2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Utility of Inter-subject Transfer Learning for Wearable-Sensor-Based Joint Torque Prediction Models

Jennifer Sloboda[1], Paul Stegall[2], Ryan J. McKindles[1], Leia Stirling[3], Ho Chit Siu[1]

Abstract— **Generalizability between individuals and groups is often a significant hurdle in model development for human subjects research. In the domain of wearable-sensor-controlled exoskeleton devices, the ability to generalize models across subjects or fine-tune more general models to individual subjects is key to enabling widespread adoption of these technologies. Transfer learning techniques applied to machine learning models afford the ability to apply and investigate the viability and utility such knowledge-transfer scenarios. This paper investigates the utility of single- and multi-subject based parameter transfer on LSTM models trained for "sensor-to-joint torque" prediction tasks, with regards to task performance and computational resources required for network training. We find that parameter transfer between both single- and multi-subject models provide useful knowledge transfer, with varying results across specific "source" and "target" subject pairings. This could be leveraged to lower model training time or computational cost in compute-constrained environments or, with further study to understand causal factors of the observed variance in performance across source and target pairings, to minimize data collection and model retraining requirements to select and personalize a generic model for personalized wearable-sensor-based joint torque prediction technologies.**

I. INTRODUCTION

Wearable physiological sensors may support decision-making, training, and device control across a variety of clinical, health monitoring, fitness, and wearable robotics tasks [1], [2], [3], [4]. The ability to generalize physiological sensing models between individuals is often a significant hurdle in model development for more widespread adoption in these domains. Among these use cases, wearable robotics (e.g. exoskeletons) have stricter requirements for temporal granularity and specificity of state sensing than other types of activity monitoring, given the need for real-time detection of user movement and intent [5]. For this work, we focus specifically on the requirements for wearable sensing models for the wearable robotics and other real-time joint dynamics estimation use cases.

[1]Lincoln Laboratory, Massachusetts Institute of Technology, 244 Wood Street, Lexington, MA 02421, USA {jennifer.sloboda, ryan.mckindles, hochit.siu}@ll.mit.edu

[2]Department of Aeronautics and Astronautics, Massachusetts Institute of Technology 77 Massachusetts Avenue, Cambridge, MA 02139, USA stegall@mit.edu

[3]Department of Industrial and Operations Engineering, Robotics Institute, University of Michigan 1205 Beal Avenue, Ann Arbor, MI 48109, USA leias@umich.edu

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This material is based upon work supported by the United States Air Force under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force.

In this context, the issue of model transferability and generalizability is encountered on several dimensions: Does a model developed with data from a single subject in one data collection session generalize to another sensor placement/data collection session? Can a model from one subject be applied to another? Can a population-based model be applied to individuals? What is the minimum amount of data and computational resources required to train a model for a single subject? Can a "general model" with minimal subject-specific fine-tuning be developed? These questions directly relate to the practicalities of experimentation with and commercialization of such systems. For example, an exoskeleton device that could be manufactured and shipped to consumers with a general model that then requires a short amount of wearer-specific retraining, with compute resources available on-board or on an at-home accessible cloud-based device, enables distribution of such a technology to large populations. Similar concerns apply to experimentation during research or product development phases, and to adoption into clinical usage: minimizing training time allows for more subjects and different experiments to be run, and understanding the limitations of such techniques are key.

Transfer learning techniques from machine learning can be leveraged to transfer knowledge between individual- and/or group-based models. The rest of this paper outlines a set of experiments which investigate the viability and utility of parameter-based transfer learning for neural network-based regression models mapping sequences of wearable-sensor-based features to ankle torque values. Section II covers background information on transfer learning, particularly as applied to recurrent neural networks and human subject sensor data, and a literature review of related work. Section III describes the data collection, model development, and experiments evaluating the impact of one-to-one and three-to-one subject based transfer learning on task performance and network training resources. Sections IV and V present and discuss the results of those experiments, followed by a summary of conclusions in Section VI.

II. BACKGROUND

Transfer learning is a machine learning method that uses a model developed for one task as a starting point for a second task, often to save training time and/or to achieve better performance. These methods are frequently employed in applications where data is limited for the second task (the "target domain") but not for the first (the "source domain"). More generally, transfer learning has often been shown to reduce the initial model error, train faster, and have a smaller final error than training machine learning models *de novo* with randomized weights.

