

Correlating Data-Driven Muscle Selection Approaches to Synergies for Gait Prediction

Annika Guez¹, C. Sebastian Mancero Castillo¹, Balint Hodossy¹, Dario Farina¹, *Fellow, IEEE*,
and Ravi Vaidyanathan¹, *Member, IEEE*

Abstract—Optimizing sensors for physiological input is critical to enhance performance as well as minimize the cost and complexity of assistive devices (e.g. lower-limb exoskeletons). Electromyography (EMG) data can trace muscle activation for gait kinematics prediction. However, identifying optimal muscle groups for electrode placement and the potential variance between users has not yet been established. In this study, we use data-driven channel selection techniques on EMG signals to find muscle group combinations that maximize prediction performance. We apply greedy search (Recursive Feature Elimination, RFE) and variance-based (Principal Component Analysis, PCA) methods to select muscle groups during gait, without prior knowledge of musculoskeletal inter-connectivity. The selected muscle subsets are evaluated using the normalized accuracy of a Multi-Layer Perceptron (MLP), mapping muscle activity to knee flexion angle in a one-step-ahead scheme. The RFE selection led to an average predicted knee angle validation accuracy of 4.52 ± 1.85 % higher than the PCA approach, suggesting that dynamic search is more appropriate than a variance analysis of the signals. Whilst the RFE-selected muscle groups differed across subjects, the selected muscles were consistently spread out over more than 80% of the extracted synergy groups. This study underlines the value of incorporating synergistic information when developing gait prediction models, and reveals that maximizing the number of synergy groups could constitute the basis of muscle selection frameworks.

Index Terms—Electromyography (EMG), channel selection, recursive feature elimination, multi-layer perceptron, principal component analysis.

I. INTRODUCTION

ASSISTIVE robotics, such as lower-limb exoskeletons (LLEs), enable users to regain mobility and independence while reducing the workload on medical staff [1]. However, cost, size, and complexity limit their accessibility. Accurate prediction of the user's intended motion is also necessary to improve the efficacy and safety of these wearable devices in rehabilitation [2]. However, it remains unclear which musculoskeletal inputs are most appropriate to achieve this [3].

Electromyography (EMG) is a sensing modality that enables the tracking of user intent, and has often been used as an input signal for predictive models of gait [4], [5], [6]. Whilst multi-channel electromyography (EMG) data across lower-limb muscles is common, there are few studies that investigate which muscle groups are most suited as inputs to these models. When working with multivariate data, channel selection can (a) reduce the computational complexity by processing only relevant channels, (b) mitigate the risk of over-fitting, and (c) reduce the duration of the experimental setup by decreasing the amount of hardware required [7]. Channel reduction also implies the removal of noisy channels, which can improve performance and reduce computational costs by handling the *Curse of Dimensionality*; for LLEs, this would limit cost by reducing the number of sensors [3], [8].

Multiple static selection algorithms based on variance have been previously investigated to select channels with complementary information, with some studies hypothesizing that channels with higher variance lead to better predictions [9]. A common variance-based reduction method is Principal Component Analysis (PCA). This is a well-established technique in multivariate studies that transforms a dataset into ranked principal components, where channels with greater contributions towards the variance in the dataset are likely to exhibit superior performance [10], [11], [12]. While this rationale is commonly used for channel selection of high-density signals, its propriety when applied to signals obtained across the musculoskeletal system, to select muscles for example, has yet to be assessed.

Received 12 December 2023; revised 5 August 2024 and 30 December 2024; accepted 30 December 2024. Date of current version 4 March 2025. This work was supported in part by UK Research and Innovation Centre(s) For Doctoral Training (UKRI CDT) in Artificial Intelligence (AI) for Healthcare under Grant EP/S023283/1; in part by CDT in Prosthetics and Orthotics under Grant EP/S02249X/1; in part by U.K. Dementia Research Institute through the Medical Research Council, in Partnership with Alzheimer's Society under Grant UK DRI-7003 and Grant UK DRI-7005; in part by the Natural BionicS Initiative under Agreement 810346; and in part by the Hamlyn Centre for Robotic Surgery of Imperial College London. (*Corresponding author: Annika Guez.*)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Imperial College Research Ethics Committee under ICREC No. 21IC7204.

Annika Guez, C. Sebastian Mancero Castillo, and Ravi Vaidyanathan are with the Department of Mechanical Engineering and the DRI-Care Research and Technology Centre (CR&T), Imperial College London, SW7 2AZ London, U.K. (e-mail: ag6016@ic.ac.uk).

Balint Hodossy and Dario Farina are with the Department of Bioengineering, Imperial College London, SW7 2AZ London, U.K.

Digital Object Identifier 10.1109/TNSRE.2025.3543743

Dynamic channel selection has been proposed as an alternative approach to PCA, using model performance to evaluate the channels at each epoch and selecting those that maintain the highest accuracy [13]. Greedy search algorithms iteratively evaluate each combination to identify the most relevant input features for a task, resulting in an optimal channel subset. Despite their “brute force” nature, they are computationally efficient compared to exhaustive methods, enabling thorough analysis of the combinations whilst remaining robust against over-fitting [14]. An example of this type of algorithm is the Recursive Feature Elimination (RFE) structure, which uses a drop-out loop to iteratively remove the channel with the least impact on model performance until the desired number of channels is reached [15], [16].

Channel reduction is prevalent in electroencephalography (EEG)-based studies since their multi-channel nature facilitates the simultaneous acquisition of data for comparison [17], [18]. However, despite its effectiveness in other areas, the utilization of this technique for EMG has been limited. The majority of models have been used either with high-density EMG (HD-EMG) or on classification tasks [9], [19], with only a handful of studies exploring channel selection for joint angle regression.

In the context of channel selection, it is crucial to maximize the initial number of channels to ensure the selection is driven by the computational analysis rather than pre-imposed restrictions. Whilst most studies focus on muscles directly involved in joint flexion and extension, trunk muscles are typically considered in the context of gait analysis; clinical studies perceive forward walking as the translation of the body’s centre of mass (CoM) [20], making the lateral weight shift of the trunk essential as it contributes to balance and stability [21]. Trunk accelerometry, for example, has been widely used for gait parameter extraction [22], and numerous clinical studies have advocated for the importance of the trunk in rehabilitation and mobility performance [23], [24]. Whilst the inclusion of the trunk is still limited in EMG-based gait analysis [25], [26], it could provide a more comprehensive understanding of the mechanisms involved in walking.

Furthermore, the analysis of varied muscle groups would facilitate observations of the human body’s inter-connectivity [27]. It has been shown that the Central Nervous System (CNS) uses specific recruitment strategies, known as muscle synergies, to simplify control tasks [28], [29], [30], [31]. Since these co-contraction patterns could impose physiological constraints on muscle coordination [11], incorporating this information would be highly advantageous. Synergies, commonly extracted from EMG envelopes using non-negative matrix factorisation (NMF) [32], refer to the coordinated co-activation of functional muscles to simplify motor control. This also serves as dimensional reduction, with previous studies extracting around 4 to 8 synergy groups for 10 to 15 muscles investigated [33], [34], [35], [36]. Given the interconnected nature of the human system, we hypothesize that the muscle groups achieving the highest performances will be correlated to muscle synergies.

This study applies data-driven channel reduction methods to select optimal muscle groups for knee angle prediction

during gait. We employ a regressive Multi-Layer Perceptron (MLP)-based RFE with a one-step-ahead prediction scheme to select the channels containing the maximum information based on validation performance. This “physiologically uninformed” approach focuses purely on the performance of the investigated signals, enabling channel selection without pre-existing biases. Additionally, we compare this RFE approach against PCA for channel selection. We then investigate whether the selected muscles have any correlation to synergies and whether this physiological framework can help contextualize muscle selection.

