Transferring Gait Predictors Across EMG Acquisition Systems with Domain Adaptation

Annika Guez[1], Balint Hodossy[2], Dario Farina[2] *Fellow, IEEE* and Ravi Vaidyanathan[1] *Member, IEEE*

***Abstract*— Lower limb assistive technology (e.g. exoskeletons) can benefit significantly from higher resolution information related to physiological state. High-density electromyography (HD-EMG) grids offer valuable spatial information on muscle activity; however their hardware is impractical, and bipolar electrodes remain the standard in practice. Exploiting information rich HD-EMG datasets to train machine learning models could help overcome the spatial limitations of bipolar electrodes. Unfortunately, differences in signal characteristics across acquisition systems prevent the direct transfer of models without a drop in performance.**

This study investigated Domain Adaptation (DA) to render EMG-based models invariant to different acquisition systems. This approach was evaluated using a Temporal Convolutional Network (TCN) that mapped EMG signals to the subject's knee angle, using HD-EMG as source data and Delsys bipolar EMG as target data. Furthermore, the feature extraction learnt by the TCN was also applied across muscle groups, evaluating the transferability of the sensor agnostic features.

The DA implementation shows promise in both scenarios, with an average increase in accuracy (angular error normalised by the range of motion) of 7.36% for the Rectus Femoris, Biceps Femoris and Tibialis Anterior, as well as a cross-muscle performance increase of up to 10.80%. However, when the domain discrepancy is severe, the model is currently unable to generate a reliable walking trajectory due to inherent limitations related to the applied regression scheme and the chosen Mean Squared Error loss function. Therefore, future research should focus on exploring advanced loss functions and classification-based DA models that prioritise restoring key features of the gait.

I. Introduction

Electromyography (EMG) signals are often used as an indication of the user's motion intent for gait prediction models [1], and bipolar EMG sensors are the most commonly used modality in practice. However, in the context of assistive robotics, high-density electromyography (HD-EMG) can provide valuable spatial information. This can improve the differentiation of intended tasks and effort levels [2].

Bipolar sensors are highly sensitive to electrode placement, and may disregard important aspects of the muscle activation due to their low spatial resolution [3]. While the conventional EMG systems can describe the general activation profile of a given muscle group, HD-EMG has the ability to quantify spatial distributions of activity within the muscle itself [4]. This information is crucial when investigating muscles or regions with inhomogeneous activation patterns, such as groups of numerous small muscles that are closely clustered together, or muscle contraction behaviours for certain pathologies [5].

Despite the benefits of using HD-EMG data, no commercial myoelectric robotic system uses HD-EMG as its sensing modality; the setup required for its acquisition is complex, comprising of large and heavy devices with many connecting cables, which renders this system impractical for use in dynamic settings [6]. Therefore, investigating methods that would enable the use of HD-EMG information in practice could lead to improvements in bipolar applications.

Bipolar EMG signals can be extracted from HD-EMG grids by taking the differential of 2 monopolar signals [7]. However, due to hardware aspects of the acquisition system (built-in filters, electrode size and type, amplifier gain, input range, etc.), this simulated bipolar signal does not translate perfectly to signals obtained from commercially available bipolar devices [8], such as the wireless Delsys Trigno Avanti sensors [9]. Therefore, a model developed on the simulated bipolar signals from HD-EMG grids will suffer severe performance degradation when applied directly to real bipolar signals due to distribution shifts between datasets.

Transfer learning is a machine learning method that consists of transferring knowledge obtained from one task to another [10]. Originally developed for computer vision, Domain Adaptation (DA) is a form of transfer learning that alters a model trained on one dataset to fit another by minimising the differences in distribution between the source domain (training data) and the target domain (testing data) [11]–[13]. DA methods aim to develop domain-invariant models by extracting similar features from either domain [14]. This can be achieved through the use of Convolutional Neural Networks (CNNs), which have often been used to extract high-level feature representations [15]. Domain-specific information is separated from the domain-invariant features, and the model can therefore focus on making predictions based only on the latter.

