Lower-Limb Joint Torque Prediction Using LSTM Neural Networks and Transfer Learning

Longbin Zhang, Davit Soselia, Ruoli Wang, and Elena M. Gutierrez-Farewik

***Abstract*— Estimation of joint torque during movement provides important information in several settings, such as effect of athletes' training or of a medical intervention, or analysis of the remaining muscle strength in a wearer of an assistive device. The ability to estimate joint torque during daily activities using wearable sensors is increasingly relevant in such settings. In this study, lower limb joint torques during ten daily activities were predicted by long short-term memory (LSTM) neural networks and transfer learning. LSTM models were trained with muscle electromyography signals and lower limb joint angles. Hip flexion/extension, hip abduction/adduction, knee flexion/extension and ankle dorsiflexion/plantarflexion torques were predicted. The LSTM models' performance in predicting torque was investigated in both intra-subject and inter-subject scenarios. Each scenario was further divided into intra-task and inter-task tests. We observed that LSTM models could predict lower limb joint torques during various activities accurately with relatively low error (root mean square error $\leq$ 0.14 Nm/kg, normalized root mean square error $\leq$ 8.7%) either through a uniform model or through ten separate models in intra-subject tests. Furthermore, a transfer learning technique was adopted in the inter-task and inter-subject tests to further improve the generalizability of LSTM models by pre-training a model on multiple subjects and/or tasks and transferring the learned knowledge to a target task/subject. Particularly in the inter-subject tests, we could predict joint torques accurately in several movements after training from only a few movements from new subjects.**

***Index Terms*— LSTM, inverse dynamics, time series, generalizability, transfer learning.**

I. Introduction

JOINT torque is an important biomechanical parameter as it provides essential information about neuromotor control mechanisms during movement [1], [2]. The ability to predict joint torque is useful in numerous applications, such as assessment of athletes' training [3], [4], evaluation of surgical outcomes [5]–[7], and incorporation in controller design of active exoskeletons or prostheses [8]–[10]. Joint torques during movements are commonly computed through multi-body inverse dynamics, using 3D kinematics and ground reaction force data collected in an instrumented motion lab. While this approach is generally considered a gold standard for computing joint torque, it requires a specialized lab, marker-based biomechanical models of the body, and motions performed over instrumented force transducers. This approach has thus limited use in real-time joint torque estimates and in any out-of-lab environments.

Electromyography (EMG)-driven neuromusculoskeletal (NMS) modelling and deep learning have in recent years become two dominant approaches of real-time joint torque prediction based on EMG signal inputs [11]–[13]. In a recent study, we compared ankle joint torques in gait and isokinetic movements estimated from NMS models and from artificial neural networks (ANNs) [14]. Our results suggested that the ANN model was able to predict ankle torques more accurately when trained on a varied set of trials. In that study, however, EMG signals were only recorded in soleus, tibialis anterior, and gastrocnemius medialis muscles; inputs were thus limited to these three EMG signals along with ankle joint kinematics, during only gait and isokinetic ankle joint movements. Additionally, EMG-driven NMS models generally require specific calibration procedures and construction of complex relationships among different variables, such as muscle excitation-activation, force-length, force-velocity, and joint angle-musculotendon kinematics relationships. As a result, using NMS models can be time- and labor-intensive. Alternatively, deep learning models can provide an approach to extract features on multiple levels of inputs and to predict the complex relationships between inputs such as joint angles and EMGs, and outputs such as joint torques, on a large and varied set of trials.

Among deep learning models, long short-term memory (LSTM) networks have been reported as among the most effective and robust for predicting time-series data [15]–[18]. LSTM networks are capable of selectively remembering patterns that represent a duration of time by adding memory structure [19]. The memory structure can update the information across time steps using gates to decide what to remember, what to forget and what to output. Kim *et al.* [20] used an LSTM network to estimate, with high agreement, both

Manuscript received August 6, 2021; revised December 17, 2021 and February 1, 2022; accepted February 28, 2022. Date of publication March 3, 2022; date of current version March 17, 2022. This work was supported in part by the Promobilia Foundation under Grant 18200, Grant 18014, Grant 18202, Grant 19302, and Grant 21302; and in part by the Swedish Research Council under Grant 2018-00750 and Grant 2018-04902. *(Corresponding author: Elena M. Gutierrez-Farewik.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Swedish Ethical Review Authority (Dnr. 2020-02311).

Longbin Zhang and Davit Soselia are with the KTH MoveAbility Laboratory, Department of Engineering Mechanics, and the KTH BioMEx Center, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden.

Ruoli Wang and Elena M. Gutierrez-Farewik are with the KTH MoveAbility Laboratory, Department of Engineering Mechanics, and the KTH BioMEx Center, KTH Royal Institute of Technology, 10044 Stockholm, Sweden, and also with the Department of Women's and Children's Health, Karolinska Institutet, 171 77 Stockholm, Sweden (e-mail: lanie@kth.se).

Digital Object Identifier 10.1109/TNSRE.2022.3156786

the torques and position of the wrist joint based on EMG signals. Siu *et al.* [21] implemented an LSTM network to estimate ankle joint torque based on EMG and accelerometer signals during standing, walking, running, and sprinting. Their results showed that prediction errors tended to increase with locomotion speed, with the highest errors during sprinting and the lowest during standing or walking. Despite the progress of LSTM networks in joint torque prediction, there are few studies in which torque is predicted in multiple lower extremity joints and degrees of freedom (DOFs).

The generalizability of joint torque prediction models has also been a matter of concern. In short, generalized models undergo a performance drop when tested with unseen data from new and unpracticed circumstances. For example, Su *et al.* [22] proposed an LSTM model to predict gait trajectory and gait phase in multiple upcoming time frames, and report that model performance dropped in inter-subject tests. To overcome this limitation, transfer learning techniques have been extensively studied to improve the generalizability of LSTMs recently. Transfer learning is a machine learning technique that uses shared knowledge from previous experience/tasks to boost performance in other different but related tasks. A recent study by Soleimani *et al.* [23] proposed a generative adversarial network with cross-subject (inter-subject) knowledge transfer in human activity recognition. They report that their proposed model outperformed other state-of-the-art methods in more than 66% of experiments with less data for training, and thus can be used for a pre-trained model, applied on a new subject for whom it is not possible to collect and label enough data to re-train the model. By leveraging knowledge from multiple previous subjects, a new neural network can be trained with good performance using relatively few samples. However, the validity of the approach has not been studied for predicting torque in multiple joints and DOFs, nor across different subjects and daily tasks.

The objectives of this study are thus threefold: (1) to predict hip flexor/extensor (F/E), hip abductor/adductor (Ab/Ad), knee flexor/extensor and ankle dorsiflexor/plantarflexor (D/P) torques using LSTM neural networks trained with EMG and kinematics from single and from multiple movements; (2) to evaluate the generalizability of the LSTM models across subjects and tasks, and (3) to investigate whether a transfer learning strategy would improve prediction performance by making use of the knowledge learned from previous experience.

II. Methods

A. Data Collection

Eight non-disabled adults (4F/4M, (mean±SD) age 29±4 years; weight 65.2±17.8 kg; height 168.1±9.4 cm) participated. All participants provided informed written consent. Data collection was performed at the KTH MoveAbility Lab, Stockholm, Sweden, and the study was approved by the regional ethics committee (Dnr. 2020-02311). Each subject was asked to perform ten types of dynamic trials: slow walking, normal walking, fast walking, jump down from a 15-cm stair, jump up to a 15-cm stair, vertical jump up, land from vertical jump, squat, sit-to-stand and stand-to-sit. The sequence of the movements was randomized. During the experiments, each movement, or "task" as they are referred to in the current study, was repeated at least ten times.