Transfer learning has been a great enabler in computer vision, which is typically focused on leveraging pre-trained convolutional neural net (CNN) layers for new tasks [6]. The advantages of performing transfer learning in this domain include the existence of large, open, and standard datasets [7], [8], and the standardization of CNN layers as the initial components of most image-based neural networks. For sequence-based networks that are often used for time series data, such as recurrent neural networks (RNN) and long short-term memory networks (LSTM), the body of work is more limited. Much of the work here is with natural language models [9], [10], again, driven by standardized, large datasets [11]. RNN and LSTM networks are particularly applicable to the biomechanics domain, since biomechanics is largely concerned with time series information.

A substantial body of work exists in applying transfer learning techniques, including cross-subject transfer, to human activity recognition from wearable sensor data [12]. Specific to kinematics regression tasks, prior work includes the application of LSTMs to the estimation of skeletal muscle forces from kinematics data during walking gait cycles, using weight transfer across multi-subject simulation to improve model accuracy [13]. Despite some similar work, to the best of our knowledge, inter-subject transfer for joint dynamics regression has not been previously investigated, particularly with the goal of understanding inter-subject differences. Our work focuses on parameter transfer between networks via weight initialization. Other approaches exist, but were not explored in this study; we refer the reader to Pan et al. 2009 for a broader survey of these methods [14].

When applying transfer learning to machine learning models for a wearable-sensor data driven exoskeleton, we view the source domain as data from a single subject or population — for which data can be collected and processed, and models trained without significant time or computational constraint — and the target domain as a single human operator. In operational deployment of these models, individualization may be compute-constrained if the system needs to adapt to a new user, or if an existing user's physiological signals change during use (e.g. due to fatigue, sweat, etc). In both contexts, the system might begin with some sensor-to-movement model and need to adapt using onboard computational resources to improve (or prevent degradation of) performance. The adapt-during-use case has a stricter requirement for retraining to be done onboard, since the changes in user signals are occurring during use. Alternatively, in cases where the systems are not compute-constrained, but obtaining ground truth may be expensive (e.g. utilizing motion capture), it may be desirable to limit the amount of subject-specific information that is needed to allow a "calibration" to a new user.

The latter scenario is particularly relevant to situations such as commercialization of a product, or mass deployment of a device to personnel, such as warehouse workers or soldiers. Here, it could be advantageous in terms of time and cost to leverage transfer learning to specialize a model to an individual, while starting from a generic model known to work well for another person, or group of people.

Although models for motion estimation from accelerometer inputs are fairly generalizable between people due to the dynamical constraints of human motion [15], the same has historically not been true for electromyography sensors [16]. Surface electromyography (sEMG) senses electrical activity of human muscles, and the specific characteristics of the signals can vary significantly depending on the physiology of the subject (e.g. skin, fat, sweat) as well as on the sensor placement [16]. sEMG is, however, potentially more useful than kinematics for some applications where motion *prediction* is desirable, such as wearable robotics (since its signals precede motion), or when a force or moment is the measure of interest, since sEMG is related to these even without motion. Apart from motion prediction, the inclusion of sEMG alongside accelerometry has been shown to improve the accuracy of human activity recognition [17]. Machine learning on sEMG signals has been demonstrated for force/torque estimation to some degree in a subject-independent setting [18] and between experiment sessions [19], though most work in this domain has focused on single-session, subject-dependent models.

In an effort to better understand and address the use of transfer learning for model personalization, this paper investigates single- and multi-subject based parameter transfer on LSTM models trained for subjects' sensor-to-joint torque prediction tasks, with regards to task performance, and network training requirements in terms of computational resources and data required.

III. METHODS

In this study, we consider the problem of training neural network models that map wearable sensor inputs to human ankle joint moments during various types of locomotion. Specifically, we probe whether models that are pre-trained on one or more individuals perform well when 1) faced with data from a new subject without additional subject-specific training (zero-shot learning), 2) trained to convergence on data from a new subject (transfer learning), and 3) trained only on a small amount of data from the new subject (data-limited transfer learning).

Previous work showed that for this task, LSTMs are able to outperform a variety of other networks used in the literature for similar tasks [20]. We start with that network model as our basic training setup and test the three previously-described conditions. This section covers the methods used to collect the human locomotion data, the training of individual base models, and the transfer learning experiments that build on top of the base models.