It is worth noting that muscle recruitment patterns are not fixed, and the extracted synergies can vary across subjects and gait conditions. For example, walking speed has an impact on explicit gait features, such as joint range of motion or step length [37], influencing the phase and amplitude of muscle activity patterns [38], [39] which are the basis for synergy extraction. Since muscle co-contraction is dependent on walking velocity, which itself varies during rehabilitation, this study also investigates the muscle activation patterns across gait speeds, and how these relate to synergies.

II. METHODS

A. Subjects

Eleven non-disabled subjects (8 male and 3 female, with average ages of 26 ± 3 years, heights of 175.7 ± 10.5 cm, and weights of 68.2 ± 12.4 kg) with no history of neurological or physiological conditions took part in this study. Experiments were performed in compliance with ethical documentation approved by the Imperial College Research Ethics Committee (ICREC reference: 21IC7204).

B. Experimental Setup

1) *Bipolar EMG Sensors*: Delsys Avanti Trigno surface electromyography (sEMG) sensors were placed on 15 muscles of the trunk and lower-limbs to acquire synchronous muscle activity information during walking. For each subject, sEMG data was recorded from the muscles on their dominant leg, consisting of the Rectus Femoris (RF), Vastus Medialis (VM), Vastus Lateralis (VL), Biceps Femoris (BF), Semitendinous (ST), Tibialis Anterior (TA), Soleus (SO), Gastrocnemius Medialis (GM) and Gastrocnemius Lateralis (GL) (total of 9 sensors). The lower-limb muscles were included unilaterally following previous studies [40]. The dominant leg was determined by conducting a push test, and selecting the leg with which the user instinctively stepped forward.

For the trunk, previous studies examined both posterior and anterior trunk muscles alongside lower-limb muscles for synergy extraction [26]. However, due to noise from the fat layer, abdominal muscles were excluded from this study. Therefore, bilateral lower-back muscles were chosen to visualize the lateral shift throughout the gait cycle (the weight is shifted towards the dominant side during the stance phase, and the non-dominant side during the swing phase). Trunk bilateral muscle activity was recorded from the Erector Spinae Longissimus (ESL), Erector Spinae Iliocostalis (ESI), and Multifidus (MF) (total of 6 sensors); throughout this paper,

the muscles on the same side as the subject's dominant leg will be designated with the suffix "-Dom", and on the non-dominant side with the suffix "-Non".

The electrodes were placed according to the SENIAM standards [41]. Sensor location was cleaned, shaved, and dried prior to attaching the sensors to ensure satisfactory contact between the electrodes and the skin. The sensors were fixed using single-use Delsys adhesive interfaces and secured with either micro-porous adhesive tape or medical fabric strapping tape. The sampling frequency of the Delsys data is 2,000 Hz. For all analyses in this study, the EMG signals were band-pass filtered (20-400 Hz), rectified, and low-pass filtered at 10 Hz to extract the signal envelopes. Causal filtering was used.

2) Motion Capture: Knee joint kinematics were acquired from the dominant leg using a Vicon Motion Capture system composed of ten 120 Hz Vicon Vero v2.2 cameras placed around the workspace, all pointing inwards towards a treadmill on which the subjects walked. Motion data was internally upsampled by the system to match that of the Delsys. Marker trajectory data were labeled and gap-filled using the Vicon Nexus software, and kinematics information was extracted by the internal Nexus Plug-In Gait Kinematics pipeline.

The angular trajectory of the dominant knee joint was extracted automatically via the Nexus software, which applied the lower-body modeling of the Plug-in Gait system using the following markers [42]:

- 1) **Pelvis:** Markers were placed on the sacral, the left and right anterior superior iliac spines, and the left and right posterior superior iliac spines.
- 2) **Lower Limb:** Markers were placed on the lower lateral surface of the thigh, the lateral epicondyle, the lower lateral surface of the shank, the lateral malleolus, the calcaneus, and the second metatarsal head. Since the users wore sports shoes during data acquisition, the markers were placed on the shoe as shown in Figure 1. This was repeated for both legs. An additional marker on the medial epicondyle was placed to facilitate gap filling of markers within the knee's vicinity.

Since all the data acquisition was done simultaneously through the analog integrated version of the Delsys Avanti Trigno (EMG) and Vicon Nexus (motion capture), the two data streams were internally synchronised during recording [43].

C. Experimental Procedure

Each subject was asked to walk on the treadmill at steady-state speeds ranging from 0.9mph to 1.8mph ($\approx 1.45\text{km/h}$ to $\approx 2.9\text{km/h}$) for a duration of 2 minutes per speed, with increments of 0.1mph ($\approx 0.16\text{km/h}$). This corresponds to an average of 75 strides per speed, resulting in over 680 steady-state cycles per subject.

D. Principal Component Analysis

Principal Component Analysis (PCA) was performed on the EMG envelopes, and the output muscle selection was employed as a baseline comparison for the RFE approach. The variance-based approach was implemented for a number of principal components ranging from 4 to 8, with the selected

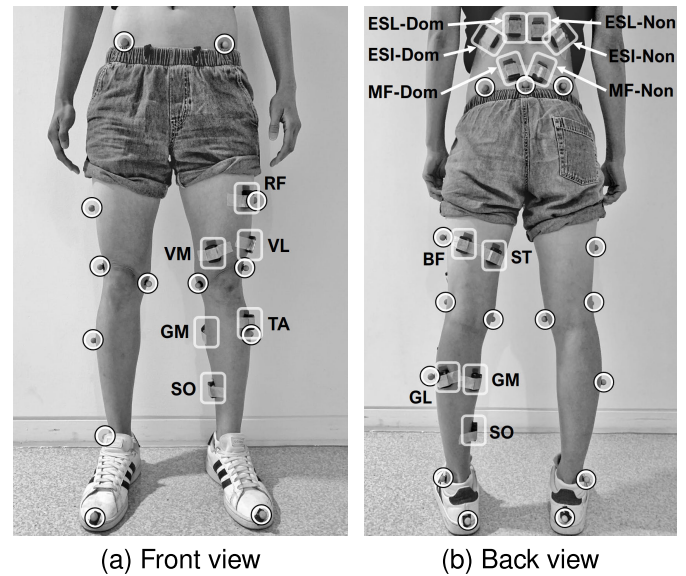


Fig. 1. Experimental setup of a subject with a left dominant leg, showing the Delsys sEMG electrode placements (white rectangles) for the RF, VM, VL, TA, Erector Spinae Longissimus on the dominant side (ESL-Dom), Erector Spinae Iliocostalis on the dominant side (ESI-Dom), Multifidus on the dominant side (MF-Dom), Erector Spinae Longissimus on the non-dominant side (ESL-Non), Erector Spinae Iliocostalis on the non-dominant side (ESI-Non), Multifidus on the non-dominant side (MF-Non), BF, ST, GL, GM, SO placed on the *a)* front and *b)* back of the subject. The Vicon Motion capture markers used to determine the joint angles are also shown, circled in white.

channels corresponding to those who had the highest coefficient of co-linearity with the extracted principal components (i.e. the channels with a coefficient closest to 1). To avoid selection duplicates, the number of extracted components were iteratively increased until the required number of unique channels was reached.

E. Recursive Feature Elimination Using an MLP

An RFE algorithm was employed to select channels based on prediction performance of the validation set. This method consists of training a model on all available channels, recording the maximum achieved performance, and then iteratively removing one channel at a time and training again; the channel with the least contribution to accuracy (i.e. whose removal leads to the smallest drop in performance) was eliminated, and the process was repeated until only 1 channel remained. For the purpose of this study, the removed channels were set to zero to keep the shape of the input data consistent throughout.

EMG envelopes were extracted (see Subsection II-B1) and normalized based on the training set, then segmented into windows of approximately 0.25 seconds in length (512 samples due to the 2 kHz sampling frequency). Each window was then paired with the knee angle value one time-step ahead of the end of the window. While previous studies have used larger prediction windows to account for the electromechanical delay between the EMG signal and the resulting movement [44], the optimal prediction horizon remains uncertain. Therefore, this one-step-ahead prediction scheme was employed to ensure that horizon selection did not bias the results, allowing muscle subset selection to depend solely on the information contained within the window preceding the motion.

