Skeleton-based Muscle Activity Prediction during Manual Material Handling

Jaekyeong Moon
Department of Industrial Engineering
Yonsei University
Seoul, Republic of Korea
Industrial Transformation Technology Department
Korea Institute of Industrial Technology
Cheonan, Republic of Korea
mesrwi@yonsei.ac.kr

Jong Hyun Kim, Chang Gi Lee, Hyunchul Tae
Industrial Transformation Technology Department
Korea Institute of Industrial Technology
Cheonan, Republic of Korea
{ddalki, leechangki, sage}@kitech.re.kr

***Abstract*— Electromyography (EMG) signals, which measure muscle electrical activity, are used to assess muscle function, activity, and fatigue levels during motion. However, collecting EMG data typically requires invasive or non-invasive sensors in controlled environments. This paper proposes a model to predict muscle activity levels from videos of people performing actions in real-world settings, enabling workplace applications. EMG signals and corresponding videos were collected from seven sites on the right side of subjects' bodies performing three actions with two weights. After preprocessing, muscle activity levels and joint coordinates were extracted to predict muscle activity levels.**

***Keywords*—Muscle acitivty prediction, Electromyography, Skeleton graph, Multimodal.**

I. Introduction

The transition to flexible, high-mix production systems has prompted the adoption of robot-assisted production to optimize workspace efficiency. In Korea, where the industrial structure is rapidly changing due to low birth rates and an aging workforce, there is an increasing need for technologies that enable humans and robots to coexist and collaborate in shared spaces. This is particularly critical for small and medium sized enterprises to enhance productivity and flexibility while compensating for labor shortages and declining workforce skills. Alongside advancements in production technologies, ensuring the safety of workers interacting with industrial robots is a key challenge. For instance, robots failing to detect workers entering their pathways may lead to physical collisions. Moreover, human workers, unlike robots that maintain consistent work intensity, are susceptible to accidents caused by reduced attention and work-related musculoskeletal disorders (WMSDs) resulting from excessive physical fatigue. Safety management technologies that mitigate these risks while maintaining productivity are therefore essential.

Improper workplace ergonomics, causing excessive workload, is one of the primary causes of WMSDs. Traditionally, workplace evaluations rely on tools such as REBA (Rapid Entire Body Assessment) and RULA (Rapid Upper Limb Assessment) to identify risk factors and suggest improvements. However, these methods have limitations: they are conducted periodically by experts, making continuous monitoring challenging, and they evaluate only representative or high-risk postures, failing to capture the broader distribution of working postures. To address these limitations, physiological signals that directly reflect workers' physical loads offer a promising solution. Electromyography (EMG) signals, which measure the electrical activity of muscles during contraction, provide detailed information about muscle activity. For example, an increase in muscle fiber activity leads to higher signal amplitude, while muscle fatigue is associated with slower recovery and a predominance of low-frequency components. These characteristics make EMG signals a valuable quantitative indicator for evaluating physical workload and posture during manual tasks.

Despite their utility, collecting EMG signals typically requires surface or invasive sensors, which may not be feasible in dynamic industrial environments. This study investigates an alternative approach, using camera-based systems like workplace-installed CCTV to estimate EMG signals. By incorporating pose estimation techniques, the proposed method aims to enable continuous monitoring of workers' physical loads in a practical and sensor-free manner.

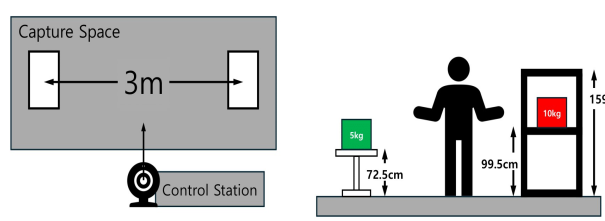

Figure 1. Schema of data collection.
(left: view from the ceiling, right: view from the front)

This paper is based on research conducted under the project 'Development of holonic manufacturing system for future industrial environment (EO240002)' funded by the Clean Production System Core Technology Research Project of the Korea Institute of Industrial Technology (KITECH) in 2024.