A. Locomotion Data Collection

Five healthy human subjects completed the data collection protocol, but only four subjects' data were usable due to sensor malfunction on one subject. The remaining four subjects were comprised of three females and one male, were 25.0 ± 4.2 years old, with body mass of 72.7 ± 21.6

Fig. 1. One of the experiment subjects with sEMG/acceleromter sensors and motion capture markers.

kg (mean and standard deviation). All participants provided written, informed consent, and the protocol was approved by the MIT Committee On the Use of Humans as Experimental Subjects (protocol #1703875483).

The data collection protocol involved self-paced locomotion on an instrumented treadmill, during which subjects were given commands to stand, walk, run, and sprint, via on-screen text, which changed every ten seconds. Each trial lasted 150 seconds, and subjects completed five trials with varying orders for the commanded speeds, with an opportunity to rest between trials.

All subjects wore eight Delsys Trigno Avanti sensors (combined surface electromyography and acceleromters, Delsys Inc, Natick, MA, USA), four on each leg (vastus medialis, hamstring, tibialis anterior, gastrocnemius medialis), along with motion capture markers forming a modified Plug-in Gait model (Vicon Industries Inc., Oxford, UK) focused on the lower body (Fig. 1). Half-second windows with 0.49 s overlap were used for feature extraction from the surface electromyography (sEMG) sensors and accelerometers. For sEMG signals, the max value and area of the rectified signal were calculated within each window, and for accelerometer signals, the median vector magnitude and the median angle in the X-Y, Y-Z, and Z-X planes were calculated for each window. These features were selected due to their low computational cost, which makes them suitable for real-time extraction with a low-power wearable computer.

The prediction labels were ankle torques over time, as estimated by the Plug-in Gait model using motion capture and force plate information, which is one of the most common methods for human gait data reduction [21]. Further experiment details are described in our previous work [20].

B. Baseline Model Training

All machine learning training and testing was done using the PyTorch framework. The model architecture used in all experiments in this study was a single-layer LSTM with 64 hidden units, followed by two feedforward layers with 16 hidden units per layer [20]. In that study, this architecture was found to perform similarly to a model more than twice the size (also from literature, [2]) for our ankle torque prediction task. For baseline training, all networks were trained for ankle torque regression, using an input sequence of 50 time steps of features, and an output sequence of 50 time steps of pairs of ankle torques. Any output sequences that contained values beyond 5σ of the mean were rejected as outliers and not used. Additionally, since the subjects spent slightly different amounts of time in each commanded activity, segments belonging to shorter activities were oversampled in the training to balance out activity representation. To facilitate both model training and post-hoc accuracy comparisons between subjects, the output torques were normalized from units of $N \cdot m/kg$ to dimensionless values between 0 and 1 using min-max normalization.

Mean squared error (MSE) loss was used to measure model performance, and models were trained until the difference in training loss between subsequent 3-epoch windows was below $0.1\ \times 10^{-3}$ (with the exception of 3-subject baseline models, which were stopped at $0.01\ \times 10^{-3}$). This will be referred to as an early stopping or convergence criteria.

As 6 trials of data were collected for each subject, two variants of a within-subject six-fold cross validation schemes were employed to evaluate the generalizability of results. In the "standard" cross-validation scheme, 5 trials were used for training, and one for testing, while in the "data-limited" scheme, 1 trial was used for training and 5 for testing. The latter was tested to probe the utility of transfer learning techniques for compute- or data-constrained applications. Model performance on the test dataset was recorded before training began ("epoch 0") and after every training epoch.

C. Transfer Learning Experiments

Three weight initialization schemes were considered: single- (1-to-1) subject transfer, multi- (3-to-1) subject transfer, and a random weight initialization baseline (Fig. 2).

A set of baseline models were trained, using a random weight initialization scheme whereby network weights are initialized by drawing from a Gaussian distribution. 24 single-subject models (four subjects, each with 6 cross-validation folds) and 24 multi-subject models (four 3-pooled-subject combinations, each with 6 cross-validation folds) were produced. Baseline models were always trained with the "standard" cross validation scheme. From this set of baseline models, one model (the third cross-validation fold) per subject or multi-subject pool was chosen to be the representative "source domain" model for all following transfer learning experiments.