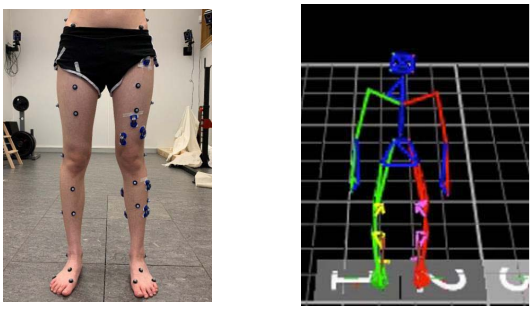

Fig. 1. (a) Marker and EMG sensor placement; (b) Kinematic model based on captured marker positions.

1) EMG Data Collection and Processing: EMG signals were recorded unilaterally using 13 EMG transmitters (aktos nano, myon, Schwarzenberg, Switzerland). The selected leg was randomized. Based on European recommendations for surface EMG [24], surface EMG electrodes were placed (Fig. 1. (a)) on the soleus (SOL), peroneus longus (PL), tibialis anterior (TA), gastrocnemius medialis (GM), gastrocnemius lateralis (GL), rectus femoris (RF), vastus medialis (VM), vastus lateralis (VL), biceps femoris (BF), semitendinosus (ST), gluteus maximus (GLMA), gluteus medius (GLME), and tensor fasciae latae (TFL) muscles. EMG signals were recorded at 1000 Hz, then band pass filtered (30-300 Hz), rectified, low pass filtered (6 Hz), and normalized to the maximum processed EMG value recorded across all trials.

2) Inverse Kinematics and Inverse Dynamics: A 3D motion capture system (V16, Vicon, Oxford, UK) was used to record the marker trajectories, which were placed according to the CGM2.3 marker set protocol (Fig. 1. (a)) [25]. Marker positions were captured at 100 Hz. During a static trial, a rigid segment model was created based on the captured marker data (Fig. 1. (b)). Inverse kinematics was used to compute the joint angles by solving a weighted root mean square problem to minimize the distance between modeled and measured markers. Ground reaction forces (GRFs) were recorded at 1000 Hz using three force plates (AMTI, MA, USA). Joint kinematics and kinetics were low-pass filtered with a fourth-order zero-lag Butterworth filter (6 Hz) [26]–[28]. Joint torques were computed using inverse dynamics computations, by solving the dynamic equations of motion [29] (Eq. (1))

$$M(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}) + \mathbf{G}(\mathbf{q}) + R(\mathbf{q})\mathbf{F}^{\mathbf{mt}} + \mathbf{F_e} = 0 \qquad (1)$$

where $\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}$ are the vector of generalised position, velocity and acceleration, respectively; $M(\mathbf{q})$ is mass matrix and $M(\mathbf{q})\ddot{\mathbf{q}}$ is a vector of inertial forces and torques; $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})$ is the vector of centripetal and Coriolis forces and torques; $\mathbf{G}(\mathbf{q})$ is the vector of gravitational forces and torques; $R(\mathbf{q})$ is the matrix of muscle moment arms; $\mathbf{F}^{\mathbf{mt}}$ is a vector of

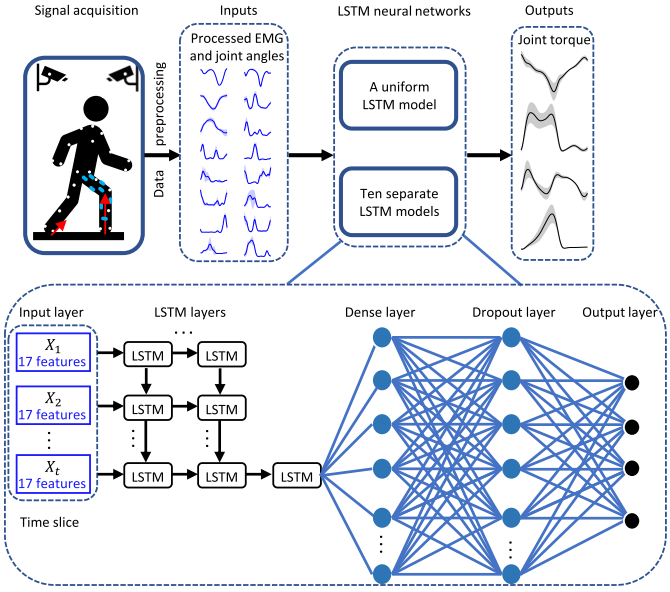

Fig. 2. Overview of joint torque prediction by LSTM neural networks. Hip flexion/extension, hip abduction/adduction, knee flexion/extension, and ankle dorsiflexion/plantarflexion joint torques were predicted through a uniform model trained by multiple-tasks or through ten separate models trained by one single task each. EMG signals and joint angles were acquired from 3D motion analysis, pre-processed, then used as input data for LSTM neural networks. For each LSTM neural network, the time-series input data were transformed as time slices data. Each time slice included five time steps and each time step had seventeen features including thirteen muscle EMG signals and four joint angles. n LSTM layers ($n = 2$ for single-task and $n = 3$ for multiple-tasks in the intra-subject scenario; $n = 3$ for single-task and $n = 4$ for multiple-tasks in the inter-subject scenario) were included, followed by a dense layer and a dropout layer. Joint torques were predicted at the output layer.

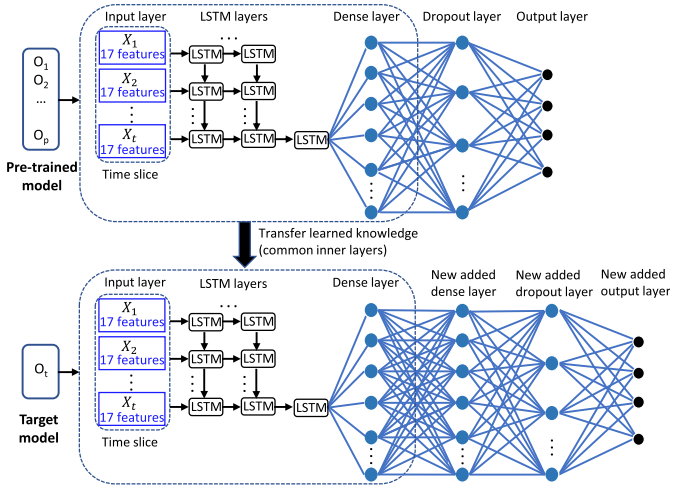

Fig. 3. The LSTM neural network architectures with transfer learning in inter-task and inter-subject tests. We extracted layers except the last two from a pre-trained model (information from previous subjects or tasks, O_1, O_2, . . . , O_p) and transferred it to the target model (O_t). In the target model, we added one dense layer, one dropout layer, and one output layer, then trained the new layers with the data from the target subject or task.

musculotendon forces and $R(\mathbf{q})\mathbf{F}^{\mathbf{mt}}$ is the vector of musculotendon torques; $\mathbf{F_e}$ is the vector of external force and torques (i.e. GRFs in this paper).

B. LSTM Neural Networks

LSTM neural networks were constructed to predict joint torque with the same inputs that are used in an EMG-driven NMS model, namely EMG signals and joint angles (Fig. 2) [30]. The EMG signals and joint angles were acquired from 3D motion analysis, pre-processed, then used as input data for the LSTM neural networks. For each LSTM neural network, the time-series input data were transformed into time slice data. Each time slice included time steps, and each time step had seventeen features, including thirteen muscle EMG signals and four joint angles. n LSTM layers ($n = 2$ for single-task and $n = 3$ for multiple-tasks in the intra-subject scenario; $n = 3$ for single-task and $n = 4$ for multiple-tasks in the inter-subject scenario) were included, followed by a dense layer and a dropout layer. The dense layer obtained the outputs from the LSTM layers to help with the regression task—joint torque prediction, whereas the dropout layer was used to prevent the neural networks from over-fitting [31]. Finally, the joint torques were predicted at the output layer.

