Generalizing Upper Limb Force Modeling With Transfer Learning: A Multimodal Approach Using EMG and IMU for New Users and Conditions

Gelareh Hajian, *Member, IEEE*, Evan Campbell, *Member, IEEE*, Mahdi Ansari,
Evelyn Morin, *Life Senior Member, IEEE*, Ali Etemad, *Senior Member, IEEE*,
Kevin Englehart, *Fellow, IEEE*, and Erik Scheme, *Senior Member, IEEE*

Abstract— In the field of EMG-based force modeling, the ability to generalize models across individuals could play a significant role in its adoption across a range of applications, including assistive devices, robotic and rehabilitation devices. However, current studies have predominately focused on intra-subject modeling, largely neglecting the burden of end-user data acquisition. In this work, we propose the use of transfer learning (TL) to generalize force modeling to a new user by first establishing a baseline model trained using other users' data, and then adapting to the end-user using a small amount of new data (only 10%, 20%, and 40% of the new user data). Using a deep multimodal convolutional neural network, consisting of two CNN models, one with high-density (HD) EMG and one with motion data recorded by an Inertial Measurement Unit (IMU), our proposed TL technique significantly improved force modeling compared to leave-one-subject-out (LOSO) and even intra-subject scenarios. The TL approach increased the average R squared values of the force modeling task by 60.81%, 190.53%, and 199.79% compared to the LOSO case, and by 13.4%, 36.88%, and 45.51% compared to the intra-subject case for isotonic, isokinetic and dynamic conditions, respectively. These results show that it is possible to adapt to a new user with minimal data while improving performance significantly compared to the intra-subject scenario. We also show that TL can be used to generalize on a new experimental condition for a new user.

Index Terms— High-density electromyography, force estimation, deep learning, convolutional neural networks, generalizing EMG-based force modeling, transfer learning, dynamic muscle contraction.

I. INTRODUCTION

ACCURATE electromyography-based force estimation is important for various applications, such as powered exoskeletons, human-robot interaction, rehabilitation systems, and human-machine interfaces (HMI). The surface electromyogram (EMG) has been used for non-invasive neural decoding of end-point force/joint torque since 1952 [1]. This paper primarily focuses on surface EMG. Therefore, all subsequent references to 'EMG' specifically pertain to surface EMG. The majority of work thus far has focused on intra-subject force modeling, where user-specific models are developed using burdensome amounts of user-supplied data [2], [3], [4], [5], [6], [7], [8], [9]. Recently, some studies have pooled data from multiple users to provide more data for deep inter-subject models [10], [11], [12], [13], [14], but the same users used to test the models were included during training. As well, their performance degrades significantly when evaluated on a new user that was not previously seen during model training (cross-user) due to strong inter-user differences in EMG characteristics and behaviours [11].

The current challenge for force estimation, therefore, remains the poor generalizability between users. However, developing a generalized model that performs well in cross-subject evaluation, or adapting an existing model to a new user, is challenging. Increasingly, transfer learning (TL) is becoming an effective deep learning technique that transfers learned knowledge from one problem to a different, but related, problem to improve performance in the new task [15]. TL is commonly used with convolutional neural networks (CNN) to leverage feature extraction kernels learned on readily available data as a baseline, then fine-tuning these weights on a new dataset [16], [17]. This means that the pretrained model is reused as the starting point for a new task.

Recently, TL has been proven to be beneficial in classification-based myoelectric control for model calibration (inter-session) [16], [18], for multi-subject models (subject-independent) [19], adapting a pretrained model to a previously

unseen end-user with minimal data (cross-subject) [20], and success in other EMG-based applications [21]. Ameri et al. demonstrated model calibration using TL, which lessened the performance degradation caused by confounding factors such as electrode shift and physiological parameters changing over multiple sessions [16], [18]. Likewise, Prahm et al. used TL to recalibrate myoelectric pattern recognition models to obtain high accuracy after a significant electrode shift [18]. TL can also be used to form a consensus across multiple subjects and establish a single usable model; despite large differences between subjects and acknowledged challenges in creating subject-independent models [22], [23]. For example, Côté-Allard et al. constructed a single subject-independent model using data from numerous subjects that significantly outperformed the intra-subject scenario for the same cohort of subjects [19]. Campbell et al. then extended this subject-independent scenario to a cross-subject scenario by adapting to a novel end-user with minimal subject-supplied data, even outperforming the intra-subject scenario despite having less end-user data [20]. Likewise, Long et al. have shown this TL strategy was valid outside EMG gesture recognition by outperforming within-subject continuous finger kinematic prediction models using subject adversarial transfer learning [21].

Despite its successful demonstration in EMG pattern recognition for myoelectric control, no previous study has used TL to generalize an EMG-based force estimation model to a new user. We hypothesize that the application of TL to force modeling can leverage data from other users to learn general and informative features via a pre-trained model, before generalizing and fine-tuning to a new user. Leveraging the additional data from other users through TL may therefore enhance a new user's performance while reducing the recording burden for that user (as a smaller portion of data is required for tuning than training).

Consequently, we propose a TL approach to investigate the feasibility of generalizing a force model to a new user using high-density (HD) surface EMG and inertial measurement unit (IMU) motion data during one degree of freedom dynamic elbow flexion and extension. The proposed TL approach is based on a deep multimodal CNN that learns from EMG (CNN_{EMG}) and motion (CNN_{IMU}) signals using twin CNN heads to extract features from each modality individually before being fused. High-density (HD) surface EMG electrodes are used to record EMG signals from the long head and short head of the biceps brachii, brachioradialis, and triceps brachii muscles. Motion data, obtained from a wearable Inertial Measurement Unit (IMU) device mounted on the forearm, and ground-truth force data are collected under quasi-dynamic (controlled force/or controlled velocity) and dynamic (no control on force and velocity) conditions. The reason for utilizing the IMU data along with EMG signals is because we have shown in our previous study [13] that incorporating the kinematic information recorded by IMU considerably contributes to the model's performance to estimate force under quasi-dynamic and dynamic conditions accurately. Our findings from [13] indicated that using only one data source, either EMG or IMU, resulted in a significant decline

in force modeling performance under quasi-dynamic and dynamic conditions. The results of the proposed TL approach, using both EMG and IMU data, show the robustness of our method in comparison to intra-subject modeling and the leave one subject out (LOSO) scenario for different training sizes. We also performed ablation experiments for the TL model to find the best configuration for our problem.

The contributions of this work can be summarized as follows. First, we propose a multi-sensor fusion TL-based solution *to generalize force estimation to a new user* during quasi-dynamic and dynamic contractions. Our solution significantly outperforms naive generalization in a leave one subject out (LOSO) scenario and even the conventional intra-subject approach. Second, we investigate the *effect of training size* on the model's performance. We show that, although increasing the amount of training data improves both the intra-subject and TL schemes, TL significantly outperforms the intra-case for all considered training sizes under all experimental conditions. Finally, we demonstrate the feasibility of leveraging TL to *generalize force modeling to a new experimental condition* (dynamic contraction).