For each subject (whose data can be considered the "target domain" for transfer learning), 6 multi-subject transfer learning models were trained (one 3-pooled-subject weight initialization scheme with 6 cross-validation folds) and 18 single-subject transfer learning models were trained (three single-subject weight initialization schemes - one per "source domain" subject - with 6 cross-validation folds). This process

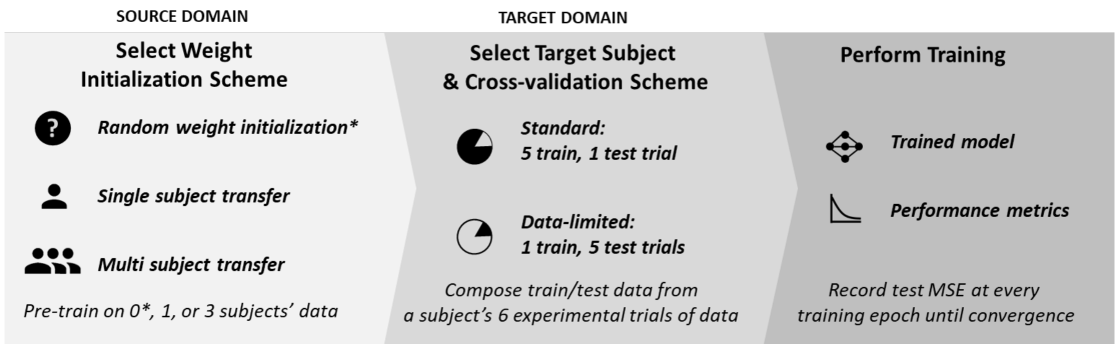

Fig. 2. Summary of the model training process, for baseline (initialized with random weights) and transfer-learned models (initialized with pre-trained model weights), in both standard and data-limited training. Six-fold cross validation is utilized, producing 6 models per source and target domain selection.

was repeated using each of the two cross-fold validation variants.

The model training process is summarized in Figure 2. This set of models was evaluated to understand the degree and utility of knowledge transfer between subjects or groups of subjects, by drawing comparisons in model performance across baseline, single-, and multi-subject transfer schemes, as well as across standard and data-limited training schemes.

To evaluate the impact of knowledge-transfer on network training and model performance, we consider three characteristics: the initial accuracy of a model, both before any training has taken place (*zero-shot learning*, "epoch 0") and after one pass ("epoch 1") through the training data; the number of passes through the training data (epochs) it takes to complete model training (reaching a predefined training error "convergence criteria"); and the final model accuracy on test data. If a particular weight initialization scheme imparts useful information for a target domain, we expect to see lower initial model error as well as potentially a lower number of epochs to converge, or lower final model error, relative to a random weight initialization scheme (illustrated in Figure 3). At epoch 0, a model without transfer-learned parameter initialization is merely a random model, thus this model performance is not considered beyond noting maximal error bounds.

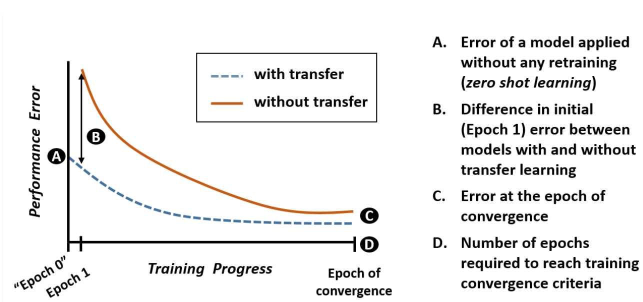

Fig. 3. Characteristics of model training and network performance, expected in the case of parameter transfer learning with utility between source and target domains.

As previously stated, the accuracy values are based on the per-subject normalized torques. This normalization allows comparisons between subjects, since the focus of this work is on relative accuracy changes due to transfer learning, rather than specific accuracy values. For context, the same model architecture and data used here with Subjects 1-3 resulted in MSE values of 0.0053 ± 0.0037 $(N \cdot m/kg)^2$ (0.073 ± 0.061 $N \cdot m/kg$ RMSE) [20].

IV. RESULTS

Boxplots for each of the aforementioned metrics for single- and multi- subject transfer weight initialization schemes, as well as a random weight initialization baseline are shown in Figures 4 to 6. Each boxplot comprises six data points: one per cross validation fold.

For both the standard and data-limited contexts, MSE at epoch 0 varies for each model and is substantially higher than MSE at epoch of convergence, indicating that target-subject-specific retraining provides a substantial improvement in model performance over zero-shot transfer.

For both single- and multi-subject transfer models in the standard cross-validation context (Fig. 4), MSE at epoch 1 is substantially lower than that of the baseline random weight initialized model. The differences in MSE between transfer learning and random weight initialized models after one training epoch was 0.022 on average for both single- and multi-subject transfer (same to this number of significant figures). This trend is not consistent in the data-limited context (Fig. 5). No clear trend in MSE at epoch 1 emerges across all single- or multi-subject initialization schemes.