C. Transfer Learning

To boost the prediction performance, a transfer learning method (Fig. 3) was used in the inter-task/subject scenarios to learn structural similarities by pre-training a model on multiple subjects/tasks ($O_1, O_2, \ldots, O_p$) and transferring the learned knowledge to a target task/subject. As described by Molnar *et al.* [32] and Brownlee *et al.* [33], deep neural networks learn high-level/complexity features in the hidden/middle layers that combine the low-level features extracted from the input layer, and layers close to the output layer interpret the extracted features in the context of a regression/classification task. Therefore, in transfer learning techniques, it is generally recommended to replace the layers close to the output layer, then re-use the pre-trained model and integrate it into entirely new models. Inspired by this approach, in this study, we extracted layers except the last two in the pre-trained model and transferred them to the model of target subjects or task (O_t). In the target model, we added one dense layer, one dropout layer and one output layer, then trained the weights of the new layers and fine-tuned the weights of other layers, with the data from the target task/subject. This type of fine-tuning technique is an optional step in transfer learning and can potentially achieve meaningful improvements by incrementally adapting the pre-trained features to the new data [34].

D. Hyper-Parameter Tuning for LSTM Neural Networks

The hyper-parameters of the LSTM models were determined by "coarse-to-fine" random search [35]. The batch size was set to 50. A Xavier weight with zero bias initializer was chosen. Two LSTM layers were used for the single-task cases and three LSTM layers for the multiple-tasks cases in intra-subject tests. Three LSTM layers were used for the single-task cases and four LSTM layers for the multiple-tasks cases in the inter-subject test. In the dense layer, 64 neurons were used for the single-task case and 128 neurons for the multiple-tasks case. The dropout rate was set to 0.4. Each LSTM layer

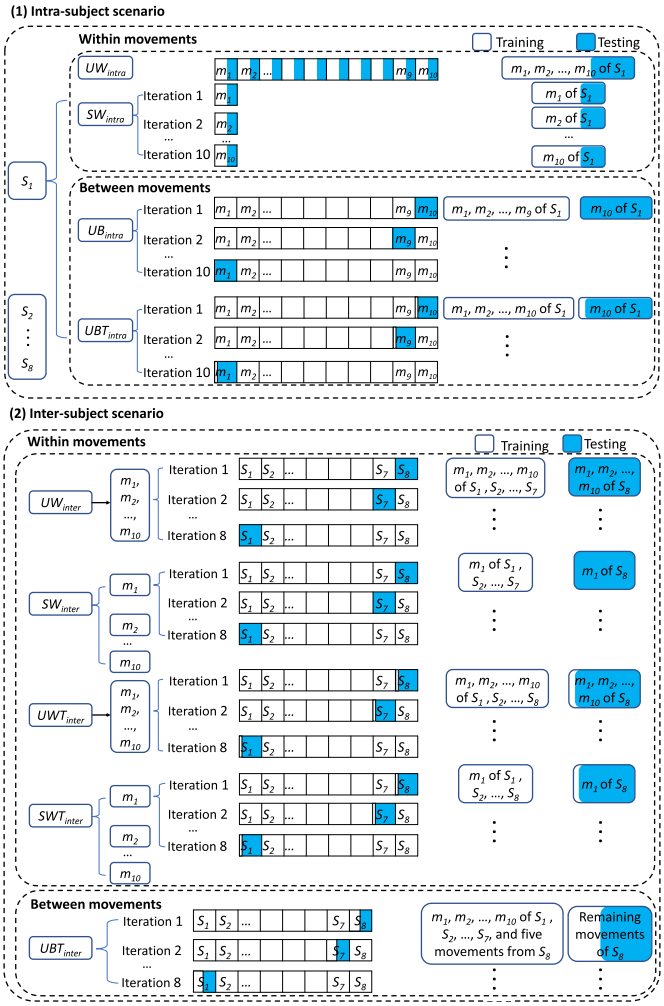

Fig. 4. The training and testing data for intra- and inter- subject (S_1, S_2, . . . , S_8) scenarios. Each scenario was further divided into intra-task (within movements m_1, m_2, . . . , m_{10}) and inter-task (between movements) tests. For each test, different cases were studied using different collected movements to train the models. Four cases were included in **intra-subject scenario**: uniform models within movements (UW_{intra}), separate models within movements (SW_{intra}), uniform models between movements (UB_{intra}), uniform models between movements with transfer learning (UBT_{intra}). Five cases were included in **inter-subject scenario**: uniform models within movements (UW_{inter}), separate models within movements (SW_{inter}), uniform models within movements with transfer learning (UWT_{inter}), separate models within movements with transfer learning (SWT_{inter}), uniform models between movements with transfer learning (UBT_{inter}).

has 50 neurons, and the tanh activation function was used. The mean square error between predicted and experimentally measured joint torque is used as the loss function during the training. The Adam optimizer was adopted with a default learning rate, 10^{-3}. In the transfer-learning approach, the learning rate was tuned to 10^{-4} and 10^{-5} for fine-tuning. The optimum model was trained for 4000 epochs with an early stop if the prediction performance did not increase in the consecutive 50 epochs.

E. Evaluation Protocol

The lower limb joint torque prediction performance of the LSTM models was investigated in intra- and inter- subject scenarios. Each scenario was further divided into intra-task (within movements) and inter-task (between movements) tests. In addition, the benefit of adding transfer learning was also investigated across tasks and subjects (Fig. 4). For each test, different cases were studied using different collected movements to train the models. For more detailed illustrations of all cases, see Fig. 4).

1) Intra-Subject Scenario:

- **Uniform models within movements** (UW_{intra}): trained models using data from all ten movements. Contains seven layers.
- **Separate models within movements** (SW_{intra}): trained separate models using data from each movement separately. Contains six layers.
  Both UW_{intra} and SW_{intra} were trained on 80% of the data and tested on the remaining 20%.
- **Uniform models between movements** (UB_{intra}): trained models using data from multiple movements except one (leave-one-out) and tested on the remaining "new" movement. Contains seven layers.
- **Uniform models between movements with transfer learning** (UBT_{intra}): used pre-trained model from UB_{intra} method as prior information to a new movement; then re-trained models using data from the new movement and tested on the remaining "new" movement. Contains eight layers.

2) Inter-Subject Scenario:

- **Uniform models within movements** (UW_{inter}): trained models using data from all ten movements; The models were trained from multiple subjects except one and tested on a remaining "new" subject. Contains eight layers.
- **Separate models within movements** (SW_{inter}): trained separate models using data at each movement separately. The models were trained from multiple subjects except one and tested on a remaining "new" subject. Contains seven layers.
- **Uniform models within movements with transfer learning** (UWT_{inter}): used pre-trained models from UW_{inter} method as prior information to new subject; then re-trained models using data from the same movements of new subject and tested on the remaining trials of the same movements of "new" subject. Contains nine layers.
- **Separate models within movements with transfer learning** (SWT_{inter}): used pre-trained models from SW_{inter} method as prior information to new subject; then re-trained models using data from the same movements of new subject and tested on the remaining trials of the same movements of "new" subject. Contains nine layers.
- **Uniform models between movements with transfer learning** (UBT_{inter}): similar to UWT_{inter} method, but only used data from five movements of the new subject to re-train the models, and tested on the remaining movements of the "new" subject. Contains nine layers.