II. EXPERIMENTAL SETUP AND DATA COLLECTION

A. Experimental Setup

Data collection took place at the Human Mobility Research Lab of Queen's University. A total of thirteen participants, comprising 6 females and 7 males, with an average age of 26 ± 9 years, were recruited for this research. The experimental protocol received approval from the Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (HSREB) at Queen's University, and participants provided informed consent before engaging in the study.

A Biodex (model 840-000) [24] was used to control arm motion and measure generated torque. It was configured for controlling elbow flexion and extension of the right arm, allowing participants to perform a series of paired flexion-extension movements. The study encompassed three distinct elbow flexion-extension movement conditions: isotonic-non-isokinetic, isokinetic-non-isotonic, and fully dynamic. Data collection was performed for a single degree of freedom.

For isotonic contractions, the torque remained constant while the movement speed varied, as there were no restrictions on movement velocity. Participants executed isotonic contractions at three different torque levels, applying 5, 8, and 12 Nm of force to the elbow joint. Isokinetic contractions involved three different velocity settings, specifically 60, 90, and 180 deg/s, without any constraints on torque and no minimum required torque. During the fully dynamic condition, participants were unrestricted in terms of applied torque levels and movement velocity, allowing them to move their arm freely with varying levels of velocity and torque. For a more comprehensive understanding of the experimental setup and its procedural details, refer to [25].

B. Data Recording

The EMG data were acquired using a EMG-USB2 HD-system [26], which operated in a referenced monopolar mode.
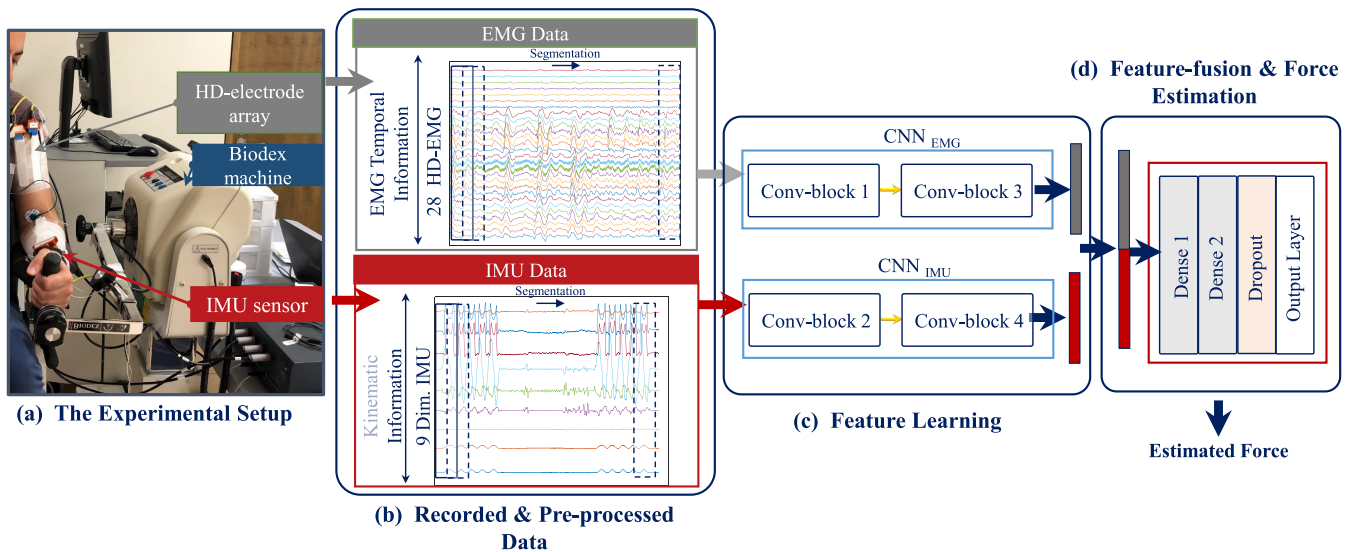
Fig. 1. The experimental setup (a), the network diagram of the proposed deep multimodal CNN is shown from (b): recorded and pre-processed data to (d): feature-fusion to force modelling.

Prior to the placement of electrodes, necessary preparations included shaving the skin (if necessary), followed by a thorough cleaning and abrasion process using an abrasive conductive gel. EMG sensor arrays were affixed to the skin using adhesive pads that featured wells filled with conductive paste, ensuring proper contact with the electrode contacts. EMG signals were recorded from the long and short heads of the biceps brachii, the brachioradialis, and the triceps brachii, using 4 linear HD-electrode arrays with 8 monopolar channels (5 mm spacing). For the biceps, the fourth electrode of each array was placed at the recommended SENIAM location [27]. For the brachioradialis, the fourth electrode was placed at one-third the length of the forearm measured from the elbow. For the long head of the triceps brachii, electrodes were placed at 50% of the distance between the posterior crista of the acromion and the olecranon at 2 finger widths medial to their midpoint. Each electrode array was connected to the EMG-USB2 via an adapter, where each adapter had its own reference electrodes. Standard ECG pre-gelled electrodes with Ag/AgCl contact were used as reference electrodes placed on regions with lower myoelectric activity. For the brachioradialis, the reference electrode was located on the wrist, while for the long and short heads of the biceps and for the triceps brachii they were placed on the elbow and fossa cubit (tendon). A driven right leg (DRL) circuit was used to reduce 60 Hz interference by attaching two reference electrodes on the right and left wrists. EMG signals were recorded with a sampling frequency of 2048 Hz, and were filtered with analog band-pass filters with cut-off frequencies of 10 and 500 Hz.

To track and record the movement of the arm, a Shimmer wearable IMU sensor [28] was placed on the back of the forearm, 4 cm from the location of the ground electrode on the wrist. This location was chosen as it yielded less movement due to muscle contraction during the experiment, which reduced the recording noise due to unwanted movement of the IMU. The IMU has three sensors, namely a triaxial accelerometer, a triaxial gyroscope, and a triaxial magnetometer, all of which were recorded at a 500 Hz sampling rate. The force data were recorded by the Biodex, with a sampling frequency of 1250 Hz.