Previous studies have attempted to apply DA methods to bio-signal inputs such as electroencephalography (EEG) and EMG, however, they mainly focused on transferring knowledge across different sensing modalities [10,16] or subjects [17]. The work presented in this paper is concerned with the transfer of knowledge across similar EMG signals acquired from the same subject but different systems,

This work was supported by the UKRI CDT in AI for Healthcare (Grant No. EP/S023283/1), CDT in Prosthetics and Orthotics (Grant No. EP/S02249X/1), the UK Dementia Research Institute Care Research & Technology Centre (DRI-CRT), the Natural BionicS initative (Grant agreement ID: 810346, and the Hamlyn Centre for Robotic Surgery of Imperial College London).

[1]A. Guez and R. Vaidyanathan are with Department of Mechanical Engineering, Imperial College London, London, United Kingdom

[2]B. Hodossy and D. Farina are with Department of Biomedical Engineering, Imperial College London, London, United Kingdom

therefore compensating for the inherent signal discrepancies caused by differences in hardware.

Since both acquisition systems investigated in this study were used to map EMG data to knee angle, the approach can be categorised as transductive learning, more specifically, as a supervised homogeneous DA architecture, where the labels for both the source and target domains are similar [18].

The main challenge when developing a domain-agnostic model is to simultaneously optimise for the task whilst minimising the impact of the domain shift on the model [19]. This study therefore investigated the use of DA to train a domain-invariant predictive model on HD-EMG data with the aim to improve performance on bipolar data. A Temporal Convolutional Network (TCN) was trained for regressive knee angle prediction during level-ground treadmill walking, with and without the incorporation of the DA enhancement to the network architecture, to evaluate the viability of this approach.

II. Methods

A. Data acquisition

EMG data was acquired from the rectus femoris (RF), biceps femoris (BF), tibialis anterior (TA) and soleus (SO) of an able-bodied subject walking on a treadmill at 0.8m/s for 3 minutes in the steady-state, corresponding to approximately 380 strides.

HD-EMG data were acquired from these muscles using a 13×5 monopolar electrode grid, with an inter-electrode distance (IED) of 8mm both horizontally and vertically (model number: GR08MM1305), connected to a wireless Sessantaquattro acquisition system [20]. The area where the sensor was located was shaved, gently abraded, cleaned with an alcoholic solution and dried prior to attaching the grid to ensure satisfactory contact between the electrodes and the skin. The grids were secured onto the subject using the adhesive surface of the electrode foam, micro-pore surgical tape on the edges, and self-adhesive bandages that were wrapped around the patient's leg until the electrode was fully covered. A key difference between the two EMG systems is that the grid uses a gelled contact, while the bipolar electrodes are dry contact, which is generally associated with a higher skin impedance [21].

This grid was positioned following SENIAM placement standards [22] such that the middle electrodes positioned vertically at the centre of the grid have the same placement as would bipolar electrodes. The simulated bipolar signals were extracted by taking the voltage differential between 2 monopolar electrodes at the centre of the grid. Since the IED of the HD-EMG grid does not exactly match that of the Delsys, bipolar signals were extracted from the central monopolar electrodes (numbers 31, 32, and 33) as the 3 unique combination pairs ([31, 32], [32, 33], [31, 33]), and merged to construct the source dataset. Figure 1 shows the coloured monopolar electrodes corresponding to the chosen central electrodes from the HD-EMG grid, juxtaposed with the positioning of the Delsys Trigno Avanti electrode (grey rectangle).

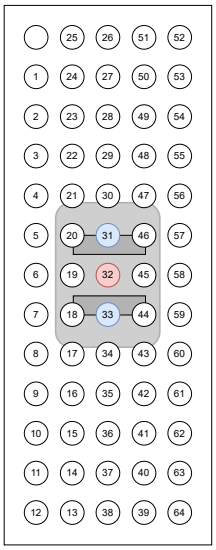


Fig. 1: Grid placement relative to the Delsys sensor.