Model performance at the epoch of convergence is similar across all target domain subjects and weight initialization schemes for the standard cross-validation results (difference of 0.0016 normalized units between single-subject transfer models and baseline; 0.0023 normalized units for multi-subject transfer models). This similarity indicates that that parameter transfer followed by target-subject-specific retraining generally neither harms nor helps final model accuracy. However, for models trained in the data-limited context, random or multi-subject weight initialization have lower errors at the epoch of convergence relative their single-subject weight initialized counterparts, across all target domain subjects.

The number of epochs required to reach convergence was lower with transfer learning in most, but not all, cases of both the 5-trial and 1-trial training (Fig. 6). Notably, when Subject 4 was the target, random initialization did not take longer than pre-trained parameter transfer to reach convergence in many cases for both 5-trial and 1-trial training. Subject 1 also did not show substantial differences in convergence time between random and pre-trained initialization. Single-subject-based parameter transfer reduced training time by five epochs on average, while multi-subject transfer took an average of six fewer epochs to converge.

Tables 1 and 2 further summarize the mean and standard deviation of MSE (Table 1) and epochs to convergence (Table 2) over cross validation folds for each combination of initialization scheme and source/target domain.

TABLE I

MSE FOR 5-TRIAL TRAINING (AT CONVERGENCE)

Source Domain	*Target Domain*			
Initialization	Subj 1	Subj 2	Subj 3	Subj 4
Subj 1		3.3 ± 0.8	5.0 ± 0.7	3.6 ± 1.4
Subj 2	7.2 ± 8.1		5.4 ± 0.5	3.8 ± 0.5
Subj 3	14.1 ± 24.1	3.0 ± 0.2		3.8 ± 0.7
Subj 4	11.5 ± 18.3	3.8 ± 0.6	6.0 ± 1.2	
3-to-1	9.2 ± 11.9	3.6 ± 0.9	4.6 ± 0.8	3.1 ± 0.7
Random	16.0 ± 20.3	3.7 ± 0.2	5.9 ± 1.5	4.4 ± 1.4

Mean & standard deviation over cross-validation folds. All values are MSE $\times 10^{-3}$.

TABLE II

EPOCHS TO CONVERGENCE FOR 5-TRIAL TRAINING

Source Domain	*Target Domain*			
Initialization	Subj 1	Subj 2	Subj 3	Subj 4
Subj 1		17.0 ± 5.8	12.0 ± 9.7	13.2 ± 5.3
Subj 2	11.7 ± 3.3		14.2 ± 5.0	13.7 ± 4.5
Subj 3	12.7 ± 4.6	18.3 ± 5.3		17.0 ± 6.2
Subj 4	14.7 ± 3.9	14.7 ± 4.8	16.7 ± 4.8	
3-to-1	16.7 ± 6.9	12.0 ± 2.8	12.0 ± 3.9	13.7 ± 5.0
Random	17 ± 7.5	20.7 ± 4.3	25.0 ± 9.7	16.0 ± 4.2

Mean & standard deviation over cross validation folds.

V. DISCUSSION

A. Transfer-Learned Model Accuracy and Training

The differences in MSE between transfer learning and baseline after one training epoch indicates that there is useful knowledge transfer from models developed for different subjects: prior information learned from mapping wearable sensor features to joint torque values from one subject has utility for that same task but a different subject.

The accuracy of single-subject models applied to target data without retraining (the zero-shot learning application) shows variability based on the source subject. For example, a model trained on subject 4 (source) has a substantially lower test error than either subject 1 or 3 (source) when applied to data from subject 2 (target). The model built for subject 1 has a significantly lower error than that of the other subjects when applied to subject 3. There may be a variety of contributing factors to inter-subject differences, ranging from physiological differences (particularly for the sEMG sensors, as previously described), to differing gait strategies [22] and muscle synergies for movement [23]. Understanding potential causes would require additional research and a larger data set, which may allow such differences to be seen. Such a capability could be useful in a clinical setting where the best model for a patient needs to be chosen from a set of pre-existing models with known or defined predictors to aid with model selection. A larger dataset may also reveal phenotype groups and allow for a library of multiple "generic" models from which to choose as a source domain when applied to a new individual.