For each subject, the data were collected in one session. For each quasi-dynamic condition, 12 trials were completed in 2 sets of 6 continuous repetitions with 30 seconds rest between sets. This was repeated for the three isotonic conditions (5, 8, and 12 Nm) and three isokinetic conditions (60, 90, and 180 deg/s), yielding a total of 36 trials for both the isotonic and the isokinetic cases. Another 36 trials were repeated for the dynamic condition (3 sets of 12 trials) to ensure the same number of trials across conditions. Appropriate rest periods - at least 10 minutes and more if needed - were provided between conditions to avoid muscle fatigue. More details on the data recording can be found in [25]. Figure 1 (a) and (b) shows the experimental setup, including the subject seated in the Biodex machine, the EMG-USB2 HD-system, the HD-electrodes, the IMU sensor.

C. Data Pre-Processing

The torque signals, initially sampled at 1250 Hz by the Biodex, underwent up-sampling through linear interpolation to reach a frequency of 2048 Hz, aligning with the EMG's sampling rate. Likewise, the IMU data underwent a similar up-sampling process, transitioning from 500 Hz to 2048 Hz. Then, differential HD-EMG signals were obtained by subtracting neighbouring channels longitudinally along each 8-channel array, resulting in 7 differential channels from each array. Each differential channel was further band-pass filtered with cut-off frequencies of 10 Hz and 500 Hz using an eighth-order Butterworth filter. The Biodex data were smoothed using a 300-point (146.48 ms) moving average filter. The IMU data were low-pass filtered using a Savitzky-Golay filter, with a window length of 400 points (195.31 ms). The data were then segmented, where the segment lengths were set to 50 ms with an overlap of half of the segment length, based on our previous findings [25]. Before processing the data obtained

during muscle contractions, any intervals corresponding to rest periods between sets were excluded from analysis.

Data normalization was carried out using the standardization approach to maintain a uniform scale and optimize model training. We first determined the mean and standard deviation of each channel of data (for both EMG and IMU) within the training dataset. The training data was then standardized using these values. Notably, when standardizing the test data, we utilized the statistics derived only from the training set to prevent potential data leakage. This approach ensures that our model is trained on harmonized data, and the test set evaluation remains unbiased. More information regarding the train and test data will be explained in the following section.

III. METHODS

A. Transfer Learning for Force Modeling Generalization

In order to generalize the results of force modeling to the new user, a model must first be developed using a cohort of users, before fine-tuning it on a new user. Pretraining on the full dataset from a cohort of users is done to capture informative features that are consistent across subjects so as to develop a good base model for the unseen subject. Then, through the fine-tuning process using a small amount of data from the end-user, the weights of certain layers of the model are tuned to extract optimal subject-specific features.

The force generated by muscles can be characterized by the EMG from which it is produced, as well as the resulting motion. To take full advantage of both sources of information, we employed a multimodal approach that uses both EMG and IMU data. As shown in our previous work [13], incorporating the kinematic information recorded by IMU considerably improves model performance when estimating force under quasi-dynamic and dynamic conditions. A base model was trained using EMG (28 channels) and motion signals (9 channels, obtained from a wearable IMU device) extracted from the cohort of training subjects. For the TL procedure, we then tuned the parameters of the base model to personalize it to a previously unseen end-user for subject-specific force modeling. In the following subsections III-A.1 and III-A.2, we explain the details of the model development process.

1) Model Architecture: CNNs are extensions of standard neural networks, originally proposed for the analysis of image and video data as they are capable of dealing with high-dimensional raw data without the need for manual feature extraction. They are used in a variety of applications related to biological signals and have been previously employed for force modelling [10], [25]. CNNs are typically made up of several types of layers: convolutional layers, batch normalization, non-linear activation functions (rectified linear unit (ReLU)), and max-pooling layers to reduce the dimensionality of the feature map, decrease computation, and to help avoid overfitting. The use of these layers for automatic feature extraction from input data in CNNs is referred to as a conv-block here. Therefore, each conv-block shown in Figure 1 consists of convolutional layers, batch normalization, ReLu activation functions, and max-pooling layer.

An overview of the proposed deep multimodal CNN framework for force modelling is shown in Figure 1 (b)-(d). The pre-processed and segmented EMG and IMU data are used as inputs to separate CNNs (CNN_{EMG} and CNN_{IMU}). For the CNN_{EMG}, inputs consisted of the preprocessed and segmented EMG data (28 differential signals). The input layer for the CNN_{IMU} was fed with the pre-processed and segmented IMU data recorded by the triaxial accelerometer, gyroscope and magnetometer (9 channels in total). The size of the input layers for CNN_{EMG} and CNN_{IMU} were thus (102×28) and (102×9), respectively. During training, a dropout layer was used at the end of each CNN to reduce the chances of overfitting to the training data. Each CNN stream was used to learn features from its respective inputs (feature learning block). Next, the feature fusion block concatenated the learned features to obtain a multimodal feature encoding of both EMG and IMU. This was then followed by fully connected (dense) layers, which acted as a shallow neural network to weight the obtained features, and a regression layer to estimate the induced force at the wrist. Two dense layers with 100 and 200 neurons, respectively, were used. The output of the network was used for all experimental conditions (isotonic, isokinetic, and dynamic contractions).

The network architecture was largely influenced by our past work on this dataset pursuing intra- and inter-subject force modelling [13], where, we determine the best number of neurons per layer in the architecture empirically (grid search over convolutional layer parameters, fully connected parameters, and window sizes). Here, a number of hyper-parameters for each CNN_{EMG} and CNN_{IMU} were tuned to achieve the best results using a validation set of the cohort data, and closely resembled the optimal hyperparameters for the inter-subject prior work [13]. The tuned hyperparameters include: the number of conv-blocks, the number of filters and their sizes for each convolutional layer, batch sizes, the number of training epochs, and dropout rates. The values for these parameters obtained for each CNN are presented in Table I. The batch size was set at 64, as smaller sizes led to extended training times without enhancing force modelling performance, and larger sizes adversely affected the performance. We chose 150 epochs because increasing the number did not further improve performance and only lengthened the training time, while fewer epochs diminished the performance. Visual inspection of the training process supported this choice, using the validation set to ensure that adequate training had been performed. We verified underfitting had not occurred by selecting a number of epochs for training that the validation loss and training loss generally decreased. Likewise, We verified overfitting did not occur for our selected number of epochs by ensuring validation loss did not begin to increase while training loss continued to decrease. The selected dropout rate for all base learners was 0.5. To avoid overfitting, $L2$ regularization was used. Additionally, dropout was performed after the final conv-block for both CNN_{EMG} and CNN_{IMU}.

