and Gavrilov AI** (2021) Ssemg-based consecutive estimation of human lower limb movement by using multi-branch neural network. *Biomedical Signal Processing and Control* **68**, 102781.
- Xu F, Zheng Y and Hu X** (2020) Real-time finger force prediction via parallel convolutional neural networks: A preliminary study. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. New York: IEEE, pp. 3126–3129.
- Yamashita R, Nishio M, Do RKG and Togashi K** (2018) Convolutional neural networks: An overview and application in radiology. *Insights into Imaging* **9**(4), 611–629.



- Yang X, Yan J and Liu H** (2019) Comparative analysis of wearable a-mode ultrasound and semg for muscle-computer interface. *IEEE Transactions on Biomedical Engineering* **67**(9), 2434–2442.
- Young AJ and Ferris DP** (2016) State of the art and future directions for lower limb robotic exoskeletons. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**(2), 171–182.
- Yu Y, Chen C, Sheng X and Zhu X** (2020) Multi-dof continuous estimation for wrist torques using stacked autoencoder. *Biomedical Signal Processing and Control*, **57**, 101733.
- Yu Y, Chen C, Zhao J, Sheng X and Zhu X** (2021) Surface electromyography image-driven torque estimation of multi-dof wrist movements. *IEEE Transactions on Industrial Electronics* **69**(1), 795–804.
- Zanotto D, Akiyama Y, Stegall P and Agrawal SK** (2015) Knee joint misalignment in exoskeletons for the lower extremities: Effects on user's gait. *IEEE Transactions on Robotics* **31**(4), 978–987.
- Zhang Q** (2021) *Ultrasound Imaging and Surface Electromyography-Based Voluntary Effort Prediction and Control Strategies for Ankle Assistance*, PhD thesis. Raleigh, NC, USA: North Carolina State University.
- Zhang Q, Clark WH, Franz JR and Sharma N** (2022a) Personalized fusion of ultrasound and electromyography-derived neuromuscular features increases prediction accuracy of ankle moment during plantarflexion. *Biomedical Signal Processing and Control* **71**, 103100.
- Zhang J, Fiers P, Witte KA, Jackson RW, Poggensee KL, Atkeson CG and Collins SH** (2017) Human-in-the-loop optimization of exoskeleton assistance during walking. *Science* **356**(6344), 1280–1284.
- Zhang Q, Fragnito N, Franz JR and Sharma N** (2022b) Fused ultrasound and electromyography-driven neuromuscular model to improve plantarflexion moment prediction across walking speeds. *Journal of Neuroengineering and Rehabilitation*, **19**, 86. <https://doi.org/10.1186/s12984-022-01061-z>.
- Zhang Q, Iyer A, Kim K and Sharma N** (2020a) Evaluation of non-invasive ankle joint effort prediction methods for use in neurorehabilitation using electromyography and ultrasound imaging. *IEEE Transactions on Biomedical Engineering* **68**(3), 1044–1055.
- Zhang Q, Iyer A, Sun Z, Kim K and Sharma N** (2021) A dual-modal approach using electromyography and sonomyography improves prediction of dynamic ankle movement: A case study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **29**, 1944–1954.
- Zhang Q, Kim K and Sharma N** (2020b) Prediction of ankle dorsiflexion moment by combined ultrasound sonography and electromyography. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28**(1), 318–327.
- Zhang F, Li P, Hou Z-G, Lu Z, Chen Y, Li Q and Tan M** (2012) Semg-based continuous estimation of joint angles of human legs by using bp neural network. *Neurocomputing* **78**(1), 139–148.
- Zhou G-Q, Huo E-Z, Yuan M, Zhou P, Wang R-L, Wang K-N, Chen Y and He X-P** (2020) A single-shot region-adaptive network for myotendinous junction segmentation in muscular ultrasound images. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **67**(12), 2531–2542.
- Zhou S, Lawson DL, Morrison WE and Fairweather I** (1995) Electromechanical delay in isometric muscle contractions evoked by voluntary, reflex and electrical stimulation. *European Journal of Applied Physiology and Occupational Physiology* **70**(2), 138–145.
- Zhou Y, Liu J, Zeng J, Li K and Liu H** (2019) Bio-signal based elbow angle and torque simultaneous prediction during isokinetic contraction. *Science China Technological Sciences* **62**(1), 21–30.

---

**Cite this article:** Zhang Q, Fragnito N, Bao X and Sharma N (2022) A deep learning method to predict ankle joint moment during walking at different speeds with ultrasound imaging: A framework for assistive devices control. *Wearable Technologies*, **3**, e20, doi:<https://doi.org/10.1017/wtc.2022.18>