Prediction accuracy was considered high when prediction error was low. The prediction error of each LSTM model was computed as the root mean square error (RMSE) between predicted (by the LSTM) and measured (calculated by inverse dynamics) joint torques, normalized by body mass. In addition,

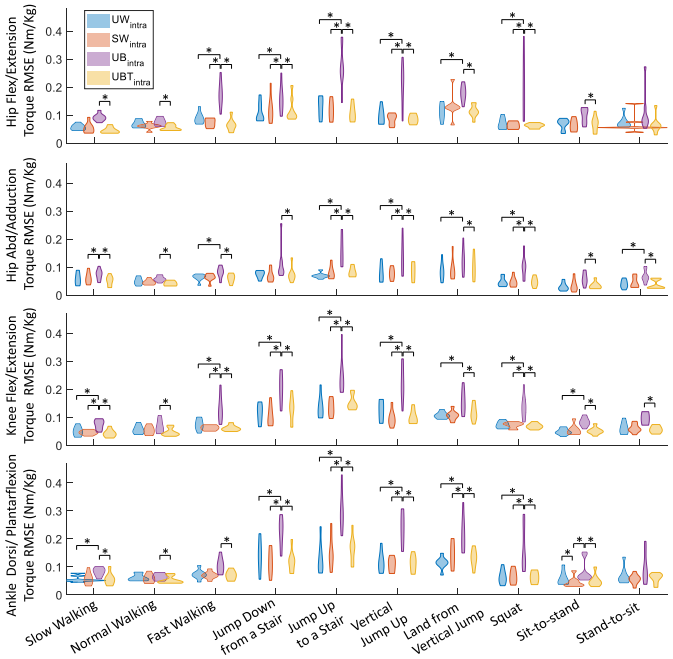

Fig. 5. Violin plots [37] depicting the distributions of RMSEs between predicted and measured joint torque (normalized by body mass) across subjects during ten movements in the intra-subject scenario, with four cases: uniform models within movements (UW_{intra}), separate models within movements (SW_{intra}), uniform models between movements (UB_{intra}), and uniform models between movements with transfer learning (UBT_{intra}). $*$ indicates a significant difference between two cases based on a Wilcoxon signed-rank test with Bonferroni correction for multiple comparisons.

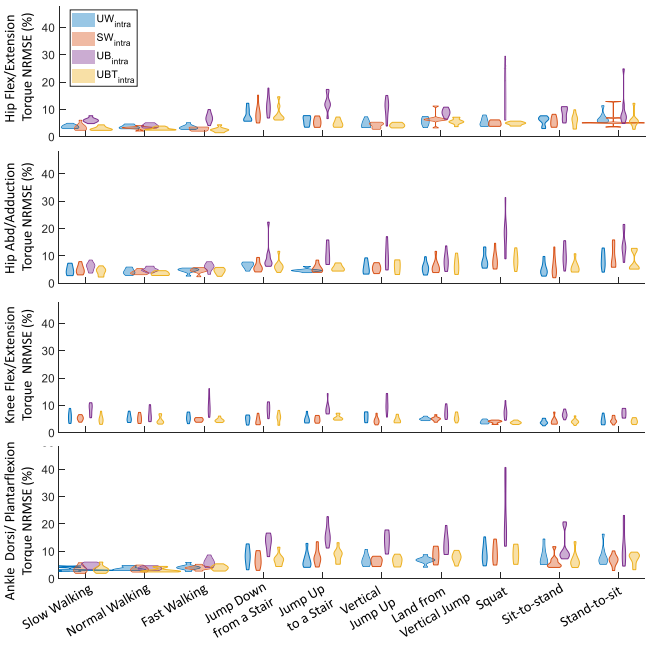

Fig. 6. Violin plots depicting the distributions of normalized root mean square error (NRMSE) between predicted and measured joint torques across subjects during ten movements in the intra-subject scenario, with four cases: uniform models within movements (UW_{intra}), separate models within movements (SW_{intra}), uniform models between movements (UB_{intra}), and uniform models between movements with transfer learning (UBT_{intra}). The NRMSE is defined as RMSE shown as % of measured joint torque during this movement.

Wilcoxon signed-rank tests with Bonferroni correction for multiple comparisons [36] were used to identify prediction error differences between torques predicted via the different models ($p < 0.05$ significance level). Normalized root mean square error (NRMSE) was also computed as RMSE divided by the range of measured joint torque during different movements. The RMSE and NRMSE were computed for each subject and then the median value was obtained across eight subjects.

III. Results

A. Intra-Subject Tests

Overall, the prediction error was similar between uniform UW_{intra} and separate SW_{intra} models within movements (Fig. 5 and Fig. 6) during all ten movements. However, the prediction errors on uniform models between movements UB_{intra} were considerably higher than those from UW_{intra} and SW_{intra} in most movements, i.e., jump down from a stair, jump up to a stair, vertical jump up, land from vertical jump and squat. Among the UW_{intra}, SW_{intra} and UB_{intra} three cases, the predicted torques (Fig. 7) all agreed well with the measured torques in walking movements, and predictions from within movements models (UW_{intra} and SW_{intra}) generally agreed better than from between movements models (UB_{intra}) in other movements.

The trained models in UW_{intra} and SW_{intra} (Fig. 5 and Fig. 6) predicted joint torques accurately with relatively low error (RMSE $\leq$ 0.14 Nm/kg, NRMSE $\leq$ 8.7%), especially in walking movements (hip F/E (RMSE $\leq$ 0.08 Nm/kg, NRMSE $\leq$ 3.6%), hip Ab/Ad (RMSE $\leq$ 0.07 Nm/kg, NRMSE $\leq$ 4.9%), knee F/E (RMSE $\leq$ 0.07 Nm/kg, NRMSE $\leq$ 5.7%) and ankle D/P (RMSE $\leq$ 0.07 Nm/kg, NRMSE $\leq$ 4.1%)). Prediction error in UB_{intra} (RMSE $\leq$ 0.29 Nm/kg, NRMSE $\leq$ 18.4%) was higher than in UW_{intra} and SW_{intra}. The trained models in UB_{intra} have relatively higher prediction error in hip F/E (RMSE: 0.18 Nm/kg, NRMSE: 13.6%), hip Ab/Ad (RMSE: 0.10 Nm/kg, NRMSE: 18.2%) and ankle D/P (RMSE: 0.13 Nm/kg, NRMSE: 18.4%) during squat than in other movements.

Among the joint torque prediction in different movements of all cases (Fig. 6), models in walking movements have relatively lower error (NRMSE $\leq$ 9.2%).

When trained with UBT_{intra} with transfer learning, trained models overall better predicted joint torques than in UB_{intra}, i.e. with a relatively low normalized error (RMSE $\leq$ 0.17 Nm/kg, NRMSE $\leq$ 9.1%, Fig. 5 and Fig. 6) and better agreement with measured torques (Fig. 7).

B. Inter-Subject Tests

When testing on uniform UW_{inter} and separate SW_{inter} models within movements (Fig. 8 and Fig. 9), the prediction errors were relatively higher (UW_{inter}: RMSE $\leq$ 0.24 Nm/kg, NRMSE $\leq$ 21%; SW_{inter}: RMSE $\leq$ 0.25 Nm/kg, NRMSE $\leq$ 19.5%) than in the intra-subject scenario. In particular, prediction errors were higher in hip Ab/Ad (NRMSE UW_{inter}: 19.4%; SW_{inter}: 19.5%) and ankle D/P moments

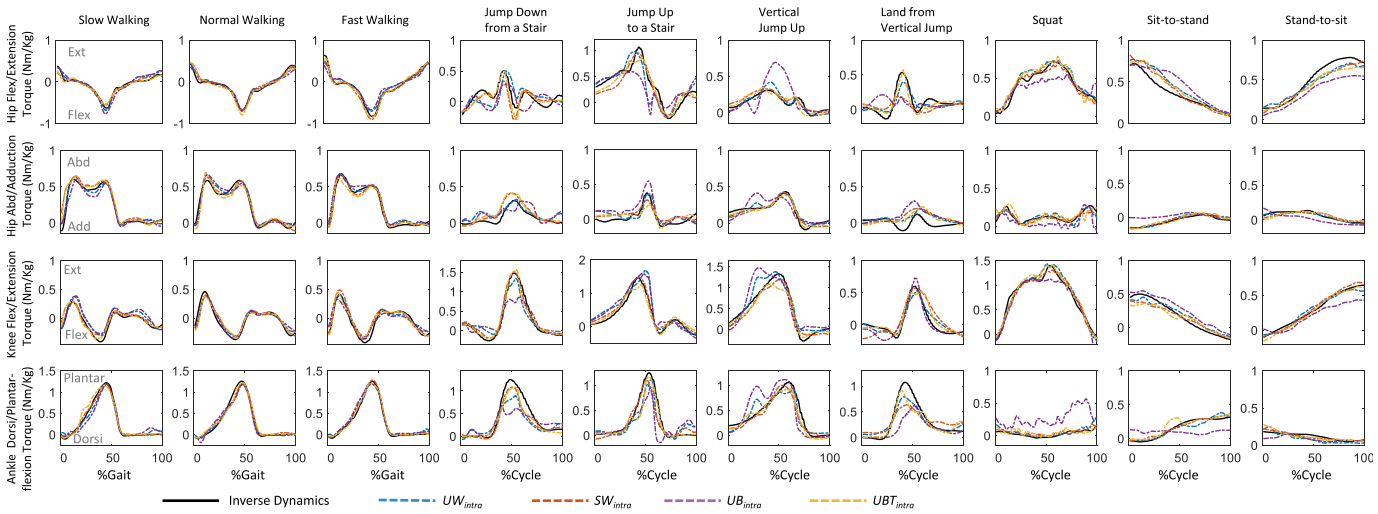

Fig. 7. one example trial of measured (computed by inverse dynamics) and estimated joint torque via models during ten movements in intra-subject scenario, with four cases: uniform models within movements (UW_{intra}), separate models within movements (SW_{intra}), uniform models between movements (UB_{intra}), and uniform models between movements with transfer learning (UBT_{intra}).