TABLE I

MODEL PARAMETERS APPLIED TO THE MLP MODELS OF THE RFE,
WITH THE INITIAL LR AND THE MAXIMUM NUMBER OF EPOCHS

Condition	Description	Initial LR	Window step
All Speeds	Trained/tested on all combined speeds, using 3-fold cross-validation	5×10^{-5}	40
Fast Cadence	Trained/tested on the fast cadence, using a 3-fold cross-validation	5×10^{-5}	40
Medium Cadence	Trained on the fast cadence and tested on the medium cadence	1×10^{-6}	20
Slow Cadence	Trained on the fast cadence and tested on the slow cadence	1×10^{-6}	20

A regressive MLP was used as the evaluating machine learning model, mapping the instantaneous sEMG envelopes to the knee angle of the subject's dominant leg. The chosen model architecture was composed of 7 fully connected layers with a Rectified Linear Unit (ReLU) activation, trained using an Mean Square Error (MSE)-based loss and an Adam optimizer, to convert the windowed EMG input into a singular one-step-ahead angular output. All models were ran for 50 epochs. Every time the learning stagnated on the validation data past 3 epochs, the learning rate (LR) was tuned to a tenth of its previous value, and the training resumed from the model parameters that had given the best validation performance up to that point. Since the model output is regressive, the percentage of error in this paper is given as the angular error normalised to the overall angular range of motion (RoM) of the knee during the investigated walking trial, which means that the performance values correspond to the maximum accuracy minus the percentage error (i.e. 100% - error percentage). This normalization step was implemented to enable meaningful comparisons between subjects who may use different ranges of knee flexion during their gait.

Model information is given in Table I, outlining the different evaluation conditions investigated, along with their chosen initial LR and the windowing stride (referred to as the "Window step"). The dataset used for the "All speeds RFE" analysis was composed of all the recorded gait speeds concatenated together, with testing data reserved for each speed to ensure a representative evaluation.

The gait data at varying speeds was segmented into the following three conditions: slow cadence (0.9mph to 1.1mph inclusive, or $\approx 1.45\text{km/h}$ to $\approx 1.77\text{km/h}$), medium cadence (1.2mph to 1.4mph inclusive, or $\approx 1.9\text{km/h}$ to $\approx 2.3\text{km/h}$) and fast cadence (1.5mph to 1.8mph inclusive, or $\approx 2.4\text{km/h}$ to $\approx 2.9\text{km/h}$). Model training and RFE were repeated for all three conditions. Since the training and testing sets are sampled from the same original dataset for both the *All Speeds* and *Fast Cadence* conditions, a 3-fold cross-validation scheme was employed to reduce the effect of the selected validation dataset on the quantified performance.

Since the majority of healthy gait data tends to be collected at a normal walking speed (corresponding to the fast cadence), the slow and medium cadences were analyzed to test the generalisability and performance of the model for a walking

speed range expected in rehabilitation environments, which would be significantly slower due to the user's physical impairments. For the *Medium* and *Slow Cadence* conditions, the window step value was reduced from 40 to 20 to increase the available data, and therefore the chances of convergence despite the domain shift between the training (fast cadence) and testing (medium and slow cadences) datasets. The initial LR was also reduced based on the effectiveness and stability of the training process.

For the *All Speeds* and *Fast Cadence* conditions, 360 models were ran with 3-fold cross-validations, resulting in a total of 3960 models trained. On the other hand, for the *Medium Cadence* and *Slow Cadence*, 120 models were trained for each subject, resulting in a total of 1320 models. All these models were ran with PyTorch. To enable faster computation time, the subjects were parallelised across multiple nodes using the High Performance Computing clusters of Imperial College London, with each node requesting 8 CPUs with 64GB memory and a RTX6000 GPU. Per subject, the cadence models were ran within 48 hours whilst the 3-fold models required 60 hours to be completed. For each trained model, the recorded performance corresponds to the highest validation accuracy achieved during training.

F. Synergy Extraction

A NMF approach was employed where the weights matrix $\mathbf{W}$ and synergistic activation coefficients matrix $\mathbf{H}$ were extracted from the multi-channel EMG input $\mathbf{E}$ in order to minimize the Euclidean distance between the matrix product and the original dataset (see Equation 1).

$$\min_{\mathbf{W}, \mathbf{H}} \left(\frac{1}{n} \sum_{n=1}^N \|\mathbf{W} \times \mathbf{H}_n - \mathbf{E}_n\|^2 \right) \\ \times \mathbf{W} \in \mathbb{R}^{C \times S}, \mathbf{H} \in \mathbb{R}^{S \times N}, \mathbf{E} \in \mathbb{R}^{C \times N} \quad (1)$$

where C is the number of EMG channels, S is the number of synergies, and N is the number of samples. The weightings were then normalized within each synergy, and a paired ranking and threshold approach was used to select the muscles belonging to each synergy, with a maximum of 5 muscles per synergy. The threshold value was set at 0.5, and would be slightly adjusted in increments of 0.05 if the total number of muscles included were less than 10 (ignoring valuable muscle groups) or above 14 (muscle groups often duplicated across synergies).

Since the number of synergies extracted in the literature often ranges from 4 to 8 [33], [34], [35], [36], the extraction procedure was repeated for 4 to 7 synergy groups for the given 15 muscle groups. The synergistic muscle subsets were then compared to the muscle subsets chosen by the RFE to assess the similarity between the two groups. For *Syn* extracted synergies, *Syn*+1 RFE-selected muscle groups were investigated; in cases, during the search, where the performance difference between two muscle groups was minimal, this narrow margin ensured that essential muscle groups were not excluded.

Considering the distribution of the muscles in the synergy groups is predetermined, a statistical analysis was conducted to quantify the significance of the overlap between the RFE

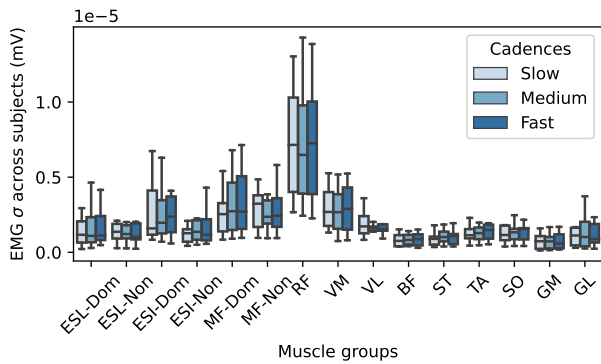


Fig. 2. Cross-cycle standard deviation (σ) of the EMG signal envelopes, where the variance was calculated specific to each speed and then merged to represent the average variance found within a given cadence. These results include all subjects.

subsets and the synergy subsets; for a given number of synergies Syn , the probability that $Syn + 1$ randomly selected muscles are spread across the synergy groups was explicitly computed, by counting occurrences for all possible overlap scenarios.

III. RESULTS

A. Principal Component Analysis

A PCA approach was investigated as it is considered a benchmark for channel selection in literature [45], and is therefore a good basis for comparison with the RFE. A subset was selected using PCA for a range of 4 to 8 muscles, with which the MLP model training and evaluation was performed on the dataset combining all steady-state speeds following the *All Speeds* condition of Table I. When looking at the muscles chosen by the PCA, although there is a lot of variation across subjects on which 5 muscle groups are preferred, the RF was the only muscle group that was consistently included in the subset, followed by some trunk muscles including the MF-Non and the ESI-Dom.

Since this is a variance-based approach, a variance analysis of the EMG's signal envelopes was performed for each muscle across all steady-state speeds, to determine whether EMG signal consistency contributed to favoring certain muscles above others when using PCA. For each subject, the standard deviation of each muscle's EMG envelope across speeds was calculated and normalized to the average, and the resulting muscle-specific variance is illustrated in Figure 2, grouped by cadence and given across all subjects.