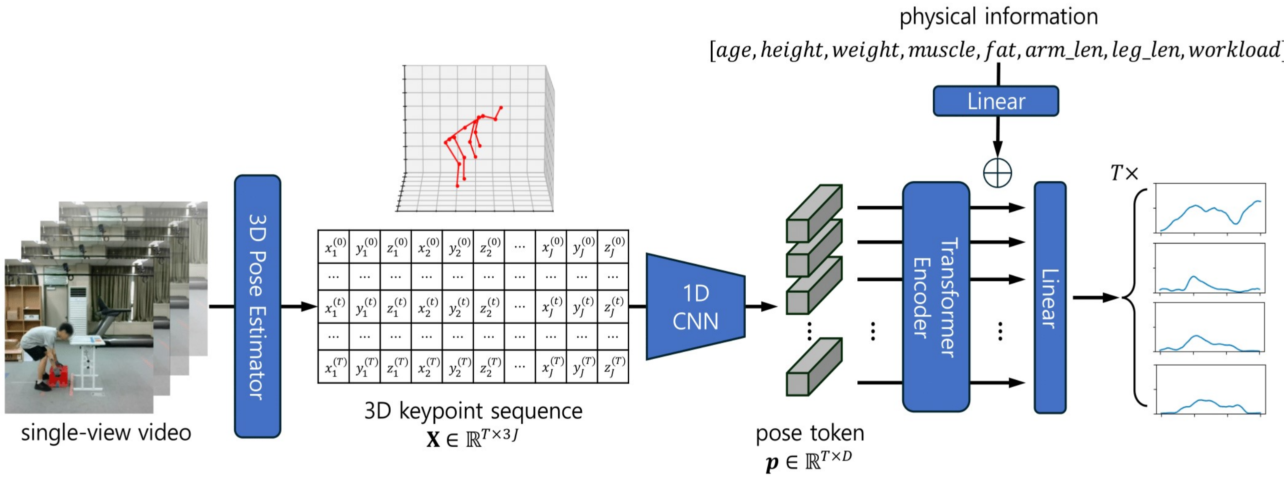

Figure 2. Framework of muscle activity prediction model.

II. Method

A. Data Collection

The lack of publicly available datasets containing synchronized EMG signals and human motion data for manual material handling tasks necessitated the development of a controlled experiment to collect relevant data. This experiment aimed to gather data that could train and evaluate the proposed prediction model, focusing specifically on predicting the amplitude of EMG signals. The schema for data collection is shown in Fig. 1.

Sixteen male participants in their 20s, with no known physical disabilities and complete limb functionality, were recruited for this study. Their demographic distributions were as follows: age ranged from 20 to 29 years, with a mean of 23.75 years (SD: 1.29); height ranged from 166 to 183 cm (mean: 179.44 cm, SD: 4.43); and weight ranged from 52 to 107 kg (mean: 73.08 kg, SD: 14.26). The experiments were conducted in a motion analysis laboratory, providing a controlled environment to minimize external variability while replicating scenarios of manual material handling tasks. Although this setup ensured consistency, it did not fully capture the complexities of real-world work environments.

Five repetitive tasks were designed to simulate common material handling activities: (1) lifting a load from the ground, (2) lowering a load to the ground, (3) carrying a load across a fixed distance, (4) placing a load on a higher platform, and (5) removing a load from a higher platform. To simplify the experimental design, tasks 1 and 2 were combined into a continuous sequence, as were tasks 4 and 5. Each task was conducted under two weight conditions, 5 kg and 10 kg, resulting in six distinct experimental setups: 5 kg and 10 kg for combined tasks 1 & 2, 5 kg and 10 kg for task 3, and 5 kg and 10 kg for combined tasks 4 & 5. Participants were instructed to perform each task in synchronization with periodic auditory signals to maintain consistency in execution. To ensure participant safety, a rest period was provided after each experimental setup, with the next task being conducted only after participants reported full recovery through an interview. Unlike predefined repetition counts or durations, the number of repetitions was determined by participants' subjective assessments of exertion, which were quantified using the Borg CR-10 scale. Tasks were terminated when participants reported a score of 10, indicating a "maximal" effort level. This approach allowed the collection of data spanning a wide range of physical states, from low fatigue to high fatigue. EMG signals were recorded from seven muscle groups associated with material handling tasks: Erector Spinae (ES), Upper Trapezius (UT), Biceps Brachii (BB), Flexor Digitorum Superficialis (FDS), Extensor Digitorum (ED), Biceps Femoris (BF), and Vastus Lateralis (VL) [1] – [5]. Surface EMG sensors (Trigno Avanti sensors, Delsys) were attached to each muscle site following SENIAM guidelines [6] to ensure accurate and reliable data collection. Simultaneously, motion data was captured using a webcam operating at 20 frames per second (fps) to record participants' movements during the tasks. The recorded motion data (images) were processed to extract 2D keypoints using AlphaPose [7], a two-dimensional pose estimation algorithm. These 2D keypoints were subsequently augmented into 3D coordinates using MotionBERT [8] , a transformer-based model designed for human motion representation. The EMG acquisition system and the webcam were synchronized to enable frame-level alignment between the motion data and the muscle activity signals. Plus, prior to the main experiment, the maximum voluntary contraction (MVC) of each participant was measured to normalize the EMG signals. Isometric con tractions were performed for each of the seven selected muscle groups. The measured MVC values were subsequently used to extract and normalize the EMG signal amplitudes recorded during the tasks. The experimental protocol was approved by the institutional review board (IRB), and all participants provided informed consent before participating in the study.