The proposed architecture and all analyses were implemented in Python using Keras with a TensorFlow backend, on an Intel(R) Core(TM) M-5Y10c CPU @ 0.80GHz, 998 MHz processor, with 8 GB of RAM. The models were trained with the adaptive moment estimation (ADAM) optimizer to update the network weights during back-propagation because it has

	No. Conv.	No. Filters	Filter Size	Maxpool
CNN_{EMG}	$m = 2$	64, 128	2×2	3×3
CNN_{IMU}	$p = 2$	64, 128	2×2	2×2

been proven to be efficient in computing the stochastic gradient problem and avoiding local minima [29]. Mean squared error loss was used for determining the accuracy of the predictions and to perform back-propagation. A learning rate (l_r) of 0.001 was used with exponential decay rates for the first and second movement estimates of $\beta 1 = 0.9$, and $\beta 2 = 0.999$, respectively. The smaller learning rate was used for the TL model fine-tuning (0.00001). Early stopping with 15 patience, and min $\delta = 0.5$ was used.

2) Transfer Learning Fine-Tuning: As the goal of this study was to perform force modelling on a new, previously unseen user, the base model explained in III-A.1 was then fine-tuned through a TL procedure to tailor the developed model to the new user. While tailoring the model to the new user, the model parameters that are adapted can be selectively chosen. In this regard, an ablation experiment was conducted to determine the optimum number of conv-blocks to freeze (remain unaltered) during the TL procedure to obtain the best force modelling performance on a new user. For the first case studied, we started by unfreezing the entire base model, which allowed all of the parameters of the model to be tuned (all conv-blocks and linear layers). As a second case, we froze the first two conv-blocks; conv-block 1, the first conv-block for EMG feature extraction, and conv-block 2, the first conv-block for IMU feature extraction (as shown in Figure 1), and tuned the parameters of the rest of the model for the TL procedure. Finally, in the third case, we froze all of the conv-blocks and only tuned the parameters of the remaining four linear layers, shown in the red box of Figure 1 (d). The selection of which blocks are frozen controls how much of the original base model learning is persisted in the final TL-tuned model.

B. Evaluation

The performance of the TL model was compared with a naive LOSO approach and the conventional intra-subject approach. As in the TL method, base model training for LOSO is conducted with all but one user, but no fine-tuning is conducted on the end-user. For intra-subject performance, a model was trained using only the end-user's own data, as is conventionally done in EMG force modelling. The LOSO scheme was meant to emulate a lower-benchmark of performance where the end-user training burden is minimal, but performance is known to be poor. In contrast, the intra-subject scheme was meant to simulate a competitive-benchmark of performance where the end-user training burden is the same as with the TL model, but it does not leverage any pretraining from the cohort of other users.

In order to determine if the TL approach was beneficial for EMG force modelling, its efficacy was compared against the established LOSO and intra-subject approaches within the isotonic, isokinetic, and dynamic conditions, independently. For each assessment, the dataset was first split into its condition, then further split into a *cohort dataset* (data from all users except the test user) and an *end-user dataset* (the data from the subject being held-out for testing). The final contiguous 50% of the end-user dataset was used as the testing dataset for all analyses to establish a fair comparison across the TL, LOSO, and intra-subject approaches. Details surrounding the portion of data used by each approach are given below:

- **Intra-subject**: From the residual 50% of the end-user dataset, $\alpha\%$ was used for the training set, and 10% was used for the validation set. To evaluate how much end-user training burden was necessary, three values of α were tested: 0%, 20%, and 40%; where using 40% of the end-user data resulted in all end-user data being involved in the analysis.
- **Leave-One-Subject-Out (LOSO)**: From the cohort dataset, 90% was used for the training set and the remaining 10% was used for the validation set. No end-user data was used in the training of this approach.
- **Transfer Learning (TL)**: The cohort dataset was used the same way as the LOSO approach to establish the base model. Afterwards, the learning rate was lowered, and $\alpha\%$ of the remaining end-user data was used for the training set of the fine-tuning and 10% for the validation set, and the same 50% of the end-user dataset as the intra-subject was used for testing. Similar to the intra-subject analyses, α was tested for values: 10%, 20%, and 40%.

Importantly, although the training and validation sets for the intra-subject training and TL fine-tuning were randomly selected from non-testing set end-user data, these selections of data were identical to ensure fairness. Additionally, since our study had three experimental conditions–namely isotonic, isokinetic, and dynamic–and under the isotonic and isokinetic categories, three respective cases were presented (three force levels for isotonic and three velocity levels for isokinetic). We ensured that each data portion (train/validation/test) consistently represented each case, meaning that they originated from identical experimental conditions while maintaining a consistent percentage across each case.

Performance was evaluated using the coefficient of determination (R^2), which is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=0}^{N}(F_i^{Est} - F_i)^2}{\sum_{i=0}^{N}(F_i - \overline{F_i})^2}, \tag{1}$$

where N is the number of data samples, F_i is the ith measured force sample, F_i^{Est} is the corresponding estimated force, and $\overline{F_i}$ is the average of F_i. The numerator in the second term of the equation is the total mean squared error (MSE) of the estimates, whereas the denominator is the total variance of the force.

C. Statistical Analysis

Statistical analysis was performed on the R^2 values using the Wilcoxon signed rank test, a nonparametric statistical test, as our evaluation metrics were not normally distributed. The
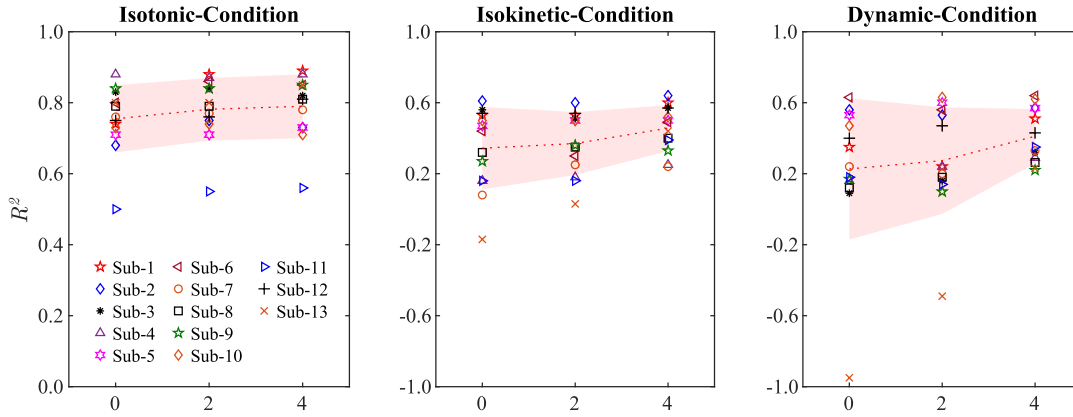
Fig. 2. R^2 values for 13 users, using transfer learning with different frozen layers: 0 (no conv-block was frozen), 2 (the first EMG conv-block and the first IMU conv-block were frozen), and 4 (all conv-blocks were frozen). The dashed line shows the mean R^2 values across all users, while the pink shading represents the standard deviation across subjects for each model, under the different experimental (isotonic, isokinetic, and dynamic) conditions.

null hypothesis was rejected when the p-value was below the critical value, which is set to 0.05. The significance level was adjusted using Bonferroni correction when multiple pairwise tests were conducted. Statistical analysis were conducted to evaluate the performance of the of TL approach relative to the naive LOSO case and traditional intra-subject modelling. The obtained R^2 values were compared between methods in a pairwise fashion, using MATLAB (MATLAB 19.1, The MathWorks Inc.).