The bipolar sensors were fixed using single-use Delsys adhesive interfaces, and secured with microporous adhesive tape and medical fabric strapping tape. The skin surface that had been covered by the grid was cleaned with the alcoholic solution and dried once more before placing the Delsys sensors, to remove any remaining conductive cream. A Vicon Motion Capture system, equipped with 10 Vicon Vero v2.2 cameras placed around the room pointing at the treadmill, was used to capture the ground truth for the knee angle kinematics. This trajectory data was labelled using the Vicon-Nexus software, with the knee angle flexion extracted using the joint coordinate system defined in [23], and was then up-sampled to match the recorded surface electromyography (sEMG) data. The HD-EMG and Vicon data were synchronised using shared reference pulses sent from a micro-controller at the start and end of the trial to align the signals, whilst the Delsys system was embedded within the Vicon system so the recordings were already synchronised. To facilitate comparison, the Delsys data was re-sampled to match the HD-EMG sampling rate of 2048Hz.

The experiments for data acquisition were performed in compliance with ethical procedures approved by the Imperial College Research Ethics Committee (ICREC reference: 21IC7204).

B. Label alignment

To enable transfer learning between datasets, an initial label alignment step was required [17]; Each gait cycle, extracted from the knee angle trajectory using the knee flexion peaks, was temporally normalised so that a given window of EMG, from either acquisition system, would map to the same gait cycle segment. This temporal normalisation was performed by resampling the knee angle and EMG of each gait cycle to a fixed number of samples larger than either cycle (4000 samples). A global shift of the EMG signals was then performed based on their envelope peaks to account for the shift in the recorded gait patterns caused by the reapplication of motion markers between experiments. Whilst this does not result in an exact mapping between signals, it enables the datasets to be interpreted comparably by the regressor, and was therefore deemed sufficient for this application.

C. Knee angle prediction using a TCN

The sEMG signals were segmented using a 250ms sliding window (corresponding to 512 samples for a 2048Hz sampling frequency) with a 40 sample step between windows.

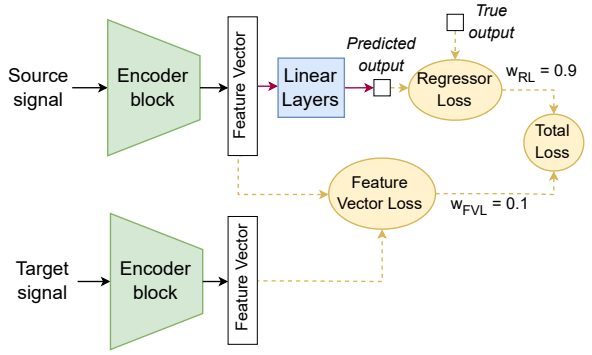


Fig. 2: Extended DA architecture with the combined FVL and RL. The encoder blocks are trained on the source and target data simultaneously to extract domain invariant features.

No preprocessing or filtering was applied other than the normalization of the signals based on the mean and standard deviation (σ) of the training dataset. A TCN was constructed to map each muscle activation window to the knee angle of the following time step, compressing the 1×512 EMG signal input to a 1×1 joint angle output. This network was composed of 5 convolutional layers with kernel and dilation sizes of 5, each with max pooling and leaky ReLU activation, followed by 5 fully connected layers with ReLU activation and dropout in between. The network weights were initialised using a Kaiming uniform distribution.

The performance drop due to a change in acquisition system was quantified by training the TCN network on the HD-EMG data, and then testing it on the bipolar data. This served as a control condition to assess the impact of applying DA. The model was run for 30 epochs, using an Adam optimizer with a learning rate (LR) of 0.0001.

D. Domain adaptation

Since this study proposed to apply models trained with HD-EMG to bipolar signals, the source signal corresponds to the simulated bipolar signals extracted from the recorded HD-EMG grids, whilst the target signal was represented by the Delsys data. The DA-augmented learning process, shown in Figure 2, was implemented on the pre-trained HD-EMG-based TCN using a LR of 10^{-6}. To ensure that the features extracted by the encoder from both the source and target datasets were similar enough to create a domain-invariant model, an additional loss, termed the feature vector loss (FVL), was calculated within the TCN and added to the regressor loss (RL). The sum of losses was then back-propagated through the network so that the difference between the encoded vectors and the knee angle prediction error are minimized simultaneously. The proportion of the FVL relative to the RL was chosen as 0.1, as it led to the best performance on the target dataset during preliminary evaluations.

