Mechanical sensors

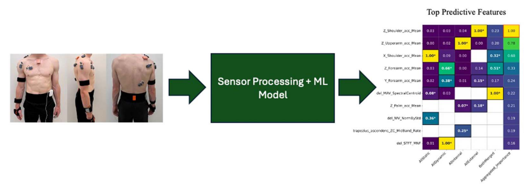
Generalizing Perceived Fatigue Estimation Across Diverse Upper Limb Tasks Using Minimal Wearable Sensors

Malik Muhammad Qirtas[1], Marco Sica[2], Merve Nur Yasar[3], Patricia O'Sullivan[1,2], Brendan O'Flynn[1,2], Salvatore Tedesco[2*], Matteo Menolotto[2*], and Andrea Visentin[1]

[1] *Insight RI Research Centre for Data Analytics, School of Computer Science, IT, University College Cork, T12 K8AF Cork, Ireland*
[2] *Tyndall National Institute, University College Cork, T12 K8AF Cork, Ireland*
[3] *Department of Coaching Education, Faculty of Sports Sciences, Sakarya University of Applied Sciences, Sakarya 54050, Türkiye*
[*] *Member, IEEE*

Manuscript received 11 July 2025; revised 24 July 2025; accepted 28 July 2025. Date of publication 7 August 2025; date of current version 21 August 2025.

Abstract—Accurately estimating perceived fatigue from wearable sensor data is a challenge, especially across diverse tasks. This letter presents a generalized framework to predict estimated fatigue scores (measured using the Borg scale) using combined electromyography and inertial measurement units data collected from two independent upper limb datasets. Our best model achieved a mean absolute error of 2.35 and a mean absolute percentage error of 18.60% using only five strategically placed sensors. A broad set of biomechanical features was extracted to capture both kinematic and neuromuscular indicators of fatigue. Vertical acceleration of the upper arm and shoulder, along with spectral features from deltoid EMG, emerged as the most consistent predictors across tasks. These findings support interpretable and generalizable fatigue detection and provide a foundation for real-time monitoring systems in sports, rehabilitation, and occupational health.



Index Terms—Mechanical sensors, EMG, fatigue estimation, IMU, machine learning, sensors.

I. INTRODUCTION

Physical fatigue is a complex physiological and psychological state that affects performance, recovery, and risk of injury in a variety of activities, from sports and rehabilitation to daily tasks and workplace demands [1]. The shoulder joint is particularly vulnerable to fatigue in overhead sports and repetitive upper limb tasks, where sustained exertion can negatively affect movement patterns, joint stability, and overall biomechanical efficiency [2].

Traditionally, fatigue has been assessed through subjective scales, such as the Borg rating of perceived exertion (RPE) [3]. Although widely used, these subjective ratings can vary significantly due to individual perceptions and personal biases. Recent developments in wearable technologies, such as Electromyography (EMG) and inertial measurement unit (IMU), have provided objective alternatives to continuously monitoring fatigue in real time [4]. EMG captures neuromuscular signals related to muscle activation and fatigue, while IMU can track movement kinematics. Other portable systems, such as near-infrared spectroscopy and ultrasound imaging, provide insights into muscle oxygenation and deformation but are often bulky, power-intensive, and sensitive to motion. In contrast, our combined IMU and EMG setup provides a lightweight and practical solution for an objective estimation during dynamic tasks.

Combining subjective fatigue ratings with objective data collected through sensors can provide a more comprehensive understanding of fatigue progression. However, achieving robust and generalizable fatigue estimation remains challenging due to variability in individual responses and task-specific characteristics. Moreover, practical deployment of fatigue monitoring systems requires minimizing sensor usage to improve user comfort, reduce costs, and extend battery life.

In this study, we address these challenges by developing machine learning models to predict Borg RPE scores from EMG and IMU sensor data across multiple upper limb resistance tasks. We used data from two independently collected datasets with multiple movement types and resistance levels. By merging these datasets, our models aim to generalize the estimation of perceived fatigue and reduce the impact of individual variability. The following are the main contributions.

1) We evaluate fatigue prediction performance across 14 distinct upper limb tasks with static and dynamic movements at varying resistance levels.
2) We develop a single fatigue prediction model trained on combined datasets (66 subjects) to show generalizability across various tasks and resistance levels.
3) We identify a minimal subset of wearable sensors sufficient to maintain high prediction accuracy to support practical deployments.

To the best of the authors' knowledge, this is the first study to integrate and generalize fatigue prediction across distinct upper limb datasets using a minimalistic sensor approach.