IV. RESULTS AND DISCUSSION

In this section, we provide an analysis of the impact of various aspects of the TL procedure, such as the effect of unfreezing the conv-blocks during transfer, the amount of training data used, and the impact of end-user supplied training data for fine-tuning. We then present the performance of the proposed TL model in conducting force estimation compared with the naive LOSO approach (with no model fine-tuning) and traditional intra-subject force modelling. We also investigate the feasibility of using TL to generalize force modeling to a new user under a new experimental condition for that user.

A. Ablation Study

An ablation study was conducted that first determined the optimum number of conv-blocks to tune for the TL procedure following the cases outlined in section III-A.2. For this analysis, the base models were trained as described in III-B, 20% of the data from a new user was used to fine-tune the TL model, 10% was used for validation, and finally, 50% was used for testing. Figure 2 shows the R^2 values obtained when freezing different numbers of conv-blocks for fine-tuning. The results indicated that freezing all four conv-blocks resulted in significantly higher performance, for isokinetic ($p < 0.0039$) and dynamic ($p < 0.0037$) compared to freezing two or no conv-blocks. For the isotonic case, there was no significant difference between freezing all four conv-blocks and freezing the first two conv-blocks, although freezing all conv-blocks was significantly better than freezing none them ($p = 0.0078$).

Freezing all conv-blocks was also much better in terms of reducing computational time, as gradients didn't need to flow back through those layers.

Therefore, in our TL approach, freezing all four conv-blocks in the model led to enhanced performance, primarily because these layers, pre-trained on a cohort of users, already captured robust, generalized features essential for force modelling. By keeping these layers frozen, we effectively prevented overfitting, a crucial consideration given the potentially limited data available for each new user. This strategy ensured the stability and relevance of the features extracted, allowing the model to retain the generalized knowledge acquired from the broader cohort while focusing the training process on fine-tuning the dense layers. These layers, more adaptable to individual variations, were specifically tuned to the characteristics of the new user. This approach not only optimized the model's performance by maintaining a balance between knowledge retention and adaptation to individual specificity but also improved computational efficiency, a significant factor in real-time applications. Thus, for the rest of the analysis presented here, TL was completed by freezing all four conv-blocks, and only the last four linear layer parameters were updated (fine-tuned) using end-user data, aimed at achieving the best balance between general feature extraction and individualized force estimation.

B. Performance and Comparison

In Figure 3, the performance of the TL, LOSO, and intra-subject approaches are shown for all experimental conditions. Again, the TL procedure used 20% of the end-user data to fine-tune the base model, 10% for validation, and 50% for testing. The intra-subject model used the same data, but trained a subject-specific model using the 20% end-user data (instead of fine-tuning the base model). The LOSO directly used the base model without fine tuning, but used the same 50% for testing. The dashed line in Figure 3 shows the average R^2 values across all subjects, and the pink shading shows the standard deviation across subjects for each method under the different experimental conditions (isotonic, isokinetic, and
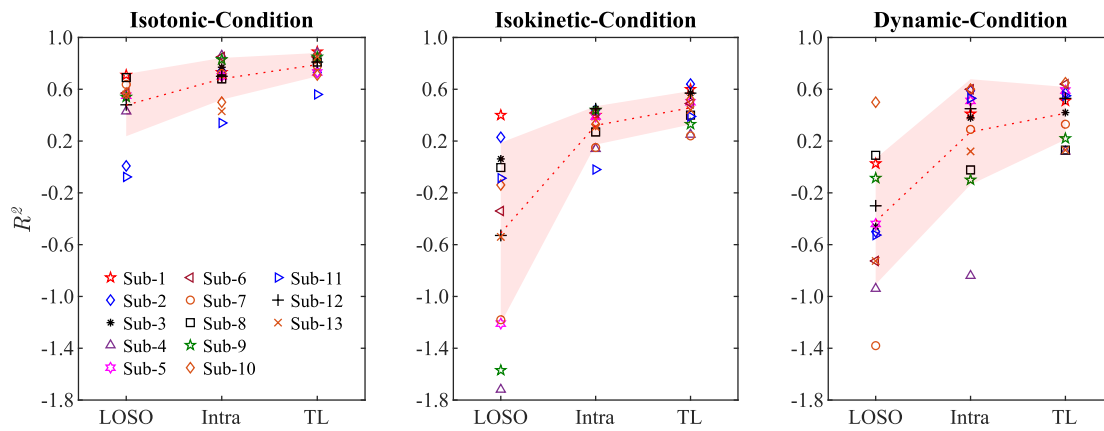
Fig. 3. R^2 values for the 13 users, using transfer learning compared to the LOSO and Intra-subject force modelling. The dashed line shows the mean R^2 values across all users, while the pink shading represents the standard deviation across subjects for each model, under different experimental conditions, isotonic, isokinetic, and dynamic. LOSO stands for leave on subject out, Intra: intra-subject modelling, TL: transfer learning.
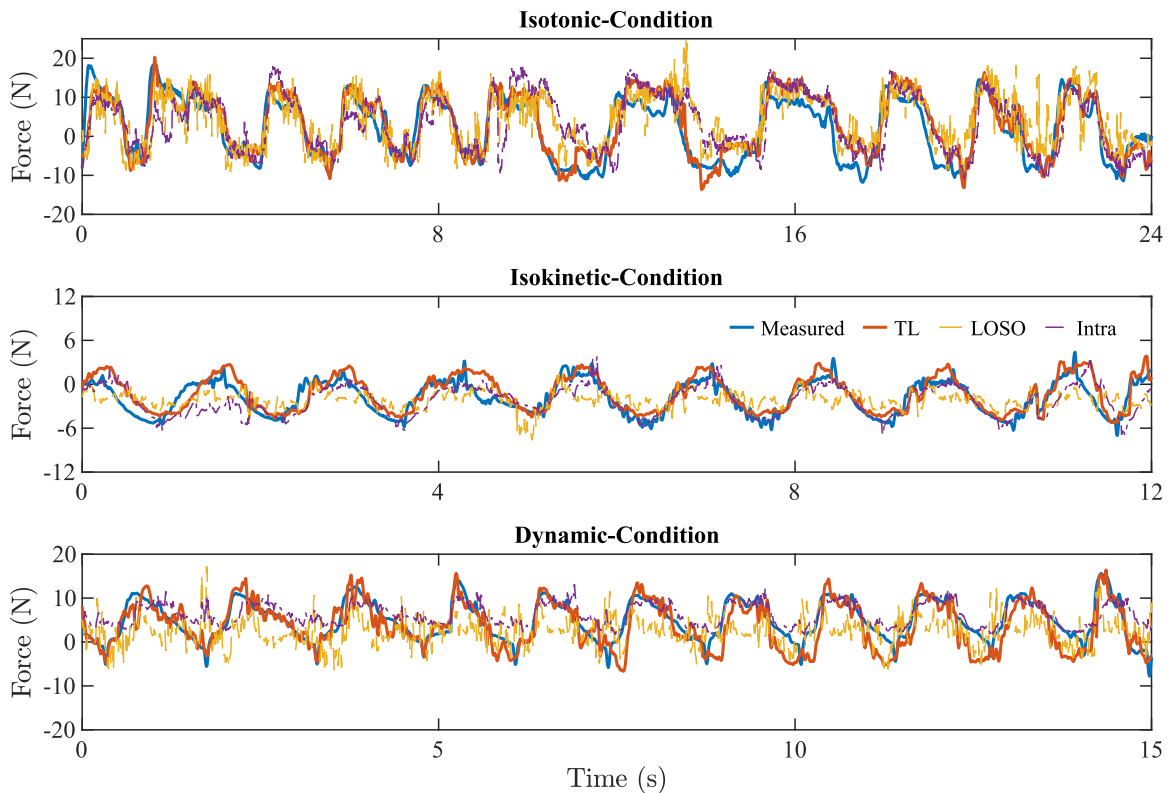