For the 5 convolutional layers of the original TCN architecture, the DA experiments were repeated using between 3 to 5 encoding layers. The FVL was calculated based on the activations of either the 3rd, 4th or 5th convolutional layer, while the RL was calculated based on the output of the full network as shown in Figure 2. Since the optimal divergence-based loss function for the FVL was undetermined, the training process was re-iterated for 3 different distribution losses to compare; for this experiment, the use of an Mean Squared Error (MSE) loss for the FVL was investigated, as well as Sinkhorn [24], and Wasserstein [25] losses that have been previously used as distribution losses for Generative Adversarial Networks (GANs).

The DA-augmented TCN models were evaluated for the following scenarios:

1) **Muscle specific evaluation**: For the 4 given muscle groups, the DA was implemented as shown in Figure 3a to evaluate whether the insertion of the FVL led to better performance on target data.
2) **Cross-muscle generalisation**: For every combination of 2 muscles, the experiment depicted in Figure 3b was conducted, where the DA encoder was pre-trained on another muscle, whereas the prediction layers are re-initialised and trained on the investigated muscle; this enabled the assessment of the domain shift's consistency across systems, regardless of the muscle group.

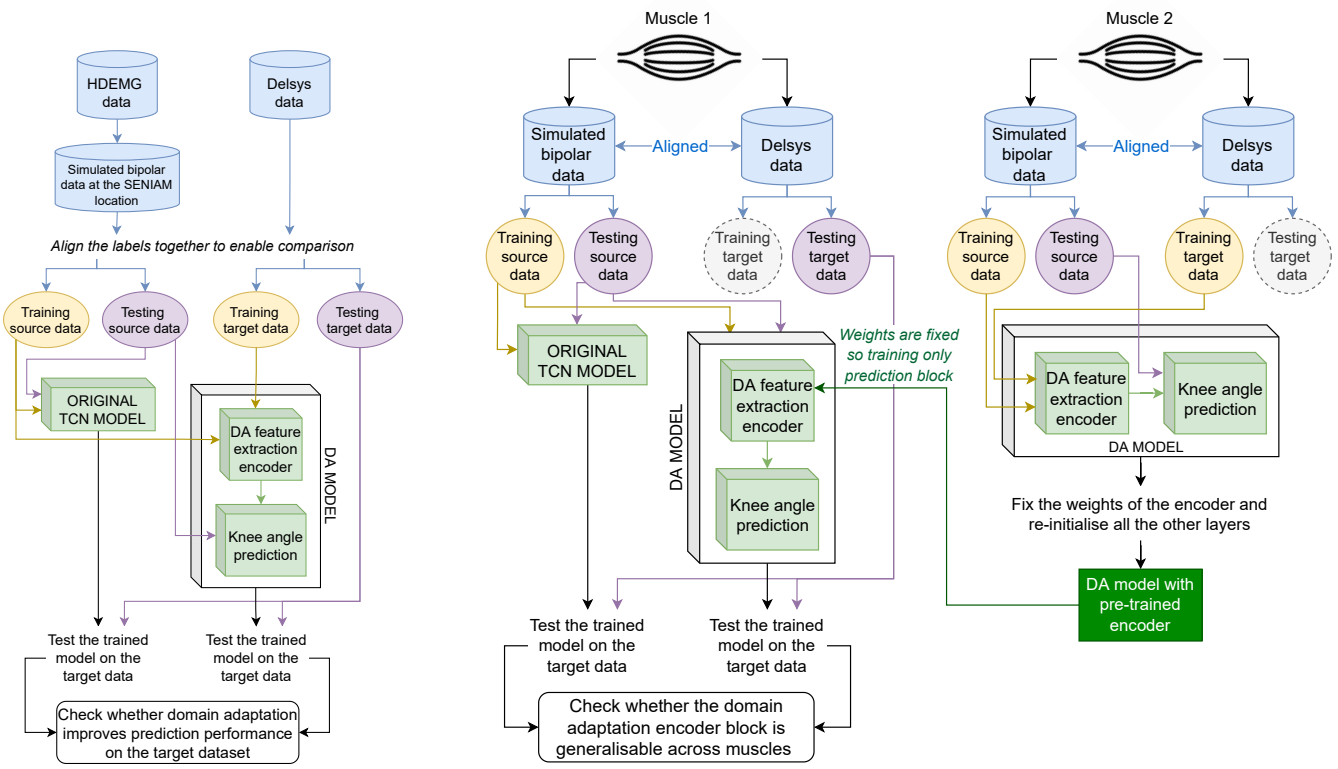
For all the experimental setups depicted in Figure 3, the train-test split used for the data was 0.7. Since the main focus of this study was the performance of the network on unseen target Delsys data, the simulated HD-EMG-based bipolar data was used as validation for the training of the networks, whereas the reserved Delsys testing data was used for model evaluation. Model accuracy was calculated as the angular error normalised by the range of motion of the knee joint, and will serve as the comparative performance metric throughout this paper.