Table II and Figure 6 indicate that parameter transfer learning reduces training time in most cases, without sacrificing the accuracy of the final trained models. The time and computational resource savings here would be most impactful in low-compute environments (e.g. retraining onboard a wearable computer) or mass deployment applications.

B. Implications for Personalized Physiological Sensing

The ability to perform accurate physiological state sensing using wearable sensors is a key capability for use cases including fitness trackers for the general population, monitoring patient progress in a healthcare setting, and controls for wearable robotics. The level of sensing and state estimation we present here would typically require a significant amount of time to both collect subject-specific data, and develop accurate models. Additionally, EMG-based state estimation has historically been difficult to translate between sensor placements, much less between people, without the use of individualized machine learning models [24], [4]. Although previous work has evaluated the zero-shot learning case for wearable sensors both across participants and between sessions, transfer learning remains underexplored in this domain [18], [19]. Thus, our results show the potential for using transfer learning as a technique to reduce both the computational and data-collection costs of personalizing sensing models, and hint at the possibility of more generalized models that work well with new users.

C. Limitations

One of the limitations of this experiment is the small size of the subject pool. Further experiments to show model transferability could be done with a larger dataset, particularly in the context of developing "general" models based on a variety of individuals, in a similar vein as our 3-to-1 transfer learning experiments.

Secondly, our experiments use a single neural network architecture, with a pre-chosen set of input features [20]. The kind of weight initialization used here is thus different from some other commonly-used techniques in the field, where pre-training is focused on feature extraction layers, as is common in transfer learning for computer vision [6].

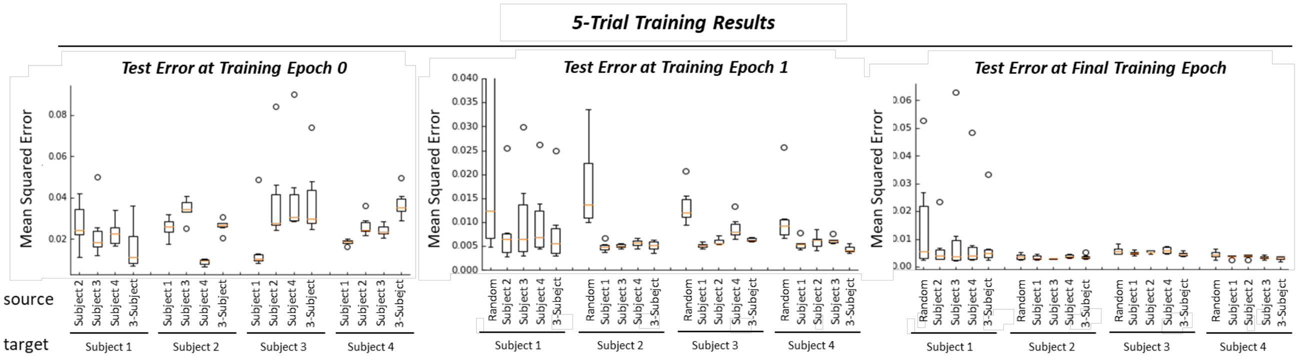

Fig. 4. One-to-one and three-to-one subject parameter transfer, standard cross-validation scheme: Test data MSE, evaluated at training epochs 0 (zero-shot learning model accuracy), 1 (initial model accuracy), and convergence (final model accuracy). Test errors for epoch 0 and random weight initialization are not shown; mean errors are ≈ 2.0 for those cases. Random source refers to random weight initialization.