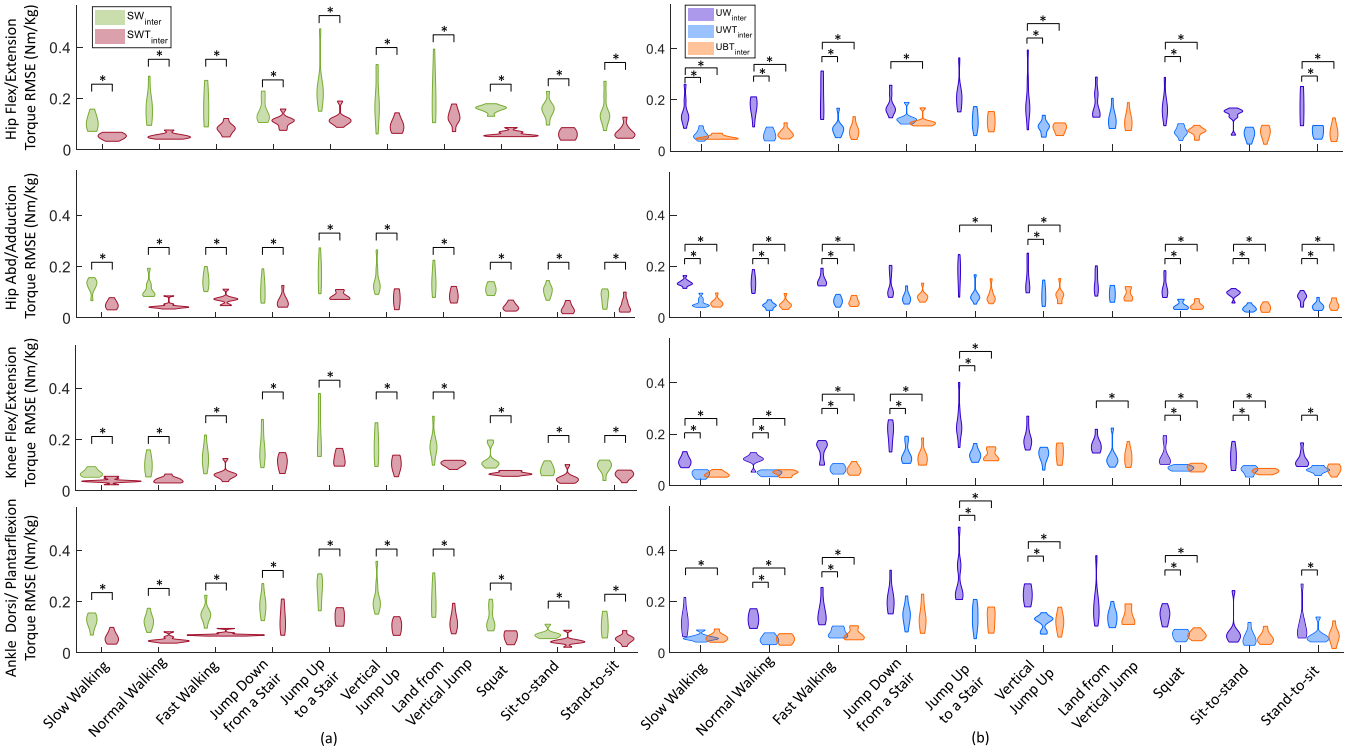

Fig. 8. Violin plots depicting the distributions of RMSEs between predicted and measured joint torque (normalized by body mass) across subjects during ten movements in intra-subject scenario with five cases: (a) separate models within movements (SW_{inter}) and separate models within movements with transfer learning (SWT_{inter}), (b) uniform models within movements (UW_{inter}), uniform models within movements with transfer learning (UWT_{inter}) and uniform models between movements with transfer learning (UBT_{inter}). $*$ indicates a significant difference between two cases based on a Wilcoxon signed-rank test with Bonferroni correction for multiple comparisons.

(UW_{inter}: 21%; SW_{inter}: 16.5%) during squat, hip F/E (UW_{inter}: 15.1%; SW_{inter}: 12.7%), and hip Ab/Ad (UW_{inter}: 16.8%; SW_{inter}: 17.5%) during stand-to-sit, as well as hip F/E: (UW_{inter}: 13.4%; SW_{inter}: 13.8%), and hip Ab/Ad (UW_{inter}: 16.7%; SW_{inter}: 18.7%) during sit-to-stand than in other joint torques during different movements.

Compared to SW_{inter}, SWT_{inter} (that used transfer learning) predicted joint torques with significantly lower error (hip F/E: RMSE $\leq$0.13 Nm/kg, NRMSE $\leq$ 8.0%; hip Ab/Ad: RMSE $\leq$0.08 Nm/kg, $\leq$ 7.8%; knee F/E: RMSE $\leq$0.11 Nm/kg, $\leq$ 5.0%; ankle D/P: RMSE $\leq$ 0.14 Nm/kg, $\leq$ 9.4%, Fig. 8 (a)).

Compared to UW_{inter} (without transfer learning), UWT_{inter} and UBT_{inter} (both with transfer learning) predicted joint torques with significantly lower errors (UWT_{inter}: RMSE $\leq$0.15 Nm/kg, NRMSE $\leq$ 10.1%;

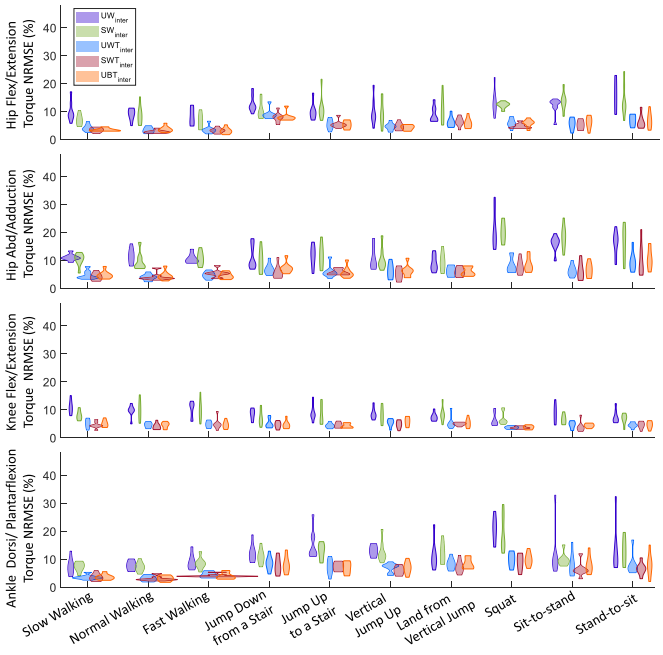

Fig. 9. Violin plots depicting the distributions of normalized root mean square error (NRMSE) between predicted and measured joint torques across subjects during ten movements in the inter-subject scenario, with five cases: uniform models within movements (UW_{inter}), separate models within movements (SW_{inter}), uniform models within movements with transfer learning (UWT_{inter}), separate models within movements with transfer learning (SWT_{inter}), and uniform models between movements with transfer learning (UBT_{inter}). The NRMSE is defined as RMSE shown as % of measured joint torque during this movement.