Whilst PCA attempts to capture the explainable variance of a dataset to extract its most representative features, it appears to be favouring muscle groups that are more inconsistent across the gait cycles and speeds. There is therefore a strong correlation between the calculated variance of a signal and its inclusion in the PCA selection, with the RF exhibiting the largest variance across all cadences.

Although previous studies have frequently used variance for channel selection, they do not assess its suitability as the optimal channel selection approach, particularly when it comes to selecting the muscle inputs for gait prediction models. Due

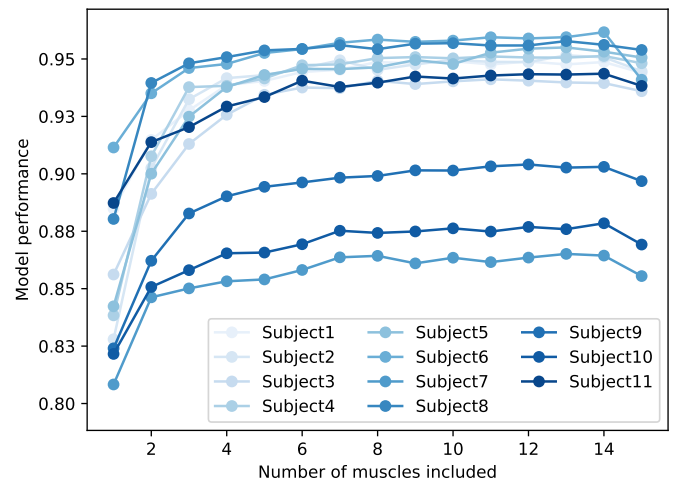


Fig. 3. Validation accuracy (Root Mean Square Error (RMSE) normalized by the RoM) across subjects of the MLPs from the RFE framework, with respect to the number of channels (i.e. muscle groups) included. These results were obtained using the *all speeds* RFE described in Table I.

to its data-driven nature, the RFE algorithm represents an unbiased extensive search tool that is capable of selecting channels based on their performance in the prediction model, through iterative evaluation of all channel combinations to determine the optimal muscle subset.

The validity of PCA was therefore evaluated by comparing the performance of the models trained using the PCA selection against those trained on the RFE selected subset. For the same number of channels included, the overall accuracy achieved was lower when using the PCA selected muscles compared to the ones chosen by the RFE, with an average validation accuracy drop of $1.81 \pm 0.26\%$. The PCA's lower performance suggests that variance-based channel selection may not be appropriate in the context where the channels correspond to different muscle groups. The proposed data-driven RFE can therefore be considered as a reliable method for muscle-subset selection of inputs to gait prediction models, and will therefore be the main channel selection approach for the rest of this study. The selected PCA subsets will nonetheless be plotted on the selection Figures 4 and 9 for illustration purposes.

B. Recursive Feature Elimination

The RFE search was performed for the *All Speeds* condition outlined in Table I to incorporate the variation of muscle activation patterns across walking speeds during channel selection. As seen in Figure 3, the MLP models achieved their highest validation accuracy when 5 or more muscle groups were included as inputs to the knee angle prediction, reaching an average cross-subject validation accuracy of $93.188 \pm 3.216\%$ for 14 electrodes, and maintaining a performance within $< 1\%$ of that maximum value while reducing the number of channels down to 5.

As expected, there was a noticeable decrease in accuracy when using 4 electrodes or less, as the restricted amount of information was insufficient to achieve high model

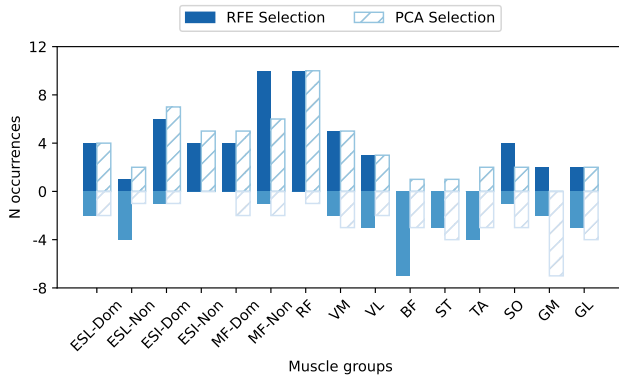


Fig. 4. The preferred (i.e. the last 5 to be eliminated) and avoided (i.e. first 3 to be eliminated) muscle groups based on the *all speeds* subset selection of the both the RFE and PCA. Here, the positive bars correspond to the preferred muscles, and the negative bars represent the avoided ones.

performance. The drop observed when using all 15 channels, however, suggests that the added channels were either noisy or redundant, and hindered the network’s generalization on validation data, therefore reinforcing the benefits of channel selection.

Figure 4 illustrates the muscle groups across subjects that tend to be preferred (belong to the 5 muscle subset selected by the RFE) or avoided (belonging to the first 3 muscles to be eliminated by the RFE). Based on the results, there seems to be a strong preference for the RF and trunk muscles (mainly ESI and MF), compared to an aversion towards the BF and TA, which could be due to their inconsistent activation across different gait speeds [46]. When looking at similarities across all subjects, only the RF was unanimously selected across all subjects in their preferred 6 muscles. If we look at the performance stability range (from 5 to 14 muscle groups), 10 out of the 11 subjects included the MF-Non in their top 5, the ESI-Dom and MF-Dom in their top 10, and the ESI-Non and SO in their top 12.

Figure 5 shows the co-activation of the MF muscles with the RF, despite cross-subject differences in activation time which could be due to varied postural control strategies. The occasionally earlier timing or narrower peak of the RF could explain its consistent inclusion, as it provides more precise information. However, the bilateral oscillation of the MF could also explain why it is often paired with the RF, as they provide complementary information to one another.

Whilst the selection is not distinct enough to conclude on the 5 muscle groups that are optimal for effective kinematics prediction out of the original 15 muscles investigated, the observed tendency in Figure 4 suggests there exists a pattern regarding which muscle groups should be included (such as the RF and the trunk). Since the RFE is purely data-driven, it is “physiologically uninformed” to any physiological information that could impact the evaluated channels. A synergy analysis was therefore performed to investigate whether underlying mechanisms of the musculoskeletal system could have influenced the selection process, therefore allowing us to better interpret the observed results.

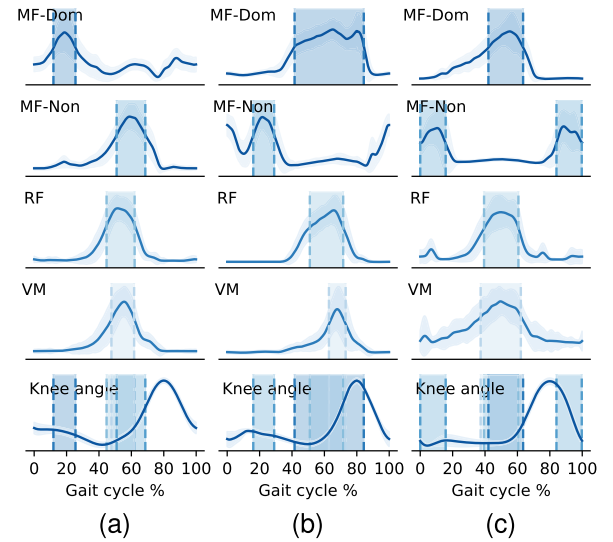


Fig. 5. Average EMG signal envelopes of the MF-Dom, MF-Non, RF, and VM muscles across a gait cycle for 3 example subjects (a, b, c). The main activation points are highlighted for each muscle group and projected onto the knee joint angle. All information is given as a function of the gait cycle percentage.