B. Pose2Muscle

The proposed model, Pose2Muscle, predicts frame-wise muscle activity levels from sequences of 3D keypoints. It is based on the framework introduced in the [9], which demonstrated the feasibility of estimating muscle activity from skeletal motion. However, this study extends the approach to account for interactions with external objects, such as lifting or carrying weights, making it suitable for manual material handling tasks.

An overview of the architecture is shown in Fig. 2. The input to the model is a sequence of 3D keypoints, $x \in \mathbb{R}^{T \times 3J}$, where T is the number of frames in the sequence, and J is the number of keypoints in the 3D skeleton. First, the model applies 1D convolution operations to generate pose tokens, $p \in \mathbb{R}^{T \times D}$, where each token represents a frame enriched with temporal information from adjacent frames. Here, D denotes the dimensionality of the model. These pose tokens are then passed through a Transformer encoder with self-attention mechanisms to model temporal relationships across frames. The output of the model consists of T predicted muscle activity values for each of the seven predefined muscle groups, corresponding to the input sequence.

While [9], which focused on bodyweight-only movements, incorporated participant IDs as conditioning features to account for individual differences in muscle activity, relying solely on IDs may be insufficient when considering the diverse impacts of external objects across participants. For instance, lifting a 10 kg weight requires greater muscle force compared to lifting a 5 kg weight, leading to increased muscle fiber recruitment. This physiological response is reflected in higher EMG signal amplitudes, which are quantified as muscle activity levels. For this reason, the Pose2Muscle model integrates additional participant-specific physical attributes, including age, height, weight, muscle mass, fat mass, forearm length, and shin length. These physical attributes are processed through a linear layer to generate embeddings, which are subsequently concatenated with the outputs of the Transformer encoder before being passed to the final prediction layer. By combining skeletal motion data with physical attributes, the model achieves a more comprehensive representation of the factors influencing muscle activity.

III. EXPERIMENTS

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Experimental Setup

The Pose2Muscle model was trained and evaluated using the dataset collected in the experimental study. For each of the six continuous task sequences collected per participant, 20% of the data was randomly extracted and reserved for validation and testing, while the remaining 80% was used for training. This ensured that the training and evaluation datasets were strictly separated for every task across all participants. The model was trained for 300 epochs using the Adam optimizer, with a batch size of 128 and a learning rate of 0.0001. The mean squared error (MSE) was used as the loss function to optimize the regression task. All training and evaluation processes were performed on a single NVIDIA A6000 GPU.

B. Evaluation Metrics

To assess the model's performance, two evaluation metrics were employed:

- R^2 Score: This metric quantifies how well the model explains the variance in the ground truth muscle activation values, providing a comprehensive measure of prediction quality for regression tasks.
- Segment Match: Given that repeated movements often exhibit small differences in posture but noticeable variations in muscle activation, the predicted and actual muscle activation values were categorized into five predefined segments: below 5%, 5–15%, 15–35%, 35–50%, and above 50%. The segment match metric calculates the proportion of predictions where the predicted and actual values fall within the same segment.