UBT_{inter}: RMSE $\leq$0.14 Nm/kg, NRMSE $\leq$ 9.9%) and better agreement with measured torques (Fig. 10). Prediction errors between UWT_{inter} (re-trained with all movements) and UBT_{inter} (re-trained with a subset of five movements) did not significantly differ (Fig. 8 (b)).

IV. Discussion

In this study, ankle and knee joint torques in the sagittal plane, as well as hip joint torques in both sagittal and frontal planes during ten different motions were predicted using LSTM neural networks and transfer learning. We evaluated the generalizability of LSTM models to predict torques across tasks and subjects and studied whether generalizability improved with the adoption of a transfer learning technique, which takes advantage of information extracted from previous tasks and/or subjects. We observed that the models predicted lower limb joint torques during various activities accurately through a uniform model trained by multiple tasks or through ten separate models trained by a single task each, in intra-subject tests. The joint torque prediction performance dropped when generalizing the models across tasks and subjects. To overcome this, we adopted transfer learning, which significantly improved the prediction accuracy and thus the generalizability of the LSTM models. Particularly in the inter-subject tests, we could predict joint torques accurately with only a few movements from "new" subjects. Our findings, particularly those relating to how to train generalizable models with inputs from few movements, can provide useful guidelines for training models when only minimal amounts of movement input are feasible, and thus holds great promise in applications such as incorporation in control strategy design of active prostheses or exoskeletons, real-time assessment of athletes training and evaluation of surgical outcomes.

Among ANNs, LSTM neural networks have frequently been used to predict joint torques and other sequential data in biomechanics applications, as an alternative model-free method to EMG-driven NMS models to map EMG signals to joint torque for real-time applications. Compared to NMS models, LSTM networks map the relationship of EMG to joint torque without explicitly modelling the relationship of different physiological variables, such as muscle excitation-activation, muscle force-length, and joint angles-musculotendon kinematics relationships. In addition, an NMS model is usually subject-specific, and it is usually recommended to calibrate it with experimental data to obtain subject-specific NMS parameters, such as optimal fiber length and tendon slack length, after which joint torque can be further estimated with the subject-specific model [9], [30], [38]. Since NMSs are subject-specific, few studies have investigated their generalizability across subjects. If, however, these subject-specific relationships are sought, then an NMS model-based approach is clearly more appropriate. In our previous study [14], we estimated ankle joint torques in gait and isokinetic movements using both the NMS model and ANN-based approaches. Our results suggested that the ANN model was able to deliver a better prediction of ankle joint torques when it could be trained on a varied set of motion trials. It is therefore useful to define guidelines that maximize the prediction accuracy if deep learning algorithms are to be used to predict joint torques. In the current paper, we constructed different LSTM networks in different cases to predict joint torques in multiple joints of anatomical planes and investigated their generalizability across movements and subjects.

The LSTM models predicted multiple joint torques accurately with relatively low error (RMSE $\leq$ 0.14 Nm/kg, NRMSE $\leq$ 8.7%), both through a uniform model trained by multiple tasks and through ten separate models trained by a single task each, in intra-subject tests (UW_{intra} and SW_{intra}, Fig. 5). Both of these models are tested with the unseen, "new" trials of the same movement types that were used for training. Generally, deep learning can be expected to show good prediction performance when the same motions are used to train and test the model. The constructed LSTM model with multiple layers and neurons are capable of selectively remembering patterns for a duration of time and could extract the complex nonlinear relationship between EMG signals and joint torques, and thus accurately predict torques in unseen trials of the same movement types, regardless of whether the network consists of a uniform model or of ten separate models. A recent study by Xiong *et al.* [39] also predicted hip F/E, hip Ab/Ad, knee F/E and ankle D/P torques accurately (NRMSE$\leq$ 7.9%) using a feed forward ANN with 5-6 EMG signals. However, the only movements tested were walking at four different speeds, and models were only investigated in the intra-subject scenario. In our study, we predicted torques with better accuracy in similar scenarios (in walking

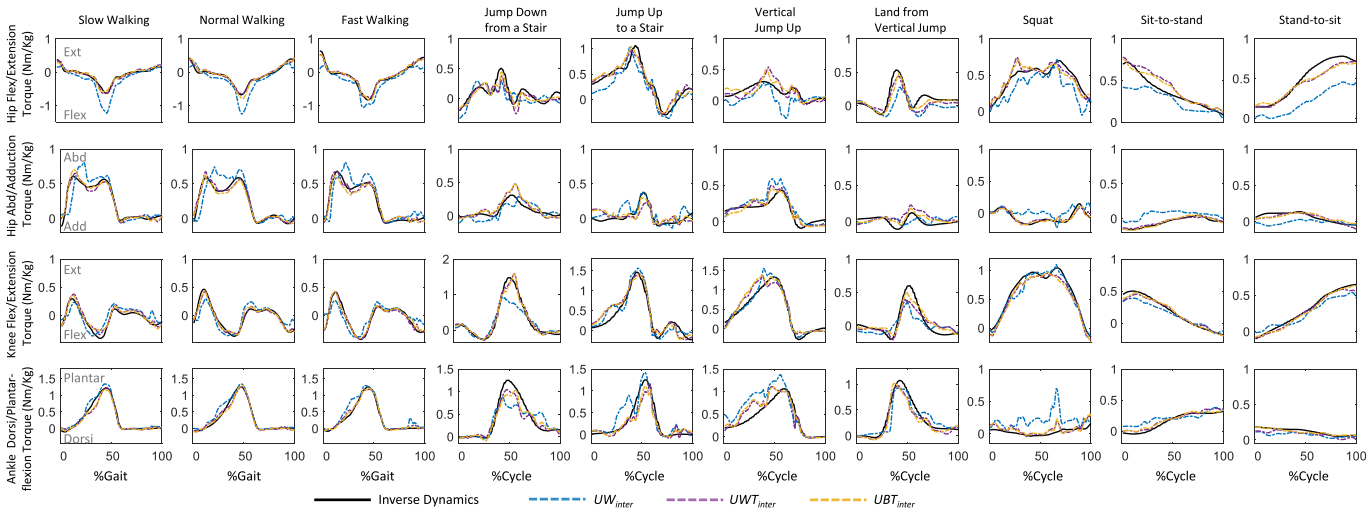

Fig. 10. Example results in one subject of measured (through inverse dynamics) and predicted joint torques in hip flexion/extension, hip abd/adduction, knee flexion/extension, and ankle dorsi/plantarflexion, during all ten movements inter-subject scenario with three cases: uniform models within movements (UW_{inter}), uniform models within movements with transfer learning (UWT_{inter}) and uniform models between movements with transfer learning (UBT_{inter}).

movements, hip F/E NRMSE$\leq$ 3.6%; hip Ab/Ad NRMSE$\leq$ 4.9%; knee F/E NRMSE$\leq$ 5.7%; and ankle D/P NRMSE$\leq$ 4.1%), and evaluated their generalizability across subjects and across ten movements/tasks. Another study about EMG-driven NMS model proposed by Pizzolato *et al.* [30] was able to predict knee F/E and ankle D/P torques accurately (RMSE 0.18 Nm/kg and 0.24 Nm/kg, respectively) but predicted hip F/E torques somewhat less accurately (RMSE 0.39 Nm/kg) in walking movements. Specifically, it underestimated the hip flexion moment in the second half of stance phase, likely attributable to the inability to acquire surface EMG data on the iliopsoas, a deep muscle. In our study, there is no apparent different prediction accuracy between joints of LSTM models in walking movements (RMSE hip F/E $\leq$0.08 Nm/kg; hip Ab/Ad $\leq$ 0.07 Nm/kg; knee F/E $\leq$0.07 Nm/kg; and ankle D/P $\leq$ 0.07 Nm/kg). In the same study by Pizzolato *et al.* [30], they further proposed an EMG-assisted NMS model that includes a static optimization to synthesize EMG signals that could not be acquired experimentally, and allows muscle excitation adjustments that better track joint torque. This approach improved the prediction accuracy in torques in which deep muscle activity is prominent. While the prediction accuracy in the EMG-assisted model was improved, especially at the hip (RMSE 0.07 Nm/kg in hip F/E), the optimization procedure can be time-consuming, thus not currently suitable for real-time applications.