III. Results

A. Muscle specific evaluation

As outlined in Subsection II-D, the DA portion of the TCN model was first implemented on individual muscle groups and evaluated on the corresponding target Delsys data (see sub-figure 3a). This evaluation was repeated for the different distribution losses of the FVL and for the proportion of encoding layers used in the DA ranging from 3 to 5 out of the 5 convolutional layers available in the TCN, shown in Figure 4.

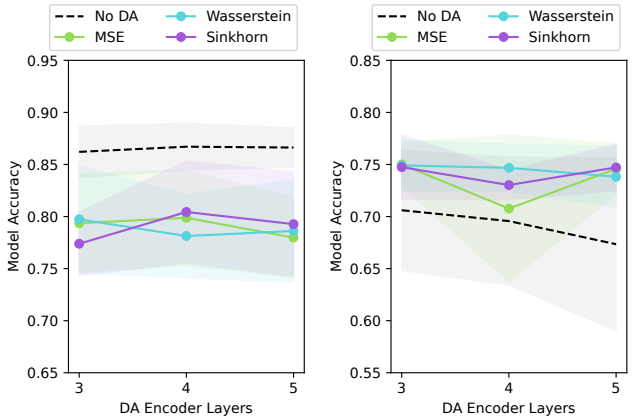
As can be seen in Figure 4, the implementation of the DA severely affected the performance of the predictive model on the HD-EMG source data, which was to be expected. The deterioration in prediction accuracy on source data is present across all investigated FVLs and proportions of convolutional layers used for feature extraction, with an average drop in accuracy of 7.442% using MSE, 7.482% using Sinkhorn, and 7.689% when using Wasserstein. This performance drop on HD-EMG data nevertheless remains tolerable, since our main focus is the impact of the model on the target Delsys



(a) Evaluating DA implementation

(b) Evaluating DA generalisation across muscles

Fig. 3: Experimental set-up to evaluate DA, for (a) the muscle specific case and the (b) muscle generalisation case.



(a) Source data

(b) Target data

Fig. 4: Performance of the original TCN model vs. the DA model for different FVL loss functions. The error bands plotted correspond to the cross-muscle variance.

data, with the HD-EMG data only used for training. To this effect, not only has the incorporation of DA in the model architecture improved the average prediction performance on target data by 4.278%, 4.990% and 5.299% for the MSE, Sinkhorn and Wasserstein losses respectively, but it has also rendered model performance a lot more consistent across muscles, with drops in σ of 2.237%, 4.494%, and 4.371% across muscle performances. Furthermore, it appears that the Sinkhorn loss performed best overall when evaluated on the target data using all 5 convolutional layers for the DA feature extraction, reducing the prediction mapping part of the network to a simple feed-forward multi-layer perceptron (MLP).