Fig. 4. The estimated versus the measured force using TL, LOSO and Intra, for one user.

dynamic). Pairwise comparisons were performed using the Wilcoxon test on the R^2 values of 13 participants, for each condition (isotonic, isokinetic, and dynamic) separately. The experimental results show that intra-subject performance is significantly better than LOSO for all experimental conditions as $p < 0.0024$. More importantly, TL achieves significantly better performance than both LOSO and intra-subject for all experimental conditions, with posthoc p-values of $p < 0.00024$ for all conditions of TL versus LOSO and $p < 0.0017$ for all conditions of TL versus intra. The TL result demonstrates that the TL approach can not only increase the performance of LOSO force estimation by fine-tuning to the

end-user, but that it can effectively leverage data from other users to outperform within-subject training alone.

Figure 4 shows examples of the measured force vs. estimated force using the TL, LOSO, and intra cases, for one subject under the different experimental conditions. The TL approach is clearly able to estimate the force for a new user more accurately in each of the different experimental conditions than the other approaches. The R^2 values for the data shown in Figure 4 are TL: 0.71, LOSO: 0.44, and intra: 0.58 (for isotonic), TL: 0.51, LOSO: 0.15, and intra:0.34 (for isokinetic), and TL: 0.65, LOSO: 0.097, and intra: 0.51 (for dynamic).

1) Effect of Experimental Conditions on Model Performance:
We also conducted a statistical analysis to investigate the effect of experimental conditions on the performance of the TL model. Specifically, we compared the R^2 values of three experimental conditions: isotonic, isokinetic, and dynamic contractions. Our results indicate that the TL model performed significantly better for the isotonic condition compared to the isokinetic ($p = 0.0002$) and dynamic ($p = 0.0002$) conditions. However, there was no significant difference in the TL model's performance between the isokinetic and dynamic conditions ($p = 0.5$).

Additionally, we compared the intra-subject force modeling approach across different conditions and found that isotonic performance was better than isokinetic ($p = 0.0002$) and dynamic ($p = 0.002$) conditions, with no difference between isokinetic and dynamic conditions ($p = 0.9$). For the LOSO scenario, we found that isotonic performance was statistically better than isokinetic ($p = 0.0007$) and dynamic ($p = 0.0002$) conditions, while dynamic performance was not significantly better than isokinetic. The higher performance under isotonic conditions may stem from the relatively consistent force over time and across participants, in contrast to the inconsistency observed in isokinetic and dynamic scenarios. As highlighted in previous research [13], the variability in force levels during isokinetic and dynamic conditions plays a more significant role in performance reduction compared to velocity variability seen in isotonic and dynamic conditions. This is because the inclusion of IMU data makes the model more robust to changes in velocity, as the IMU can track the varying speeds of the limb. However, in the dynamic and isokinetic conditions, we rely on EMG to account for the variability in the force.

C. Impact of Fine-Tuning Data Amounts on Performance

Although deep learning benefits from more training data, the amount of data required from an end-user for EMG-based force modelling should be minimized for practicality and user convenience. To help overcome this conflict, in this study, the multimodal deep CNN base model was pre-trained using a larger set of data recorded from other users. This base model was then fine-tuned through TL using some amount of end-user data.

To investigate the impact of the amount of data available from the end-user for fine-tuning on TL performance, we considered using 10%, 20%, and 40% of the available end-user data for each experimental condition. We also compare the TL results with the intra-subject approach, using the same amount of data to directly train a new subject-specific model. In all cases, the validation and test sizes were kept consistent, with 10% used for validation, and 50% for testing.

Figure 5 shows the impact of training size on the R^2 for the different experimental conditions, for both the TL and intra-subject approaches. As shown in Figure 5, increasing the amount of data used to fine-tune the TL and train the intra-subject models improves the performance of both. For all experimental conditions, however, the performance of the TL approach is significantly higher than that of the intra-subject scheme when using the same amount of end-user data, with p values of $p < 0.01$, $p < 0.0049$, and $p < 0.02$ for isotonic,
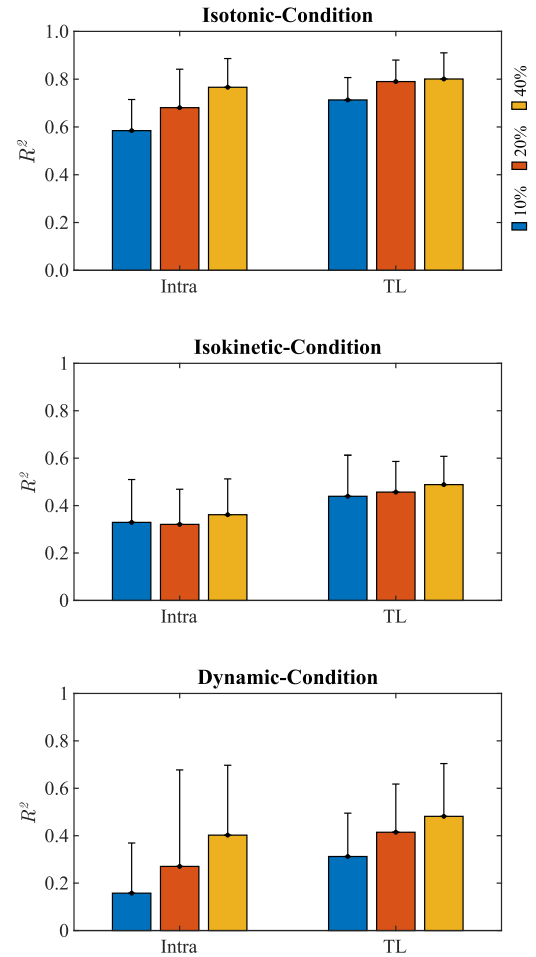