As could be expected, the LSTMs' generalizability was not good in inter-task tests of the intra-subject scenario (UB_{intra}, Fig. 5) and intra-task tests of the inter-subject scenario (UW_{inter} and SW_{inter}, Fig. 8) when transfer learning was not included. Without transfer learning, LSTM models were trained with only information from previous subjects/tasks but none from the new participant/tasks. This performance drop was not surprising, as muscle coordination patterns may be quite different between multiple tasks and subjects [40], [41]. Traditional neural networks generally work well under a common assumption that the training and testing data have the same feature space and distribution. However, when distribution differs, the prediction results are likely to degrade due to the differences in domain data. As an example in inter-task tests, relatively high prediction errors were observed during squat (NRMSE $\leq$ 18.4%, UB_{intra}, Fig. 6) but lower errors were found during walking movements (NRMSE $\leq$ 9.2%, UB_{intra}, Fig. 6). We attribute this finding to the larger discrepancy in muscle coordination patterns during in squat vs. in other motions, compared to the smaller discrepancy in muscle coordination patterns during gait at one speed vs. those in gait at other speeds. It is important to note that in inter-subject tests, higher prediction errors were also observed in squats. Different squat techniques may be used by different subjects. One previous study by Slater *et al.* [42] demonstrated that differently aligned squats in the frontal and sagittal planes result in significantly different muscle activation patterns. It is therefore possible that, due to the different muscle coordination patterns across some tasks and subjects, it may not be possible to achieve good generalizability without at least some training data from a new subject or task.

To improve the generalizability of neural networks, transfer learning is a popular technique for cross-subject or cross-task validation. Transfer learning is inspired by the fact that humans can learn new tasks more quickly if they have learned similar knowledge previously, especially with limited data [43]. A recent study by Kian *et al.* [44] stated that the effectiveness of EMG-driven NMS model calibration is task-dependent and recommends using a broad range of contrasting tasks to produce the most accurate and physiologically relevant musculotendon and EMG-to-activation parameters. However, for persons with disabilities, it would quite likely be very difficult to perform a broad range of tasks. It is especially beneficial for situations in which a large amount of data

would be required, and when collecting such data on a new subject/movement would be expensive, time-consuming, and possibly difficult, for instance in persons with motor disabilities. When transfer learning was implemented, the joint torque prediction performance was significantly improved in all cases. For instance, the maximum prediction errors in squat in inter-task (NRMSE 18.4%, UB_{intra}, Fig. 6) and inter-subject (21.0%, UW_{inter} and SW_{inter}, Fig. 9) scenarios reduced (8.7%, (UBT_{intra}, Fig. 6) and 9.6% (UWT_{inter}, SWT_{inter}, and UBT_{inter}, Fig. 9), respectively) when transfer learning was included. It is important to note that the generalized model could predict joint torques with similar accuracy in inter-subject UWT_{inter} (re-trained on all movements) and UBT_{inter} (re-trained on a subset of 5 movements), indicating that the model's generalizability when trained on fewer movement types is not compromised. This is likely attributable to the shared similarities in the pre-trained model based on previous subjects and tasks, which was used as the starting point to train a new model for the new subject. The starting point would influence the performance of the trained neural network model wherein initial weights that are closer to the target solution will result in better network prediction performance [45], [46]. In short, with the incorporation of transfer learning in inter-subject/task scenarios, LSTM models can predict joint torques accurately with less effort from new subjects, which increases their feasibility when acquiring a large amount of training data may not be possible.

Though we only investigate hip, knee and ankle sagittal plane moments and hip frontal plane moments in the current study, our workflow could be further applied to other joints and degrees of freedom in different applications. For example, predicting hip internal/external rotator torques during a change in direction while walking [47] and knee abductor/adductor torque during an evaluation of a patient-specific cost function [7] or a real-time assessment to prevent anterior cruciate ligament injury [48] could all be of interest.

There are some limitations in this study. The major outcome measurement by which we determine prediction accuracy was computed errors of predicted to the 'ground truth' torques from experimental motion capture and inverse dynamics. While low error should certainly be a goal in a prediction model, the timing of prediction inaccuracies can have large implications on exoskeleton control; prediction inaccuracies during, for example, swing vs. stance may have different practical consequences. Furthermore, inaccurately predicted torques will alter the user's dynamics, further complicating the task of creating a stable gait. These should be topics of future studies before realistic implementation can be achieved. We predicted joint torques using only LSTM neural networks. While our aim was to predict joint torques using different trained LSTM models and to improve and evaluate their generalizability across tasks and subjects, other types of neural networks may have had even higher prediction accuracy, which warrants further investigation. Also, in UBT_{inter}, we only use five movements from a new subject to re-train the LSTM model for predicting unseen movements of that subject. A subsequent sensitivity analysis could be performed to determine how many and which movements are optimal for re-training the model, more specifically to increase feasibility without sacrificing prediction accuracy. Furthermore, one could consider standardizing the chair height in relation to each subject's height during sit-to-stand and stand-to-sit, and could also take into account different subjects' squat techniques, as they can result in different muscle activation patterns [42]. Also, the subjects enrolled in this study were relatively homogeneous with regards to age and overall health. Further studies using a subject group with a wider spread of both age and of neuromuscular function is warranted.

V. Conclusion

In this paper, we predicted ankle and knee joint torques in the sagittal plane, and hip joint torques in both sagittal and frontal planes during several different motions using LSTM neural networks and transfer learning. Our results illustrate that both a uniform model trained on all ten movements and ten separate LSTM models trained on one movement each could predict lower limb joint torques accurately with low prediction error in intra-subject scenarios where both of these models were tested with the unseen, "new" trials of the same movement types that were used for training. Adopting transfer learning further improved the LSTMs' generalizability across tasks and subjects, even if re-trained with a smaller subset of movements from the "new" subject. Detailed information relating to how to construct LSTM models with good generalizability can provide useful guidelines to train neural networks with minimal data in different circumstances, and thus holds great promise in applications ranging from exoskeleton control to real-time assessment of interventions.

References

[1] R. F. Zernicke and J. L. Smith, "Biomechanical insights into neural control of movement," in *Handbook of Physiology*. New York, NY, USA: Oxford Univ. Press, 1996, pp. 293–330.

[2] Y. Liu, S.-M. Shih, S.-L. Tian, Y.-J. Zhong, and L. Li, "Lower extremity joint torque predicted by using artificial neural network during vertical jump," *J. Biomech.*, vol. 42, no. 7, pp. 906–911, May 2009.

[3] G. D. Myer, K. R. Ford, O. P. Palumbo, and T. E. Hewett, "Neuromuscular training improves performance and lower-extremity biomechanics in female athletes," *J. Strength Conditioning Res.*, vol. 19, no. 1, pp. 51–60, Feb. 2005.

[4] A. J. Harrison, S. P. Keane, and J. Coglan, "Force-velocity relationship and stretch-shortening cycle function in sprint and endurance athletes," *J. Strength Conditioning Res.*, vol. 18, no. 3, pp. 473–479, Aug. 2004.

[5] J. A. Reinbolt, R. T. Haftka, T. L. Chmielewski, and B. J. Fregly, "A computational framework to predict post-treatment outcome for gait-related disorders," *Med. Eng. Phys.*, vol. 30, no. 4, pp. 434–443, May 2008.

[6] G. Lenaerts *et al.*, "Subject-specific hip geometry affects predicted hip joint contact forces during gait," *J. Biomech.*, vol. 41, no. 6, pp. 1243–1252, 2008.