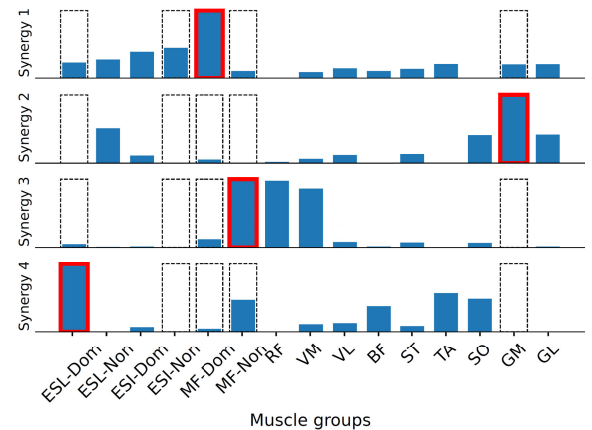


Fig. 6. Example of the overlap between the RFE selected muscles and the extracted synergy groups that determines the inclusion percentage, in the case of 4 extracted synergies for Subject 3. The rectangular boxes illustrate the RFE selected subset, with a red outline when the chosen muscle coincides with one of the synergies. Here, Synergy 1 relates to the stabilization during the stance phase, Synergy 2 to the propulsion at the end of the stance phase, Synergy 3 to the stabilization of the knee during heel strike, and Synergy 4 facilitates foot clearance during the swing phase.

C. Correlating Results to Synergies

For 15 muscle groups investigated, we chose to extract between 4 to 7 synergies per subject (similarly to [33]), and compared them to the RFE selection. Figure 6 illustrates how the selected subset of 5 muscle groups correlate to the 4 synergies extracted from the EMG signals of each subject. In this example, the RFE selection consistently included the dominant muscle in every extracted synergy group, exhibiting a strong correlation between synergies and the RFE subset.

The inclusion percentage within the RFE (i.e. the number of synergies included in the selection, where we define inclusion based on whether one of the synergy’s muscles was selected)

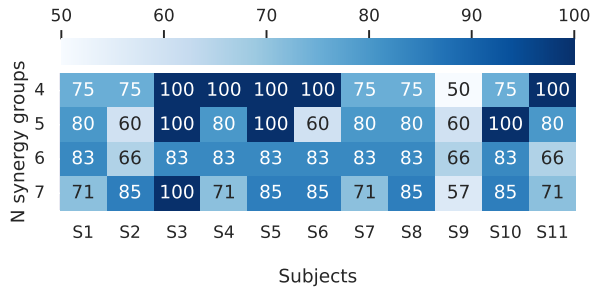


Fig. 7. Inclusion percentage of the extracted synergy groups in the RFE selected subset, given for all the subjects (S1 - S11) and number of extracted synergy groups (4-7).

was calculated for all subjects and number of extracted synergy groups to assess this correlation further. The results are tabulated in Figure 7, with “100%” inclusion defined as the case where all the synergies are included in the RFE (for example the case shown in Figure 6).

The average percentages of extracted synergies included in the muscle selection of the RFE are summarised below:

- $84.09 \pm 16.85\%$ for 4 synergies
- $80.00 \pm 15.49\%$ for 5 synergies
- $78.79 \pm 7.78\%$ for 6 synergies
- $79.22 \pm 11.72\%$ for 7 synergies

As seen in Figure 7, the synergy inclusion percentage of Subject 9 (i.e. S9) appears to be noticeably lower than for other subjects. If we omit Subject 9, the average inclusion percentages become: $87.50 \pm 13.17\%$ for 4 synergies, $82.00 \pm 14.76\%$ for 5 synergies, $80.00 \pm 7.03\%$ for 6 synergies, and $81.43 \pm 9.64\%$ for 7 synergies.

Following the statistical analysis outlined in Subsection II-F, the probability that a randomly selected muscle subset would include a large number of the extracted synergies was calculated. Assuming that the $Syn + 1$ selected muscles from the RFE were completely at random, the likelihood of them lying in $> Syn - 1$ synergies for $4 \leq Syn \leq 7$ is given below:

- **For 4 extracted synergies:** $35.70 \pm 5.69\%$ chance of including 75% of the synergies (i.e. to 3 out of 4)
- **For 5 extracted synergies:** $47.99 \pm 3.60\%$ chance of including 60% of the synergies (i.e. 3 out of 5), and $24.09 \pm 6.25\%$ chance of including 80% of the synergies (i.e. 4 out of 5)
- **For 6 extracted synergies:** $43.23 \pm 5.11\%$ chance of including 66.6% of the synergies (i.e. 4 out of 6), and $15.12 \pm 5.41\%$ chance of including 83.3% of the synergies (i.e. 5 out of 6)
- **For 7 extracted synergies:** $35.47 \pm 7.32\%$ chance of including 71.43% of the synergies (i.e. 5 out of 7), and $11.84 \pm 6.67\%$ chance of including 85.71% of the synergies (i.e. 6 out of 7)

Considering that the RFE’s selected muscles were consistently spread across more than 80% of the synergy groups, we can conclude that this discrepancy is significant enough to rule out randomness, rendering the correlation between muscle subsets and synergies meaningful. Using synergies for muscle selection could therefore help simplify the identification

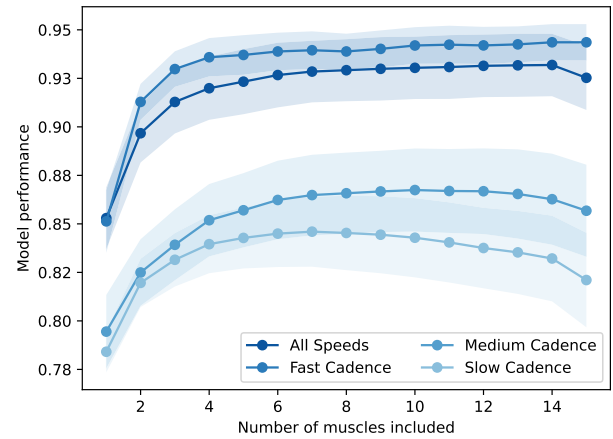


Fig. 8. Validation accuracy of the RFE with respect to the number of channels included. This plot shows the comparative performance of the RFE models for the 3 conditions outlined in Table I, with half the cross-subject standard deviation σ .

of optimal muscle inputs for gait predictors, whilst achieving comparable prediction accuracies to those from the RFE (which consistently outperformed PCA-selected muscles) since they account for the complex musculoskeletal dynamics present in the data.

D. Impact of Gait Speeds

It has been shown that the body’s kinematics change across walking speeds, potentially impacting the extracted muscle synergies [47]. The investigation into the generalizability of models, when trained on normal walking speeds and tested on slower cadences, is particularly relevant as it illustrates the expected behavior of patients during rehabilitation. This study therefore investigated a range of walking speeds that were lower than that of the subject’s natural gait speed, to simulate the expected application environment. Furthermore, slower walking speeds could challenge muscle synergies that are shared between postural control and walking, since slower speeds force the patient to put a larger emphasis on balance rather than propulsion [48]. As the subjects walking speed varied from 0.9mph to 1.8mph ($\approx 1.45\text{km/h} - 2.9\text{km/h}$), the muscle activation patterns underwent both an amplitude and phase shift, which impacted the EMG signals individually as well as their relationship to one another.

1) *On the RFE:* The RFE process was repeated when trained on the fast cadence and then tested on either the medium or slow cadences, as described in Table I as the *Medium Cadence* and *Slow Cadence* conditions. Similarly to the *All Speeds* RFE, there was a noticeable decrease in model performance when there were both too few (< 5) or too many (> 13) electrodes included, as shown in Figure 8. The impact of the cross-speed variability in the EMG signals on model performance is also apparent, as the model trained and tested purely on the fast cadence outperforms all the other conditions, and its accuracy is already established when using only 4 muscle groups.