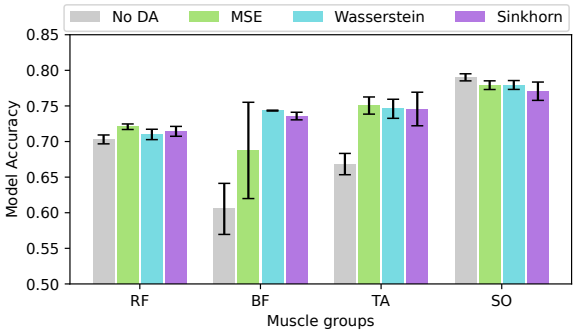


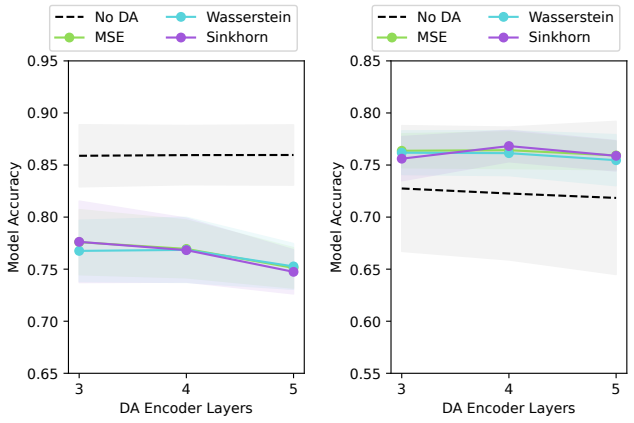
Fig. 5: Performance of the model when evaluated on Delsys target data for all muscles and FVL loss functions. The error bars correspond to the variance in performance across the number of layers in the DA encoder.

The DA consistently improves model performance on the target data for the RF, BF and TA, with an average accuracy increase of 6.072% for the MSE, 7.303% for the Sinkhorn, and 7.423% for the Wasserstein, whereas the SO suffers

instead a slight decrease in model performance. As seen Figure 5, this could be due to the SO's exceedingly high performance when trained on HD-EMG data and tested on Delsys data, hence the impact of the DA is rendered negligible.

B. Cross-Muscle generalisation

This DA implementation was repeated across muscle groups (see Figure 3b) to assess whether the translation from one acquisition system to another is constant, regardless of the muscle activation pattern examined.



(a) Source data (b) Target data

Fig. 6: Performance of the original TCN model vs. the generalised version of the DA implementation for different FVL loss functions. The error bands plotted correspond to the cross-muscle variance.

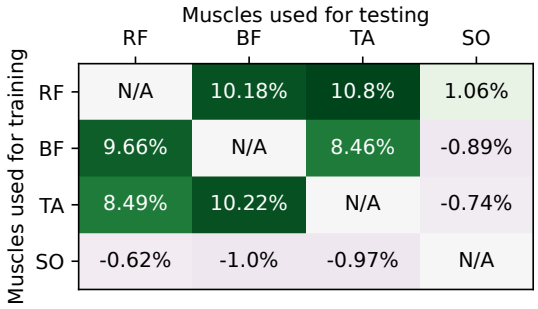


Fig. 7: Increase in prediction accuracy of the model after the generalised DA implementation. The "Muscles used for testing" correspond to the Muscle 1 from Figure 3b and the "Muscles used for training" correspond to Muscle 2. This plot is for the Sinkhorn FVL using 4 DA encoding layers.

The overall impact of the DA was very similar to that of the muscle specific experiments presented in Subsection III-A; based on Figure 6, the number of layers for each FVL that led to the highest performance were recorded, with an average accuracy increase of 2.717% when using the MSE with 3 layers, 3.417% for the Sinkhorn with 4 layers, and 2.577% for the Wasserstein with 3 layers, when the model was evaluated on the target Delsys data. The σ across all muscle combinations decreased by 4.694%, 4.872% and 4.898% for the 3 scenarios respectively.

Once more, the SO does not seem to be receptive to the added DA module (see Figure 7), and does not benefit the other muscle groups when used to train the DA feature extraction block, suggesting that its feature extraction process is non-transferable. When omitting the bias incurred by the SO outlier, the average performance increase becomes 9.463% for the MSE, 8.959% for the Sinkhorn, and 8.949% for the Wasserstein FVL when implementing cross-muscle DA.

IV. Discussion and Conclusion

Results show that the DA addition led to an increase of both performance and consistency when applied to the Delsys bipolar data of the RF, BF and TA, with an average performance increase of 7.364% paired with an average σ drop of 2.725% across the Sinkhorn and Wasserstein FVL functions. Furthermore, it led to significant improvements for the majority of cross-muscle combinations, further reinforcing the benefits of DA-augmented network training. However, this added generalisability was associated with an accuracy decrease on the source HD-EMG data, and the performance of the SO on the target data did not improve for the muscle-specific (see Figure 5) or generalisation DA (see Figure 7), which could be due to its high baseline performance on the target dataset without DA.