C. Results

The average R^2 score on the validation dataset was 0.66535, demonstrating the model's ability to effectively predict muscle activation levels. The R^2 scores for each of the seven muscle groups are shown in Fig. 3. The R^2 scores revealed notable differences in the prediction performance across muscle groups. High R^2 scores were observed for ES and VL, which may be attributed to the relatively simple activation patterns of these muscles in response to large joint movements. For example, ES shows consistent activation when the spine flexes during forward bending, while VL activates when the knees flex during squatting movements. In contrast, lower R^2 scores were recorded for UT and BF. These muscles exhibit more variable activation patterns across participants. For instance, during object lifting, activation of BF depends on whether the spine remains neutral, while activation of UT varies based on the individual's posture and upper body mechanics. Additionally, muscles such as BB demonstrated moderate prediction performance, potentially due to the diverse range of arm movements, which increases the complexity of modeling their activation patterns. These results suggest that while the model

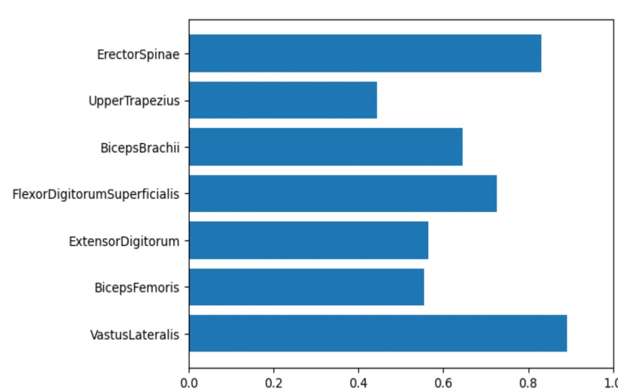

Figure 3. R^2 scores across muscle groups.

effectively captures activation patterns for muscles with consistent activation mechanics, its performance is hindered for muscles with high inter-participant variability or diverse movement patterns.

To further evaluate the model's performance, three participants were randomly selected, and the segment match was visualized for the three primary experimental conditions (lifting, carrying, and lowering tasks). Table I provides the segment match scores for each task, and Fig. 4 – 6 visualizes the segments to which the predicted and actual muscle activation values belong. These visualizations demonstrate that the model achieves a high level of match across repeated movements.

IV. Conclusion

This study proposed Pose2Muscle, a framework for predicting muscle activation levels during manual material handling tasks from 3D pose sequences. By leveraging advanced pose estimation and temporal modeling techniques, the model effectively captures the complex relationships between skeletal motion and muscle activation. Unlike prior studies, which focused solely on bodyweight-only movements, this research addresses the additional complexities of interacting with ex ternal objects, such as lifting or carrying weights, making it more applicable to real-world industrial settings. Through experiments, the Pose2Muscle model achieved an average R^2 score of 0.66535 across all muscle groups, with particularly strong performance for muscles such as the Erector Spinae and Vastus Lateralis.

The findings of this study have practical implications for workplace safety and ergonomic assessments. By predicting muscle activation levels without requiring sensors, the proposed framework offers a scalable solution for monitoring workers' physical states and identifying potential risks of musculoskeletal disorders. This represents a significant step toward integrating motion analysis and muscle activation prediction, providing a foundation for developing safer and more efficient collaborative environments in industrial settings.

TABLE I. SEGMENT MATCH ACROSS MUSCLE GROUPS AND TASKS

Muscle Group	Segment Match (%)		
	Lifting (Task 1 & 2)	***Carrying (Task 3)***	***Lowering (Task 4 & 5)***
Erector Spinae	90.65	69.38	88.5
Upper Trapezius	97.32	83.23	85.75
Biceps Brachii	88.69	79.17	85.88
Flexor Digitorum Superficialis	92.32	90.42	83.38
Extensor Digitorum	92.2	89.38	81.12
Biceps Femoris	100.0	85.42	92.38
Vastus Lateralis	90.89	85.42	74.5
Average	93.15	83.20	84.5

Meanwhile, the model faced challenges in predicting muscle activation for muscles with high inter-participant variability, such as the Upper Trapezius and Biceps Femoris. This highlights the importance of incorporating additional factors, such as movement dynamics or contextual features, to improve the model's predictive accuracy for such muscles. Future research could explore further extensions of the model, such as integrating dynamic task-level features or expanding the dataset to include more diverse populations and working conditions.