Compared to the RFE trained on all steady-state speeds, the overall performance of the *Slow* and *Medium Cadence* conditions was noticeably lower. For the model tested on the

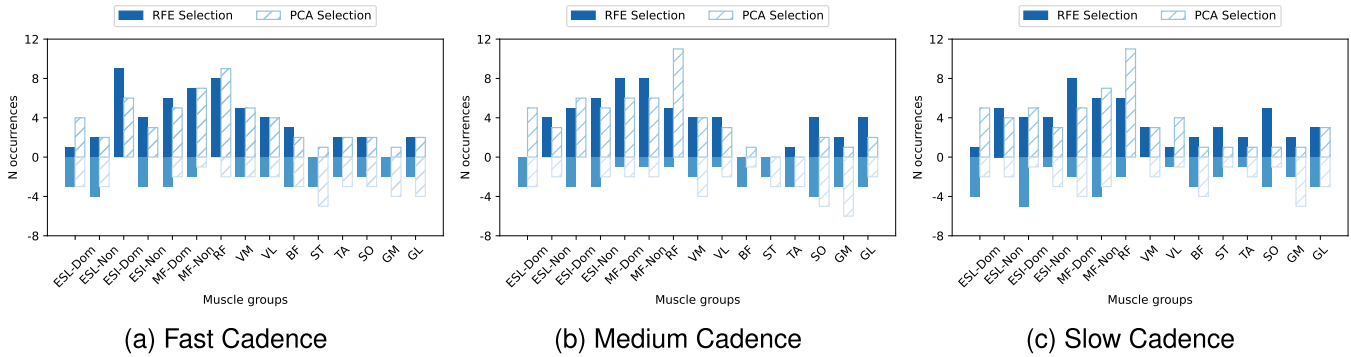


Fig. 9. The preferred (i.e. those belonging to the selected subset of 5 muscles) and avoided (i.e. the first 3 muscles to be eliminated) muscle groups based on the cadence-based RFE's output selection, showing the (a) *Fast cadence*, (b) *Medium cadence*, and (c) *Slow cadence*.

medium cadence, results yielded a maximum average cross-subject performance of $86.743 \pm 4.282\%$ for 10 channels, which was maintained within a $< 1\%$ margin between 6 to 14 channels. For the *Slow Cadence* condition, the RFE resulted in a maximum accuracy of $84.599 \pm 3.610\%$ for 7 channels, which was maintained within the 1% margin between 4 to 12 channels. For both these cases, the performance degradation could be a consequence of the shift in activation patterns across speeds; the observed discrepancy in both the shape and amplitude of the EMG envelopes created a domain mismatch between the training and testing datasets, therefore hindering the model's ability to generalize.

It is interesting to note that, as the speeds of the testing dataset progressively differed from the training one, the maximum number of channels for which the model's performance is stable decreased accordingly; for the *All Speeds* and *Fast Cadence* conditions, performance is maintained within the margin when including all 15 channels, whereas this number decreases to 14 channels for the *Medium Cadence*, and then only 12 for the *Slow Cadence*. This would suggest that the uncertainty induced by the EMG signal's variability across speeds further reinforces the need for channel selection.

Despite the drop in performance when attempting to generalize across cadences, the RFE selected muscles still seem to lean once more towards the inclusion of trunk muscles and the RF (see Figure 9), with a slight emphasis on shank muscles at slower speeds. Interestingly, the PCA subsets are comparable to those of the RFE at higher speeds, but progressively differ as the walking speed slows down, suggesting that simple variance analysis cannot capture the complexity of muscle recruitment strategies across speeds.

2) On the Synergies: The number of synergies included in the muscle subset of the RFE also changed as the muscle activation patterns were altered. When extracting the synergies from the different cadences, the weighting of certain muscle groups changed within their respective synergistic groups; the trunk muscles gained importance at slower speeds in synergies related to stability and posture, whilst there was an increase of shank muscle weights at higher speeds during propulsion. The discrepancy between the synergy weightings and RFE selection (see Figure 8) suggests that muscles with higher synergistic weightings may not be more effective predictors due to potential signal inconsistency.

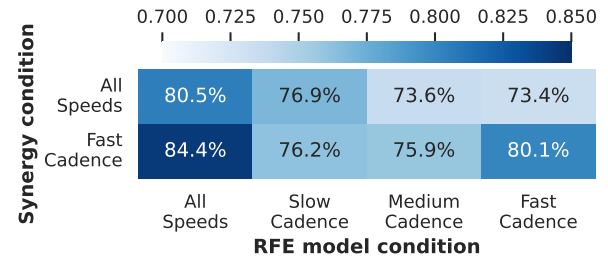


Fig. 10. Percentage of synergies included in the RFE selection across all investigated conditions. The y-axis refers to the condition of the dataset from which the synergy groups were extracted, whilst the x-axis refers to the RFE model condition, as specific in Table I. These results are averages across all subjects and synergy groups ranging from 4 to 7 extracted synergies.

Figure 10 shows the percentage of synergy groups that were included in the RFE selection, for all investigated conditions, when compared with the synergy groups extracted from the combined speeds and the fast cadence. As expected, the inclusion percentage of the synergies in the RFE-selected subsets decreased for the cadence based conditions compared to the *All Speeds* condition. However, when extracting synergies from the fast cadence and comparing them with the RFE subsets across the investigated conditions, the inclusion percentage was found to be higher in every case, suggesting that the synergy groups extracted when the subject is walking at a natural pace are more reliable, corroborating the notion of unnatural muscle recruitment strategies at slower speeds.

Whilst these synergy inclusion percentages remain significantly higher than the random probabilities calculated earlier in the statistical analysis, this preliminary observation underlines the need for further focus on cross-speed analysis in the context of synergy extraction optimization. Understanding and integrating the way gait speed affects muscle recruitment, and therefore predictors trained on muscle activity data would facilitate the development of a muscle selection strategy that is robust to gait speeds. This would therefore enable us to select the most appropriate speed range when conducting data acquisition experiments for reliable synergy extraction, which can then serve as a baseline for optimal muscle selection.

IV. DISCUSSION

The aim of this study was to determine which muscle subsets, from an initial 15 EMG signals extracted from both the trunk and the lower-limb were the most reliable as inputs to gait prediction models. Following previous studies on channel selection methods, a variance-based PCA approach was employed as well as a data-driven MLP-based RFE algorithm to ascertain the optimal muscle combination. Despite the popularity of PCA as a benchmark for channel-selection, the RFE selection yielded an overall higher performance accuracy than the PCA subsets, and was therefore deemed the more reliable approach in this context. This would imply that variance-based methods are not suitable when the channels correspond to different muscle groups, which could be due to the physiological interactions and dependencies across the musculoskeletal system that are not present when performing channel selection on high-density set-ups that investigate a single muscle separately.

The performance achieved by the MLP models within the RFE architecture were similar across subjects and behaved comparably, reaching their maximum performance between 5 and 14 channels for all steady-state speeds, corroborating the notion that 5 muscle groups can be sufficient to reach high predictive performance of gait kinematics [49].

Similarly to the PCA, the RFE selection also gravitated towards the RF, however consistently paired it with the trunk muscles. This could indicate that the pairing offered complementary information to the model [14], which cannot be identified using a simple PCA approach. Moreover, this preference towards the inclusion of trunk muscles would suggest that they contain valuable information for gait models, as seen in previous clinical studies [20]. Considering the importance of the trunk in both clinical and rehabilitative settings, as well as the ease of its usability in LLEs since the muscles are located away from the investigated limb, future works should further investigate the combination of trunk and lower-limb muscles as gait predictors.

Despite certain patterns and preferences visible in the selected RFE subset (see Figure 4), these remained inconclusive across subjects, and the mechanistic reasoning behind the chosen muscles groups were further investigated using synergies. Despite being given no physiological information or constraints, the physiologically uninformed data-driven RFE had a propensity for synergistic selections (see Figure 7); this not only reinforces the notion that synergies provide valuable information on body kinematics and motion intent [50], but also reveals the importance of including a mechanistic context when identifying optimal neuromuscular inputs to predict gait. This correlation of data-driven approaches to physiological elements is similar to that of previous studies, where the RFE method managed to extract physiologically meaningful channels from EEG signals corresponding to the left and right hands respectively [16]. Therefore, by determining antagonist muscle subsets derived from synergies, one could potentially achieve higher prediction performance by maximizing the number of included synergies.