Fig. 5. The average and standard deviation of R^2 values across all users, using transfer learning (TL) compared to Intra-subject (Intra) force modelling, for different amounts of end-user data used for fine-tuning and training (10%, 20%, and 40%), respectively. Results are shown for the different experimental conditions, isotonic, isokinetic, and dynamic.

isokinetic, and dynamic, respectively. Furthermore, increasing the training size from 10% to 40% significantly improves the TL performance only for the isotonic condition, while for the intra-subject scheme, the performance improved significantly for all experimental conditions. This indicates the need for more training data to develop a model from scratch for a new user, whereas when using the TL approach, the model can benefit from the learned features from other users.

1) Effect of Experimental Conditions on Model Performance for Different Amounts of Training Data: We investigated the impact of experimental conditions on the performance of the TL for different amounts of data available for fine-tuning: 10%, 20%, and 40%. Our statistical analysis indicates that the TL model performed significantly better for the isotonic condition compared to the isokinetic ($p < 0.0007$) and dynamic ($p < 0.0012$) conditions for all considered training sizes. However, there was no significant difference in performance between the isokinetic and dynamic conditions ($p > 0.05$).

In the intra-subject case, we found that isotonic condition performance was significantly better than the other conditions, including isokinetic ($p < 0.005$) and dynamic ($p < 0.002$).

When comparing isokinetic versus dynamic, there were no significant differences in performance except for the 10% training size ($p = 0.0017$). Despite variations in training size, our findings demonstrate that both TL and intra-subject modeling exhibit higher performance levels for the isotonic condition compared to the isokinetic and dynamic conditions.

The amount of data required for the TL approach to reach a plateau in performance was lower for the isotonic condition compared to the isokinetic and dynamic conditions, as only a negligible improvement was found when doubling the amount of data for fine-tuning from 20% to 40% for isotonic conditions. This early plateau was not found in the intra-subject models for the isotonic condition. Given that the LOSO approach alone did not outperform intra-subject models for the isotonic condition (Figure 3), this suggests that the early plateau was in response to both the isotonic nature of the motions, and having access to the cohort's data. We suspect this plateau indicates that when the user is performing variable speed motions, the TL model was better able to leverage the modality that contributes kinematic information and does not have subject-specific variability (IMU). Given that the isokinetic and dynamic conditions did not plateau, this indicates that when users perform force varying motions, the TL model was able to leverage information from the cohort to outperform the intra-subject model; however, because of the subject-specific nature of EMG signals more end-user data was required. This relationship indicates that strategies that attempt to regularize out subject-specific variability would be beneficial for EMG force modelling in isokinetic and dynamic settings to improve performance while minimizing training burden.

D. Generalizability to Dynamic Condition

Finally, we investigated the feasibility of using TL to not only generalize force modeling to a new user for a given condition but also to generalize to a new experimental condition for that user as well. To do so, we trained the base TL model using all of the quasi-dynamic data from a cohort of training users (from the isotonic and isokinetic conditions) and fine-tuned it using the same isotonic and isokinetic data from the end-user, with 80% used for training and 20% for validation. We then tested it using the fully dynamic condition data from the end-user. We compared this to the intra-subject and LOSO schemes. In the intra-subject case, using the same end-user data availability, a new subject-specific model was trained using the isotonic and isokinetic condition data for that user, and tested using their dynamic data. For the LOSO, data from other users during the quasi-dynamic conditions were used to train the model, and testing was conducted on the new user. Figure 6 shows the R^2 values of the TL, intra-subject, and LOSO schemes, for each participant. The dashed line in Figure 3 shows the average R^2 value and the pink shading shows the standard deviation across all subjects for each method, TL, intra and LOSO. Leveraging TL significantly outperformed intra-subject force modeling, with average R^2 values for TL of 0.29, intra-subject models of -0.12, and -0.73 for LOSO. TL resulted in significantly better performance compared to the intra- ($p = 0.00073$) and
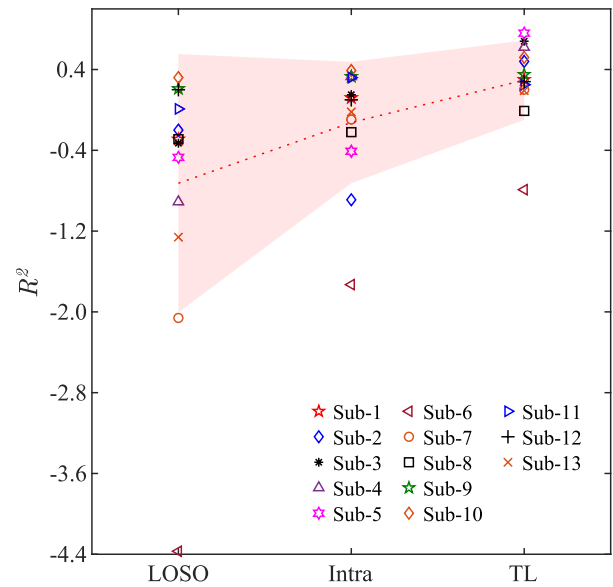

Fig. 6. R^2 values for the 13 users, using transfer learning compared to the LOSO and Intra-subject force modelling for generalizing the model to the dynamic condition. The dashed line shows the mean R^2 values across all users, while the pink shading represents the standard deviation across subjects for each model.

LOSO ($p = 0.00024$) cases. No significant difference was found between the performance of intra- and LOSO when generalizing to a new experimental condition.

V. CONCLUSION AND FUTURE WORK

The purpose of this study was to explore the potential benefit of leveraging transfer learning to reduce the training burden on an end-user in EMG-based wrist force modeling. A novel state of the art TL approach was compared against a naive LOSO approach and the conventional intra-subject approach for isotonic, isokinetic, and fully dynamic elbow flexion and extension conditions. The proposed TL method, using a deep multimodal CNN pipeline as a base model which extracted features from EMG and IMU data from multiple users, and then tuned its parameters to personalize the model for a new user, using a small portion of data. The TL approach outperformed LOSO and intra-subject force modeling methods for all experimental conditions. Our results confirmed the effectiveness of using the TL approach over developing a model from scratch for a new user, for different training sizes. We also showed TL required less data than intra-subject models and achieves significantly higher performance across all experimental conditions. Further, while marginal improvements can be made by including more end-user data, TL was found to not significantly improve by including more than 10% new user data for most conditions, indicating a viable approach with minimal training burden. We also showed that TL can be used to generalize even to a new experimental condition for a new user. Thus, it has great potential for generalizing a developed model not only to a new user, but also to a new experimental condition, which is highly needed in force modeling applications, such as assistive devices, robotics, and rehabilitation.