Acknowledgment

This paper is based on research conducted under the project 'Development of holonic manufacturing system for future industrial environment (EO240002)' funded by the Clean Production System Core Technology Research Project of the Korea Institute of Industrial Technology (KITECH) in 2024.

References

[1] S. Kim, et al., "Assessing the influence of a passive, upper extremity exoskeletal vest for tasks requiring arm elevation: Part I–"Expected" effects on discomfort, shoulder muscle activity, and work task performance," Applied ergonomics, 70, pp. 315-322, 2018.

[2] K. Hyusamen, et al., "Evaluation of a passive exoskeleton for static upper limb activities," Applied ergonomics, 70, pp. 148-155, 2018.

[3] P. Maurice, et al., "Objective and subjective effects of a passive exoskeleton on overhead work," IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28(1), pp. 152-164, 2019.

[4] M. Claramunt-Molet, et al., "Biomechanical evaluation of upper limb exoskeletons in automotive assembly using EMG," Reunión Del Capítulo Español de La Sociedad Europea de Biomecánica, pp. 25-26, 2019.

[5] S. Iranzo, A. Piedrabuena, D. Iordanov, U. Martinez-Iranzo, and J. M. Belda-Lois, "Ergonomics assessment of passive upper-limb exoskeletons in an automotive assembly plant," Applied ergonomics, 87, 103120, 2020.

[6] H. J. Hermens, et al., "European recommendations for surface electromyography," Roessingh research and development, 8(2), pp. 13-54, 1999.

[7] H. S. Fang, et al., "Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time," IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(6), pp. 7157-7173, 2022.

[8] W. Zhu, et al., "Motionbert: A unified perspective on learning human motion representations," In Proceedings of the IEEE/CVF International Confernce on Computer Vision, pp. 15085-15099, 2023.

[9] M. Chiquier and C. Vondrick, "Muscles in action," In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.22091-22101, 2023.

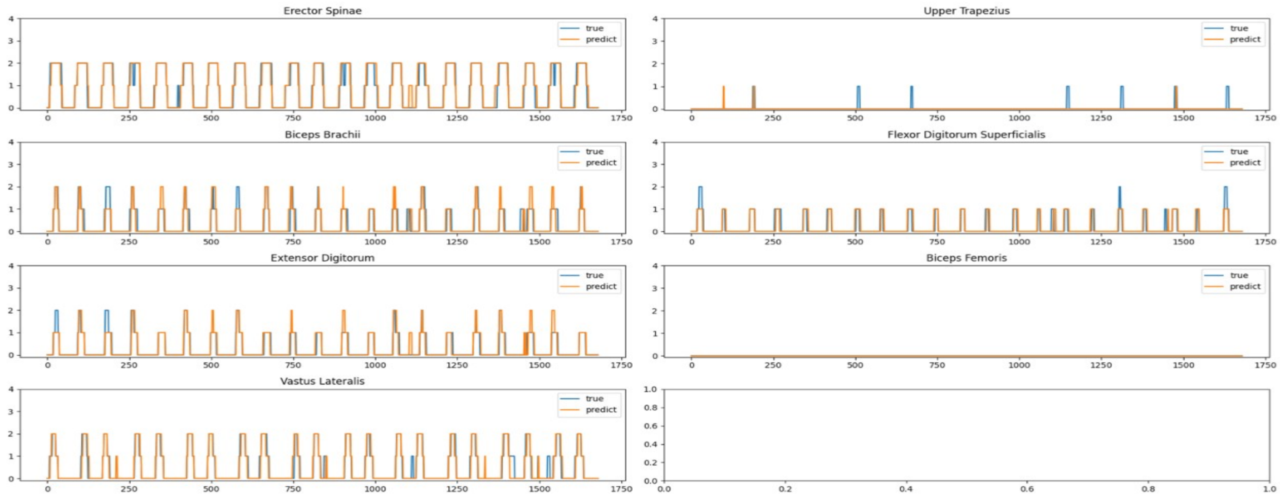

Figure 4. Actual(blue) and predicted(orange) segment across muscle groups during lifting task.