This is particularly advantageous in the context of LLE control, especially if the muscles within the same synergy

group can be used interchangeably; knowledge of the subject's recruitment strategy could constitute a reliable framework for muscle selection, offering flexibility on which specific muscles to include, provided they are distributed across the extracted synergies. This, therefore, offers a channel reduction strategy that is highly adaptable and straightforward to implement, enabling us to efficiently select reliable EMG signals for gait prediction whilst the remaining muscle groups could be included for passive processes such as monitoring or diagnosis.

Whilst the correlation between optimal muscle inputs and key physiological features is promising, it should be noted that not all possible combinations of muscle groups across synergies were evaluated. This would be a crucial step to verify whether performance remains consistent regardless of the muscle used within a synergy, and whether there are some muscle groups within a given synergy that are consistently preferred or avoided, as suggested by Figure 4. Future works should conduct an exhaustive evaluation of all muscle combinations that include all investigated synergy groups to assess the coherence of their performances, hence evaluating their interchangeability. This would require the extraction of synergies that are reliable, for example by using appropriate gait speeds as suggested in Subsection III-D, and generalisable across subjects in order to streamline this approach. It should also be noted that walking speed has a strong impact on both the model performance and the subject's recruitment strategies, which in turn affects the reliability of synergy-guided channel selection. Future works should therefore focus on performing large clinical studies to determine the synergistic groups relevant to gait prediction that are generalizable across both subjects and conditions.

Whilst previous works have used larger prediction horizons [44], [51], this study focused on a one-step-ahead prediction scheme, relying on the window size to capture sufficient information prior to the motion. Future works should investigate the impact of different prediction horizons on the selection. Moreover, as this current study focused on steady-state walking, they should also explore more challenging locomotion tasks that could amplify performance discrepancies, and therefore potentially change the subsets obtained from the RFE.

V. CONCLUSION

The aim of this study was to determine whether there exists a muscle subset that would invariably achieve high prediction accuracy when mapping the EMG signals to the knee angle. We employed data-driven methods to ensure the selection was unbiased, focusing solely on the performance of the signals without any additional context. Using a MLP as the evaluating model, both PCA and RFE approaches were investigated for input channel selection, with each channel corresponding to a muscle group from the trunk or the dominant lower-limb. For all relevant channel subsets, RFE-selected muscles outperformed the PCA selections, revealing that variance-based methods are not appropriate when selecting muscle groups, despite their success in EEG and HD-EMG studies. This could be due to the physiological implications, such as interactions and dependencies, that arise when investigating

multiple muscle groups concurrently. Despite inconclusive results regarding RFE muscle subsets, a preference for the RF paired with trunk muscles was observed, along with an avoidance of the BF and TA. This suggests that the combination of trunk and lower-limb muscles should be investigated further, as they appear to provide complementary information that could be highly beneficial to gait predictors. Synergies were extracted and compared with the RFE muscles as a way to account for recruitment strategies and the inter-connectivity of the human body system, rather than evaluating them independently. Despite having given no prior physiological context to the model, results show a strong correlation between the RFE's muscle selection and the subject's synergies, with the selected subset distributed across more than 80% of the extracted synergy groups. This implies synergies could be used as a basis for input optimization when predicting user kinematics, and that one should aim to maximize the number of synergy groups included in model inputs to have a reliable and representative dataset and achieve higher performance. Future works should therefore assess the interchangeability of muscles within synergies, and identify whether specific muscles are consistently favored over others within a given synergy. This should be performed on large clinical datasets to additionally assess the impact of different conditions on these muscle recruitment strategies (such as speed, gait mode, or pathology) to assess the generalisability of this synergy-based technique.

REFERENCES

- [1] W. Huo, S. Mohammed, J. C. Moreno, and Y. Amirat, "Lower limb wearable robots for assistance and rehabilitation: A state of the art," *IEEE Syst. J.*, vol. 10, no. 3, pp. 1068–1081, Sep. 2016.
- [2] M. R. Tucker et al., "Control strategies for active lower extremity prosthetics and orthotics: A review," *J. NeuroEng. Rehabil.*, vol. 12, no. 1, p. 1, 2015.
- [3] B. Wei et al., "A novel sEMG-based gait phase-kinematics-coupled predictor and its interaction with exoskeletons," *Frontiers Neurobotics*, vol. 15, Aug. 2021, Art. no. 704226.
- [4] C. Caulcrick, W. Huo, W. Hoult, and R. Vaidyanathan, "Human joint torque modelling with MMG and EMG during lower limb human-exoskeleton interaction," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 7185–7192, Oct. 2021.
- [5] C. Caulcrick, W. Huo, E. Franco, S. Mohammed, W. Hoult, and R. Vaidyanathan, "Model predictive control for human-centred lower limb robotic assistance," *IEEE Trans. Med. Robot. Bionics*, vol. 3, no. 4, pp. 980–991, Nov. 2021.
- [6] C. Frigo and R. Shiavi, *Applications in Movement and Gait Analysis*. Hoboken, NJ, USA: Wiley, 2004, ch. 15, pp. 381–401.
- [7] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, pp. 1–21, Dec. 2015.
- [8] W.-F. Xu, Y.-F. Fang, G.-Y. Zhang, Z.-J. Ju, G.-F. Li, and H.-H. Liu, "Surface Emg channel selection for thumb motion classification signal," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, vol. 2, Jul. 2018, pp. 662–666.
- [9] Y. Geng, X. Zhang, Y.-T. Zhang, and G. Li, "A novel channel selection method for multiple motion classification using high-density electromyography," *Biomed. Eng. OnLine*, vol. 13, no. 1, p. 102, Dec. 2014.
- [10] I. S. Dhindsa, R. Agarwal, and H. S. Ryait, "Principal component analysis-based muscle identification for myoelectric-controlled exoskeleton knee," *J. Appl. Statist.*, vol. 44, no. 10, pp. 1707–1720, Jul. 2017.
- [11] A. Guo, "Channel selection in emg-based hand gesture estimation," Ph.D. dissertation, Dept. Comput. Sci., Eng., Northeastern Univ. Boston, Boston, MA, USA, 2019.
- [12] 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.
- [13] A. Al-Ani, I. Koprinska, G. R. Naik, and R. N. Khushaba, "A dynamic channel selection algorithm for the classification of EEG and EMG data," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 4076–4081.
- [14] I. Guyon, C. Aliferis, and A. Elisseev, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.
- [15] E. L. Glassman and J. V. Gutttag, "Reducing the number of channels for an ambulatory patient-specific EEG-based epileptic seizure detector by applying recursive feature elimination," in *Proc. 28th IEEE EMBS Annu. Int. Conf.*, Sep. 2006, pp. 2175–2178.
- [16] T. N. Lal et al., "Support vector channel selection in BCI," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003–1010, Jun. 2004.
- [17] A. Soler, L. A. Moctezuma, E. Giraldo, and M. Molinas, "Automated methodology for optimal selection of minimum electrode subsets for accurate EEG source estimation based on genetic algorithm optimization," *Sci. Rep.*, vol. 12, no. 1, p. 11221, Jul. 2022.
- [18] A. Goshvarpour and A. Goshvarpour, "A novel approach for EEG electrode selection in automated emotion recognition based on lagged poincare's indices and sLORETA," *Cognit. Comput.*, vol. 12, no. 3, pp. 602–618, May 2019.
- [19] J. Liu, X. Li, G. Li, and P. Zhou, "EMG feature assessment for myoelectric pattern recognition and channel selection: A study with incomplete spinal cord injury," *Med. Eng. Phys.*, vol. 36, no. 7, pp. 975–980, Jul. 2014.
- [20] D. Jarchi, J. Pope, T. K. M. Lee, L. Tamjidi, A. Mirzaei, and S. Sanei, "A review on accelerometry-based gait analysis and emerging clinical applications," *IEEE Rev. Biomed. Eng.*, vol. 11, pp. 177–194, 2018.
- [21] H. Hsiao, V. L. Gray, R. A. Creath, S. A. Binder-Macleod, and M. W. Rogers, "Control of lateral weight transfer is associated with walking speed in individuals post-stroke," *J. Biomech.*, vol. 60, pp. 72–78, Jul. 2017.
- [22] L. Tesio and V. Rota, "The motion of body center of mass during walking: A review oriented to clinical applications," *Frontiers Neurol.*, vol. 10, p. 999, Sep. 2019.
- [23] M. M. Mañago, J. R. Hebert, J. Kittelson, and M. Schenkman, "Contributions of ankle, knee, hip, and trunk muscle function to gait performance in people with multiple sclerosis: A cross-sectional analysis," *Phys. Therapy*, vol. 98, no. 7, pp. 595–604, Jul. 2018.
- [24] L. F. Quintino, J. Franco, A. F. M. Gusmão, P. F. D. S. Silva, and C. D. C. D. M. Faria, "Trunk flexor and extensor muscle performance in chronic stroke patients: A case-control study," *Brazilian J. Phys. Therapy*, vol. 22, no. 3, pp. 231–237, May 2017.
- [25] J.-C. Ceccato, M. de Sèze, C. Azevedo, and J.-R. Cazalets, "Comparison of trunk activity during gait initiation and walking in humans," *PLoS ONE*, vol. 4, no. 12, Dec. 2009, Art. no. e8193.
- [26] T. Wojtara, F. Alnajjar, S. Shimoda, and H. Kimura, "Muscle synergy stability and human balance maintenance," *J. NeuroEng. Rehabil.*, vol. 11, no. 1, p. 129, 2014.
- [27] A. Oshima, Y. Nakamura, and K. Kamibayashi, "Modulation of muscle synergies in lower-limb muscles associated with split-belt locomotor adaptation," *Frontiers Human Neurosci.*, vol. 16, Jun. 2022, Art. no. 852530.
- [28] P. Azaripasand, A. Maleki, and A. Fallah, "The extraction of time-varying muscle synergies during hand-reaching movement with a K-means assisted approach," in *Proc. 23rd Iranian Conf. Electr. Eng.*, May 2015, pp. 84–87.
- [29] C. S. M. Castillo, R. Vaidyanathan, and S. F. Atashzar, "Synergistic upper-limb functional muscle connectivity using acoustic mechanomyography," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 8, pp. 2569–2580, Aug. 2022.
- [30] A. d'Avella, P. Saltiel, and E. Bizzi, "Combinations of muscle synergies in the construction of a natural motor behavior," *Nature Neurosci.*, vol. 6, no. 3, pp. 300–308, Mar. 2003.
- [31] M. C. Tresch, V. C. Cheung, and A. d'Avella, "Matrix factorization algorithms for the identification of muscle synergies: Evaluation on simulated and experimental data sets," *J. Neurophysiol.*, vol. 95, no. 4, pp. 2199–2212, Apr. 2006.
- [32] S. Tanzarella, S. Muceli, M. Santello, and D. Farina, "Synergistic organization of neural inputs from spinal motor neurons to extrinsic and intrinsic hand muscles," *J. Neurosci.*, vol. 41, no. 32, pp. 6878–6891, Aug. 2021.