In order to assess the robustness and generalizability of the proposed method for more practical applications, such

as assistive devices or rehabilitation, the effectiveness of the TL approach should be evaluated with larger datasets, including individuals with disabilities. As this study focused on a relatively small and healthy number of participants, conducting further research with a more diverse population would enable more comprehensive conclusions. We recognize that this collection was constrained where users performed contractions along only a single degree of freedom, and future work should explore less enforced movements with more irregular patterns. Also, further research should explore settings with larger training burdens such as predicting force for multiple degrees of freedom (including wrist flexion/extension alongside elbow flexion/extension) and surveying more muscles (flexor carpi radialis, extensor carpi ulnaris) to assess whether TL can provide greater benefit under greater challenges.

REFERENCES

[1] V. T. Inman, H. J. Ralston, J. B. De C. M. Saunders, M. B. Bertram Feinstein, and E. W. Wright, "Relation of human electromyogram to muscular tension," *Electroencephalogr. Clin. Neurophysiology*, vol. 4, no. 2, pp. 187–194, May 1952.

[2] D. Staudenmann, K. Roeleveld, D. F. Stegeman, and J. H. van Dieën, "Methodological aspects of SEMG recordings for force estimation—A tutorial and review," *J. Electromyogr. Kinesiol.*, vol. 20, no. 3, pp. 375–387, Jun. 2010.

[3] F. Bai and C.-M. Chew, "Muscle force estimation with surface EMG during dynamic muscle contractions: A wavelet and ANN based approach," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 4589–4592.

[4] G. Hajian, A. Etemad, and E. Morin, "Automated channel selection in high-density sEMG for improved force estimation," *Sensors*, vol. 20, no. 17, p. 4858, Aug. 2020.

[5] C. Dai, B. Bardizbanian, and E. A. Clancy, "Comparison of constant-posture force-varying EMG-force dynamic models about the elbow," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 9, pp. 1529–1538, Sep. 2017.

[6] G. Hajian, E. Morin, and A. Etemad, "PCA-based channel selection in high-density EMG for improving force estimation," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 652–655.

[7] J. Luo, C. Liu, and C. Yang, "Estimation of EMG-based force using a neural-network-based approach," *IEEE Access*, vol. 7, pp. 64856–64865, 2019.

[8] F. Mobasser and K. Hashtrudi-Zaad, "A comparative approach to hand force estimation using artificial neural networks," *Biomed. Eng. Comput. Biol.*, vol. 4, Jan. 2012, Art. no. BECB.S9335.

[9] G. Hajian, E. Morin, and A. Etemad, "Convolutional neural network approach for elbow torque estimation during quasi-dynamic and dynamic contractions," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 665–668.

[10] G. Hajian, A. Etemad, and E. Morin, "Generalized EMG-based isometric contact force estimation using a deep learning approach," *Biomed. Signal Process. Control*, vol. 70, Sep. 2021, Art. no. 103012.

[11] L. Xu, X. Chen, S. Cao, X. Zhang, and X. Chen, "Feasibility study of advanced neural networks applied to sEMG-based force estimation," *Sensors*, vol. 18, no. 10, p. 3226, Sep. 2018.

[12] G. Hajian, E. Morin, and A. Etemad, "EMG-based force estimation using artificial neural networks," in *Proc. Can. Med. Biol. Eng. Conf.*, vol. 42, 2019, pp. 1–4.

[13] G. Hajian, E. Morin, and A. Etemad, "Multimodal estimation of endpoint force during quasi-dynamic and dynamic muscle contractions using deep learning," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[14] G. Hajian, B. Behinaein, A. Etemad, and E. Morin, "Bagged tree ensemble modelling with feature selection for isometric EMG-based force estimation," *Biomed. Signal Process. Control*, vol. 78, Sep. 2022, Art. no. 104012.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[16] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "A deep transfer learning approach to reducing the effect of electrode shift in EMG pattern recognition-based control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 370–379, Feb. 2020.

[17] U. Côté-Allard et al., "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019.

[18] C. Prahm, B. Paassen, A. Schulz, B. Hammer, and O. Aszmann, "Transfer learning for rapid re-calibration of a myoelectric prosthesis after electrode shift," in *Converging Clinical and Engineering Research on Neurorehabilitation II* (Biosystems & Biorobotics), vol. 15, J. Ibáñez, J. González-Vargas, J. Azorín, M. Akay, and J. Pons, Eds. Cham, Switzerland: Springer, 2017, doi: 10.1007/978-3-319-46669-9_28.

[19] U. Côté-Allard, E. Campbell, A. Phinyomark, F. Laviolette, B. Gosselin, and E. Scheme, "Interpreting deep learning features for myoelectric control: A comparison with handcrafted features," *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 158, Mar. 2020.

[20] E. Campbell, A. Phinyomark, and E. Scheme, "Deep cross-user models reduce the training burden in myoelectric control," *Frontiers Neurosci.*, vol. 15, p. 595, May 2021.

[21] Y. Long, Y. Geng, C. Dai, and G. Li, "A transfer learning based cross-subject generic model for continuous estimation of finger joint angles from a new user," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 4, pp. 1914–1925, 2023.

[22] T. S. Saponas, D. S. Tan, D. Morris, and R. Balakrishnan, "Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2008, pp. 515–524.

[23] E. Campbell, J. Chang, A. Phinyomark, and E. Scheme, "A comparison of amputee and able-bodied inter-subject variability in myoelectric control," 2020, *arXiv:2003.03481*.

[24] *Biodex*. Accessed: Sep. 6, 2020. [Online]. Available: https://www.biodex.com/

[25] G. Hajian and E. Morin, "Deep multi-scale fusion of convolutional neural networks for EMG-based movement estimation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 486–495, 2022.

[26] *Ot Bioelettronica*. Accessed: Sep. 6, 2020. [Online]. Available: https://www.otbioelettronica.it/en/

[27] H. J. Hermens, B. Freriks, C. Disselhorst-Klug, and G. Rau, "Development of recommendations for SEMG sensors and sensor placement procedures," *J. Electromyogr. Kinesiol.*, vol. 10, no. 5, pp. 361–374, Oct. 2000.

[28] *Shimmer IMU*. Accessed: Sep. 6, 2020. [Online]. Available: http://www.shimmersensing.com/products/shimmer3-imu-sensor

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.