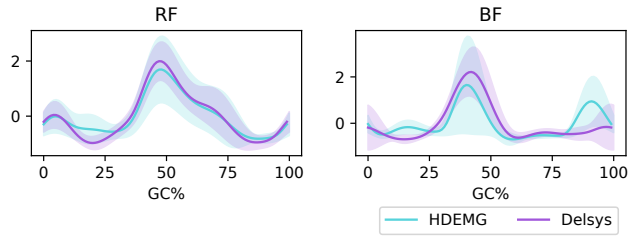


Fig. 8: Average normalised EMG signal envelopes of the RF and BF from both the HD-EMG and the Delsys systems, expressed as a function of the gait cycle (GC) percentage.

Whilst the MSE-based prediction performance improved with DA, this study revealed limitations in ensuring the practical reliability of the predicted output; certain muscle groups, such as the RF and SO, transferred relatively well to the Delsys data, maintaining shape and periodicity, however the predicted trajectory for the BF and TA was affected by the domain change (see Figure 9). As can be seen in Figure 8, there was significant cross-system discrepancy of the BF's EMG signals compared to the RF; a distribution shift like this could have been too severe to be restored by the DA, deteriorating the predicted trajectory profile when tested on the target data. Inherent hardware differences such as noise levels, frequency response or dynamic range, could be the cause of this lack of transferability, as they would impact the recorded signals non-linearly. Future works should therefore analyse sensor specifications to minimise domain discrep-

ancy, which could be achieved through calibration or pre-processing techniques.

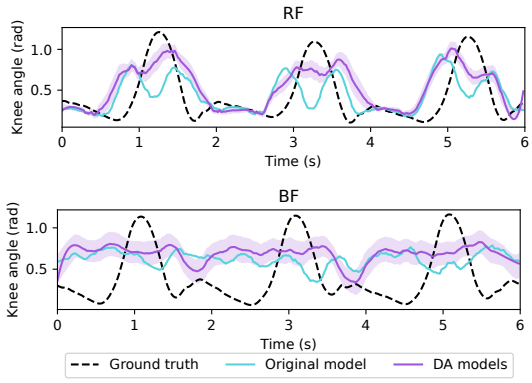


Fig. 9: Predicted knee angle when tested on Delsys for the model trained on HD-EMG only (Original model) vs. the average of the muscle-specific DA models with 3 encoding layers (DA models)

This study reveals the potential of DA-enhanced networks to achieve robustness across different sensor systems, which would enable the exploitation of rich high-density spatial information in applications, such as assistive robotics, where bipolar sensors are preferred. Although MSE loss is commonly employed for regressive mapping, it primarily focuses on the vertical difference between data points rather than preserving the overall shape and physiological meaning of the output. For practical gait trajectory predictions, it is vital to prioritise certain features such as presence of distinct gait events, range of motion, or periodicity, in order to maintain the physiological meaningfulness of the output. Therefore, future works should investigate alternative evaluation metrics or network architectures that can better preserve the signal morphology during DA in order to achieve more meaningful results - this could involve more complex DA frameworks, such as adversarial- [26] or reconstruction-based DA [27]. Considering DA was developed for image classification, exploring class-based gait analysis, specifically gait phase classification, would be a promising direction as it prioritises the physiological significance of the output.

REFERENCES

[1] B. Sheng, L. Tang, S. Xie, C. Deng, and Y. Zhang, "Alterations in muscle activation patterns during robot-assisted bilateral training: A pilot study," *Proceedings of the Institution of Mechanical Engineers. Part H, Journal of engineering in medicine*, vol. 233, no. 2, pp. 219–231, 2019.

[2] M. Rojas-Martínez, M. A. Mañanas, and J. F. Alonso, "High-density surface emg maps from upper-arm and forearm muscles," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, no. 1, p. 85, 2012.

[3] M. J. Zwarts and D. F. Stegeman, "Multichannel surface emg: Basic aspects and clinical utility," *Muscle & nerve*, vol. 28, no. 1, pp. 1–17, 2003.

[4] D. Falla and A. Gallina, "New insights into pain-related changes in muscle activation revealed by high-density surface electromyography," *Journal of electromyography and kinesiology*, vol. 52, p. 102422, 2020.