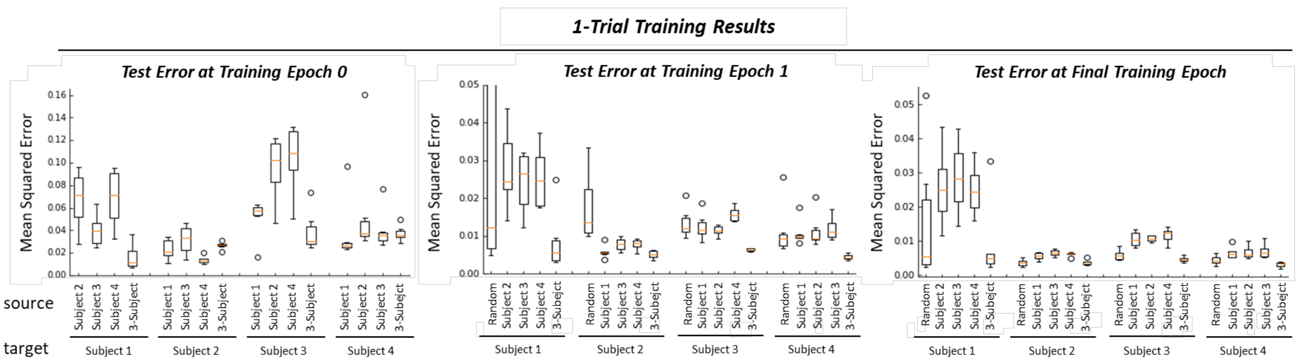

Fig. 5. One-to-one and three-to-one subject parameter transfer, data-limited cross-validation scheme: Test data MSE, evaluated at training epochs 0 (zero-shot learning model accuracy), 1 (initial model accuracy), and convergence (final model accuracy). Test errors for epoch 0 and random weight initialization are not show; mean errors are ≈ 2.0 for those cases. Random source refers to random weight initialization.

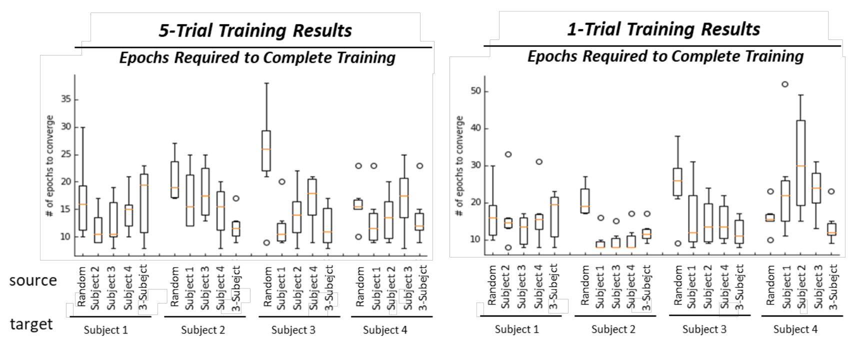

Fig. 6. Convergence time (epochs) for one-to-one and three-to-one subject parameter transfer learning.

VI. Future Work

Broad categories for potential future work include exploring model generalizability, using alternative neural architectures, and transfer learning techniques focusing on feature extraction.

For greater model generalizability, a larger subject pool would be desirable. Given the general trends of data usage in deep neural networks, it is possible that a model trained on a sufficiently large dataset may have superior zero-shot performance than our 3-subject pooled models. On the other hand, specific phenotypes of groups of individuals with similar physiological signals may also emerge with a sufficiently large subject pool. Future work may explore whether learned models are able to disambiguate between such phenotypes when faced with a new subject simply through the normal training process, or if such information needs to be explicitly encoded through feature engineering or clustering techniques.

For alternative neural architectures, the use of models that attempt to learn the forms of mechanistic models for physical systems ("scientific machine learning" [25]), are an attractive avenue of exploration. These models allow for a blend of the universal approximation properties of neural networks and physical system constraints. For human motion, constraints exist in dynamical models of limbs, and the limits of human motion. Given the dynamical nature of wearable sensor inputs and human state outputs, these approaches may present a way to decrease training time, improve model interpretability, and increase robustness to variations between individuals. An exploration of inter-subject variability and transferablility here may be a useful avenue for future research.

Finally, we picked handcrafted input features in this work, but a type of transfer learning not performed here focuses on feature extraction, and is a common approach in general machine learning. Under this framework, instead of allowing an entire pre-trained network to be retrained (as we do here), early network layers, which are primarily responsible for feature extraction, are frozen, while later layers are further tuned [6]. This technique might remove the need to choose handcrafted features. Although common networks for general time-series feature extraction do not really exist as they do for image recognition, there does exist some work that uses CNNs to do feature extraction for physiological input data similar to ours, so assessing the transferability of such models is a reasonable next step [18], [19].

VII. CONCLUSIONS

This paper demonstrates the utility of single- and multi-subject-based parameter transfer on LSTM models trained for "sensor-to-joint torque" prediction tasks during locomotion, with regards to task performance and computational resources required for network training. The results illustrate useful knowledge transfer from both paradigms, with varying results across specific "source" and "target" subject pairings. Comparable accuracy of full trained models, relative to baseline models without transfer learning, achieved in fewer training epochs show that such techniques can be leveraged to lower model training time or computational cost in compute-constrained environments. Target-subject-specific training beyond the zero-shot application substantially reduced model error, which suggests transfer learning techniques could be used to minimize data collection and model retraining requirements.