[7] B. J. Fregly, J. A. Reinbolt, and T. L. Chmielewski, "Evaluation of a patient-specific cost function to predict the influence of foot path on the knee adduction torque during gait," *Comput. Methods Biomech. Biomed. Eng.*, vol. 11, no. 1, pp. 63–71, Feb. 2008.

[8] C. Pizzolato *et al.*, "Neuromusculoskeletal modeling-based prostheses for recovery after spinal cord injury," *Frontiers Neurorobotics*, vol. 13, pp. 1–10, Dec. 2019.

[9] M. Sartori, G. Durandau, S. Došen, and D. Farina, "Robust simultaneous myoelectric control of multiple degrees of freedom in wrist-hand prostheses by real-time neuromusculoskeletal modeling," *J. Neural Eng.*, vol. 15, no. 6, Dec. 2018, Art. no. 066026.

[10] S. Yao, Y. Zhuang, Z. Li, and R. Song, "Adaptive admittance control for an ankle exoskeleton using an EMG-driven musculoskeletal model," *Frontiers Neurorobotics*, vol. 12, no. 16, pp. 1–12, Apr. 2018.

[11] G. Durandau, D. Farina, and M. Sartori, "Robust real-time musculoskeletal modeling driven by electromyograms," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 3, pp. 556–564, Mar. 2018.

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

[13] M. Ezati, B. Ghannadi, and J. McPhee, "A review of simulation methods for human movement dynamics with emphasis on gait," *Multibody Syst. Dyn.*, vol. 47, no. 3, pp. 265–292, Nov. 2019.

[14] L. Zhang, Z. Li, Y. Hu, C. Smith, E. M. G. Farewik, and R. Wang, "Ankle joint torque estimation using an EMG-driven neuromusculoskeletal model and an artificial neural network model," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 564–573, Apr. 2021.

[15] A. Choi, H. Jung, and J. H. Mun, "Single inertial sensor-based neural networks to estimate COM-COP inclination angle during walking," *Sensors*, vol. 19, no. 13, pp. 1–12, 2019.

[16] H. Su, Y. Hu, H. R. Karimi, A. Knoll, G. Ferrigno, and E. D. Momi, "Improved recurrent neural network-based manipulator control with remote center of motion constraints: Experimental results," *Neural Netw.*, vol. 131, pp. 291–299, Nov. 2020.

[17] Z. Li, Y. Xia, D. Wang, D.-H. Zhai, C.-Y. Su, and X. Zhao, "Neural network-based control of networked trilateral teleoperation with geometrically unknown constraints," *IEEE Trans. Cybern.*, vol. 46, no. 5, pp. 1051–1064, May 2016.

[18] B. Wu, J. Zhong, and C. Yang, "A visual-based gesture prediction framework applied in social robots," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 3, pp. 510–519, Mar. 2022.

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

[20] D. Kim *et al.*, "Simultaneous estimations of joint angle and torque in interactions with environments using EMG," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 3818–3824.

[21] H. C. Siu, J. Sloboda, R. J. McKindles, and L. A. Stirling, "Ankle torque estimation during locomotion from surface electromyography and accelerometry," in *Proc. 8th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, Nov. 2020, pp. 80–87.

[22] B. Su and E. M. Gutierrez-Farewik, "Gait trajectory and gait phase prediction based on an LSTM network," *Sensors*, vol. 20, no. 24, pp. 1–17, 2020.

[23] E. Soleimani and E. Nazerfard, "Cross-subject transfer learning in human activity recognition systems using generative adversarial networks," *Neurocomputing*, vol. 426, pp. 26–34, Feb. 2021.

[24] H. J. Hermens *et al.*, "European recommendations for surface electromyography," *Roessingh Res. Dev.*, vol. 8, no. 2, pp. 13–54, 1999.

[25] F. Leboeuf, R. Baker, A. Barré, J. Reay, R. Jones, and M. Sangeux, "The conventional gait model, an open-source implementation that reproduces the past but prepares for the future," *Gait Posture*, vol. 69, pp. 235–241, Mar. 2019.

[26] D. A. Winter, H. G. Sidwall, and D. A. Hobson, "Measurement and reduction of noise in kinematics of locomotion," *J. Biomech.*, vol. 7, no. 2, pp. 157–159, Mar. 1974.

[27] A. Mantoan, C. Pizzolato, M. Sartori, Z. Sawacha, C. Cobelli, and M. Reggiani, "MOtoNMS: A MATLAB toolbox to process motion data for neuromusculoskeletal modeling and simulation," *Source Code Biol. Med.*, vol. 10, no. 1, pp. 1–14, Dec. 2015.

[28] T. R. Derrick, A. J. van den Bogert, A. Cereatti, R. Dumas, S. Fantozzi, and A. Leardini, "ISB recommendations on the reporting of intersegmental forces and moments during human motion analysis," *J. Biomech.*, vol. 99, pp. 1–10, Jan. 2020.

[29] M. G. Pandy, "Computer modeling and simulation of human movement," *Annu. Rev. Biomed. Eng.*, vol. 3, no. 1, pp. 245–273, 2001.

[30] C. Pizzolato *et al.*, "CEINMS: A toolbox to investigate the influence of different neural control solutions on the prediction of muscle excitation and joint moments during dynamic motor tasks," *J. Biomech.*, vol. 48, no. 14, pp. 3929–3936, Nov. 2015.

[31] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

[32] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Abu Dhabi, United Arab Emirates: Lulu, 2020.

[33] J. Brownlee, *Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python*. Vermont, VIC, Australia: Machine Learning Mastery, 2019.

[34] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[35] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. 2, pp. 1–25, 2012.

[36] R. Woolson, "Wilcoxon signed-rank test," *Wiley Encyclopedia Clin. Trials*, pp. 1–3, Mar. 2007.

[37] P. Morel, "Gramm: Grammar of graphics plotting in MATLAB," *J. Open Source Softw.*, vol. 3, no. 23, pp. 1–4, 2018.

[38] H. X. Hoang, C. Pizzolato, L. E. Diamond, and D. G. Lloyd, "Subject-specific calibration of neuromuscular parameters enables neuromusculoskeletal models to estimate physiologically plausible hip joint contact forces in healthy adults," *J. Biomech.*, vol. 80, pp. 111–120, Oct. 2018.

[39] B. Xiong *et al.*, "Intelligent prediction of human lower extremity joint moment: An artificial neural network approach," *IEEE Access*, vol. 7, pp. 29973–29980, 2019.

[40] A. de Rugy, G. E. Loeb, and T. J. Carroll, "Muscle coordination is habitual rather than optimal," *J. Neurosci.*, vol. 32, no. 21, pp. 7384–7391, May 2012.

[41] S. Safavynia, G. Torres-Oviedo, and L. Ting, "Muscle synergies: Implications for clinical evaluation and rehabilitation of movement," *Topics Spinal Cord Injury Rehabil.*, vol. 17, no. 1, pp. 16–24, Jul. 2011.

[42] L. V. Slater and J. M. Hart, "Muscle activation patterns during different squat techniques," *J. Strength Conditioning Res.*, vol. 31, no. 3, pp. 667–676, 2017.

[43] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1–40, 2016.

[44] A. Kian *et al.*, "The effectiveness of EMG-driven neuromusculoskeletal model calibration is task dependent," *J. Biomech.*, vol. 129, pp. 1–6, Dec. 2021.

[45] T. T. Dao, "From deep learning to transfer learning for the prediction of skeletal muscle forces," *Med. Biol. Eng. Comput.*, vol. 57, no. 5, pp. 1049–1058, 2019.

[46] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[47] W. Yamazaki and Y. Tanino, "Gender differences in joint torque focused on hip internal and external rotation during a change in direction while walking," *J. Phys. Therapy Sci.*, vol. 29, no. 12, pp. 2160–2164, 2017.

[48] G. D. Myer, J. L. Brent, K. R. Ford, and T. E. Hewett, "Real-time assessment and neuromuscular training feedback techniques to prevent anterior cruciate ligament injury in female athletes," *Strength Conditioning J.*, vol. 33, no. 3, pp. 21–35, 2011.