- [33] Y.-X. Liu and E. M. Gutierrez-Farewik, "Muscle synergies enable accurate joint moment prediction using few electromyography sensors," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Piscataway, NJ, USA, Sep. 27, 2021, pp. 5090–5097.
- [34] S. Hagio, M. Fukuda, and M. Kouzaki, "Identification of muscle synergies associated with gait transition in humans," *Frontiers Human Neurosci.*, vol. 9, p. 48, Feb. 2015.
- [35] H. Yokoyama, T. Ogawa, N. Kawashima, M. Shinya, and K. Nakazawa, "Distinct sets of locomotor modules control the speed and modes of human locomotion," *Sci. Rep.*, vol. 6, no. 1, p. 36275, Nov. 2016.
- [36] H. Yokoyama, N. Kaneko, T. Ogawa, N. Kawashima, K. Watanabe, and K. Nakazawa, "Cortical correlates of locomotor muscle synergy activation in humans: An electroencephalographic decoding study," *iScience*, vol. 15, pp. 623–639, May 2019.
- [37] G. Soleimani, M. E. Andani, H. Marateb, and F. Bahrami, "Speed influences on complicated kinetic, kinematic and electromyographic gait data of lower extremities among healthy subjects," in *Proc. 23rd Iranian Conf. Biomed. Eng. 1st Int. Iranian Conf. Biomed. Eng. (ICBME)*, Nov. 2016, pp. 321–326.
- [38] A. Barrett, A. E. Donnelly, and G. O'Laighin, "Tibialis anterior EMG activation pattern changes with walking speed during over ground and treadmill walking," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2007, pp. 4883–4886.
- [39] L. Chen, P. Yang, L. Zu, and Y. Geng, "The influence of walking speed on muscle activity of thigh and application in prostheses control," in *Proc. IECON-36th Annu. Conf. IEEE Ind. Electron. Soc.*, Nov. 2010, pp. 2714–2718.
- [40] N. Amrani El Yaakoubi, C. McDonald, and O. Lennon, "Prediction of gait kinematics and kinetics: A systematic review of EMG and EEG signal use and their contribution to prediction accuracy," *Bioengineering*, vol. 10, no. 10, p. 1162, Oct. 2023.
- [41] H. J. Hermens et al., "European recommendations for surface electromyography," *Roessingh Res. Develop.*, vol. 8, no. 2, pp. 13–54, 1999.
- [42] V. M. Systems, *Vicon Plug-in Gait Reference Guide*, 2nd ed. Oxford, U.K.: Vicon Motion Systems, 2023. [Online]. Available: <https://www.vicon.com/trademarks/>
- [43] *Delsys Trigno and Vicon Nexus Integration*, document 361-2-3, Delsys, Apr. 2023.
- [44] C. Yi et al., "Continuous prediction of lower-limb kinematics from multi-modal biomedical signals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2592–2602, May 2021.
- [45] K.-M. Ong, K.-H. Thung, C.-Y. Wee, and R. Paramesran, "Selection of a subset of EEG channels using PCA to classify alcoholics and non-alcoholics," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, Jan. 2005, pp. 4195–4198.
- [46] K. Fujita, H. Hori, and Y. Kobayashi, "Contribution of muscle activity at different gait phases for improving walking performance in chronic stroke patients with hemiparesis," *J. Phys. Therapy Sci.*, vol. 30, no. 11, pp. 1381–1385, 2018.
- [47] H. Takahashi et al., "Spatial and temporal correlation of muscle synergies in human walk-run transition," in *Proc. Int. Symp. Micro-NanoMechatronics Human Sci. (MHS)*, Dec. 2017, pp. 1–5.
- [48] B. Kibushi, S. Hagio, T. Moritani, and M. Kouzaki, "Speed-dependent modulation of muscle activity based on muscle synergies during treadmill walking," *Frontiers Human Neurosci.*, vol. 12, p. 4, Jan. 2018.
- [49] B. Xiong et al., "Intelligent prediction of human lower extremity joint moment: An artificial neural network approach," *IEEE Access*, vol. 7, pp. 29973–29980, 2019.
- [50] D. Xiong, D. Zhang, X. Zhao, Y. Chu, and Y. Zhao, "Synergy-based neural interface for human gait tracking with deep learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 2271–2280, 2021.
- [51] Q. Bian, M. Castellani, D. Shepherd, J. Duan, and Z. Ding, "Gait intention prediction using a lower-limb musculoskeletal model and long short-term memory neural networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 822–830, 2024.