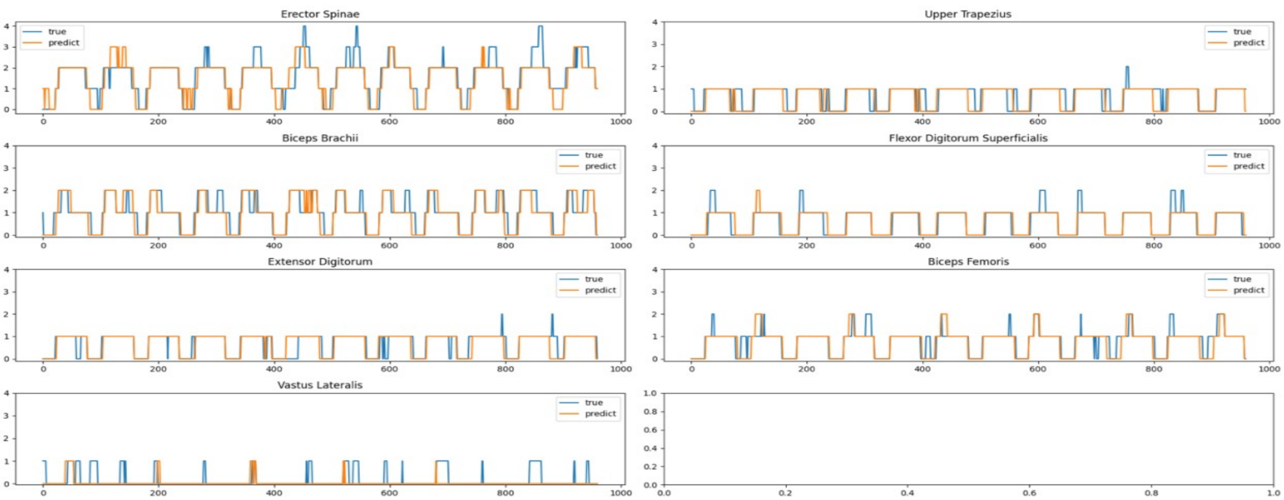

Figure 4. Actual(blue) and predicted(orange) segment across muscle groups during carrying task.

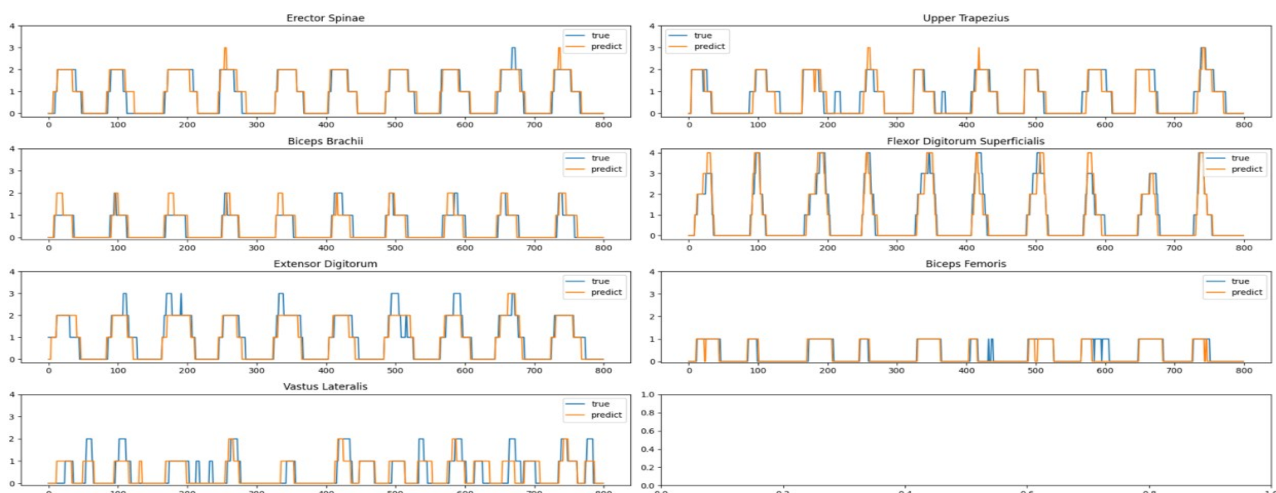

Figure 6. Actual(blue) and predicted(orange) segment across muscle groups during lowering task.