[5] K. Nizamis, N. H. M. Rijken, R. van Middelaar, J. Neto, B. F. J. M. Koopman, and M. Sartori, "Characterization of forearm muscle activation in duchenne muscular dystrophy via high-density electromyography: A case study on the implications for myoelectric control," *Frontiers in Neurology*, vol. 11, p. 231, 2020.

[6] G. L. Cerone, A. Botter, and M. Gazzoni, "A modular, smart, and wearable system for high density semg detection," *IEEE transactions on biomedical engineering*, vol. 66, no. 12, pp. 3371–3380, 2019.

[7] I. Campanini, A. Merlo, C. Disselhorst-Klug, L. Mesin, S. Muceli, and R. Merletti, "Fundamental concepts of bipolar and high-density surface emg understanding and teaching for clinical, occupational, and sport applications: Origin, detection, and main errors," 2022.

[8] S. Pizzolato, L. Tagliapietra, M. Cognolato, M. Reggiani, H. Müller, and M. Atzori, "Comparison of six electromyography acquisition setups on hand movement classification tasks," 2017.

[9] D. Incorporated, "Wearable sensors for movement sciences," 2023. [Online]. Available: http://www.delsys.com/

[10] J. J. Bird, J. Kobylarz, D. R. Faria, A. Ekárt, and E. P. Ribeiro, "Cross-domain mlp and cnn transfer learning for biological signal processing: Eeg and emg," *IEEE Access*, vol. 8, pp. 54 789–54 801, 2020.

[11] A. Arnold, R. Nallapati, and W. W. Cohen, "A comparative study of methods for transductive transfer learning," in *Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007)*, 2007, pp. 77–82.

[12] W. Xu, J. He, and Y. Shu, "Transfer learning and deep domain adaptation," 2020.

[13] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, *A Brief Review of Domain Adaptation*, ser. Advances in Data Science and Information Engineering. Cham: Springer International Publishing, 2021, pp. 877–894.

[14] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, *Domain-Adversarial Training of Neural Networks*, ser. Domain Adaptation in Computer Vision Applications. Cham: Springer International Publishing, 2017, pp. 189–209.

[15] S. Ma, W. Zhu, and L. Zhang, "Region of interest extraction based on unsupervised cross-domain adaptation for remote sensing images," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 1169–1172.

[16] P. Shi, X. Zhang, W. Li, and H. Yu, "Improving the robustness and adaptability of semg-based pattern recognition using deep domain adaptation," *IEEE journal of biomedical and health informatics*, vol. 26, no. 11, pp. 1–11, 2022.

[17] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 760–771, 2019.

[18] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing (Amsterdam)*, vol. 312, pp. 135–153, 2018.

[19] Z. Wang, Y. Luo, P.-F. Zhang, S. Wang, and Z. Huang, "Discovering domain disentanglement for generalized multi-source domain adaptation," 2022.

[20] O. B. SRL, "Ot bioelettronica," 2023. [Online]. Available: https://www.otbioelettronica.it/en/

[21] S. Day, "Important factors in surface emg measurement," *Bortec Biomedical Ltd publishers*, pp. 1–17, 2002.

[22] H. J. Hermens, B. Freriks, R. Merletti, D. Stegeman, J. Blok, G. Rau, C. Disselhorst-Klug, and G. Hägg, "European recommendations for surface electromyography," *Roessingh research and development*, vol. 8, no. 2, pp. 13–54, 1999.

[23] E. S. Grood and W. J. Suntay, "A joint coordinate system for the clinical description of three-dimensional motions: application to the knee," *Journal of biomechanical engineering*, vol. 105, no. 2, pp. 136–144, 1983.

[24] A. Genevay, G. Peyré, and M. Cuturi, "Learning generative models with sinkhorn divergences," *Proceedings of the 21 st International Conference on Artificial Intelligence and Statistics*, vol. 84, 2017.

[25] J. Wu, Z. Huang, J. Thoma, D. Acharya, and L. V. Gool, "Wasserstein divergence for gans," p. 673, 2018.

[26] X. Gu, X. Yu, Y. Yang, J. Sun, and Z. Xu, "Adversarial reweighting for partial domain adaptation," in *35th Conference on Neural Information Processing Systems*, 2021.

[27] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li, "Deep reconstruction-classification networks for unsupervised domain adaptation," vol. 9908. Springer International Publishing, 2016.