ACKNOWLEDGMENTS

The authors would like to acknowledge Emilia Grzesiak for her assistance with the machine learning code, Harvey M. Edwards and Aaron Rodriguez for assistance with collecting and processing the data, and the MIT SuperCloud for providing computational resources.

REFERENCES

[1] S. R. Simon, "Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems," *Journal of biomechanics*, vol. 37, no. 12, pp. 1869–1880, 2004.

[2] P. Slade, R. Troutman, M. J. Kochenderfer, S. H. Collins, and S. L. Delp, "Rapid energy expenditure estimation for ankle assisted and inclined loaded walking," *Journal of neuroengineering and rehabilitation*, vol. 16, no. 1, pp. 1–10, 2019.

[3] M. M. Rodgers, V. M. Pai, and R. S. Conroy, "Recent advances in wearable sensors for health monitoring," *IEEE Sensors Journal*, vol. 15, no. 6, pp. 3119–3126, 2014.

[4] H. C. Siu, A. M. Arenas, T. Sun, and L. A. Stirling, "Implementation of a surface electromyography-based upper extremity exoskeleton controller using learning from demonstration," *Sensors*, vol. 18, no. 2, p. 467, 2018.

[5] K. Anam and A. A. Al-Jumaily, "Active exoskeleton control systems: State of the art," *Procedia Engineering*, vol. 41, pp. 988–994, 2012.

[6] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[8] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.

[9] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," *arXiv preprint arXiv:1801.06146*, 2018.

[10] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018.

[11] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "Squad: 100,000+ questions for machine comprehension of text," *arXiv preprint arXiv:1606.05250*, 2016.

[12] N. Hernandez, J. Lundström, J. Favela, I. McChesney, and B. Arnrich, "Literature review on transfer learning for human activity recognition using mobile and wearable devices with environmental technology," *SN Computer Science*, vol. 1, no. 2, pp. 1–16, 2020.

[13] T. Dao, "From deep learning to transfer learning for the prediction of skeletal muscle forces," *Medical and Biological Engineering and Computing*, vol. 57, pp. 1049–1058, 2019.

[14] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[15] P. H. Veltink, H. J. Bussmann, W. De Vries, W. J. Martens, and R. C. Van Lummel, "Detection of static and dynamic activities using uniaxial accelerometers," *IEEE Transactions on Rehabilitation Engineering*, vol. 4, no. 4, pp. 375–385, 1996.

[16] C. J. De Luca, "The use of surface electromyography in biomechanics," *Journal of applied biomechanics*, vol. 13, no. 2, pp. 135–163, 1997.

[17] S. Gonzalez, P. Stegall, H. Edwards, L. Stirling, and H. C. Siu, "Ablation analysis to select wearable sensors for classifying standing, walking, and running," *Sensors*, vol. 21, no. 1, p. 194, 2021.

[18] 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, 2018.

[19] T. Bao, S. A. R. Zaidi, S. Xie, P. Yang, and Z.-Q. Zhang, "A cnn-lstm hybrid model for wrist kinematics estimation using surface electromyography," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2020.

[20] H. C. Siu, J. Sloboda, R. J. McKindles, and L. A. Stirling, "Ankle torque estimation during locomotion from surface electromyography and accelerometry," in *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, pp. 80–87.

[21] A. Ferrari, M. G. Benedetti, E. Pavan, C. Frigo, D. Bettinelli, M. Rabuffetti, P. Crenna, and A. Leardini, "Quantitative comparison of five current protocols in gait analysis," *Gait & posture*, vol. 28, no. 2, pp. 207–216, 2008.

[22] D. A. Winter, "Kinematic and kinetic patterns in human gait: variability and compensating effects," *Human movement science*, vol. 3, no. 1-2, pp. 51–76, 1984.

[23] A. B. Ajiboye and R. Weir, "Muscle synergies as a predictive framework for the emg patterns of new hand postures," *Journal of neural engineering*, vol. 6, no. 3, p. 036004, 2009.

[24] N. Beckers, R. Fineman, and L. Stirling, "Anticipatory signals in kinematics and muscle activity during functional grasp and release," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2015, pp. 1–6.

[25] C. Rackauckas and Q. Nie, "Differentialequations.jl–a performant and feature-rich ecosystem for solving differential equations in julia," *Journal of Open Research Software*, vol. 5, no. 1, 2017.