II. METHODOLOGY

A. Datasets and Preprocessing

This study used two independent datasets on upper limb movement collected at University College Cork: the Rotation Dataset and the

Corresponding author: Malik Muhammad Qirtas (e-mail: mqirtas@ucc.ie).
Associate Editor: André Eugenio E. Lazzaretti.
Digital Object Identifier 10.1109/LSENS.2025.3596719

Flexion Dataset. The datasets and code used in this study are publicly available.[1]

Rotation Dataset: The Rotation Dataset collected data from 34 healthy (age: 26 ± 4 years), physically active adults (23 males and 11 females) at the Tyndall National Institute's Wearable Laboratory, University College Cork, Ireland, between April and July 2023 [5].[2] Participants performed repeated internal (IR) and external (ER) shoulder rotation tasks at three resistance levels: 30%–40%, 40%–50%, and 50%–60% of their maximum voluntary isometric contraction (MVIC). Each of the six movement–resistance conditions was treated as a separate task and continued until participants reached maximum perceived exertion (RPE 20). EMG signals (1000 Hz) were recorded from six muscles (pectoralis major, anterior deltoid, infraspinatus, posterior deltoid, upper trapezius, and latissimus dorsi). IMU captured kinematic data (100 Hz) from the hand, forearm, upper arm, shoulder, sternum, and pelvis.

Flexion Dataset: The Flexion Dataset collected data from 32 participants (17 males and 15 females) (age: 21–36) performing shoulder and elbow flexion extension tasks [6].[3] Each subject completed eight distinct task conditions combining movement type (static or dynamic), joint (shoulder or elbow), and resistance level (20%–30% or 40%–50% of MVIC). These were StSh 20%–30%, StSh 40%–50%, StEl 20%–30%, StEl 40%–50%, DySh 20%–30%, DySh 40%–50%, DyEl 20%–30%, and DyEl 40%–50%, where "St" and "Dy" refer to static and dynamic tasks, and "Sh" and "El" refer to shoulder and elbow, respectively. EMG data (1000 Hz) were collected from five muscles: anterior deltoid, upper trapezius, biceps brachii, triceps brachii, and latissimus dorsi, and IMU sensors (100 Hz) were placed on the shoulder, upper arm, forearm, hand, and torso.

Subjective fatigue ratings: Both datasets used the Borg RPE scale to track perceived fatigue during exercises, with scores ranging from 6 ("no exertion at all") to 20 ("maximum exertion") [3]. In the Rotation Dataset, participants reported RPE before starting and every 10 s during exercises. In the Flexion Dataset, RPE was recorded every 15 s for 20%–30% MVIC tasks and every 10 s for 40%–50% MVIC tasks.

Data cleaning and processing: For both datasets, IMU and EMG signals were first cleaned to reduce noise and remove baseline offsets. For each exercise set, we subtracted the mean from each accelerometer and gyroscope axis to eliminate dc offsets. Then, a fourth-order low-pass Butterworth filter with a cutoff frequency of 20 Hz was applied to suppress high-frequency noise [7]. For EMG signals, the mean was subtracted, and then, a fourth-order Butterworth bandpass filter with cutoff frequencies at 30 Hz and 350 Hz was applied. The BTS Bioengineering S.p.A., Italy EMG system includes a built-in 50 Hz notch filter to suppress powerline noise. Dynamic repetitions were segmented based on gyroscope peaks; static tasks were windowed into fixed 3-s blocks. RPE scores were interpolated using cubic splines [8] to estimate perceived fatigue for each repetition or window to enable alignment between signal data and subjective labels.

B. Features Extraction

We extracted a comprehensive set of biomechanical features from the raw data from the IMU and EMG sensors. Only sensors common to both datasets were used, resulting in eight sensors: five IMU (hand, forearm, upper arm, shoulder, and torso) and three EMG (anterior deltoid, latissimus dorsi, and trapezius). The features were extracted separately for each segmented repetition and window to align with the corresponding Borg RPE label. To ensure consistency and comparability across datasets, we confirmed that sensor placements were matched in both studies following the same anatomical locations and setup protocols.

For IMU data, features were extracted from each accelerometer and gyroscope axis. Time domain features included movement slope, zero-crossing rates, peak velocities, cumulative displacement, and jerk (rate of change of acceleration) to capture movement smoothness and intensity. Angular metrics, such as maximum and minimum angles, angular displacement ranges, and root mean square of angular velocity, were also derived. Frequency domain features were computed using the fast Fourier transform and Power Spectral Density to capture dominant frequencies, peak energies, and energy distributions across low- and high-frequency bands.

For EMG data, time domain features, such as integrated EMG, root mean square, mean absolute value, and waveform length, were extracted to capture muscle activation intensity and complexity. The features of the frequency domain analyzed fatigue progression through the median frequency, mean frequency, spectral power, spectral centroid, and frequency band ratios using short-time Fourier transforms. As muscles fatigue, the signals they generate change in a way that the speed of signal conduction slows down. This leads to a shift of the EMG signal toward lower frequencies. Because of this, median, mean, and spectral centroid frequencies tend to decrease over time, which makes them reliable indicators of muscle fatigue. Hybrid features, such as zero-crossing rates, slopes, and peak characteristics, were also extracted to track changes in muscle activation patterns over time. We also extracted further statistical features for both IMU and EMG features using different statistics, such as mean, median, minimum, maximum, standard deviation, variance, skewness, and kurtosis.

C. Models Training

We trained machine learning models to predict Borg RPE scores across individual tasks, grouped tasks, and merged datasets. While multiple regressors were tested during early benchmarking (including support vector regression and random forest), extreme gradient boosting (XGBoost) consistently outperformed them in both accuracy and generalization. To keep this letter focused, we report only XGBoost results.

Given the large feature space, we first trained a random forest model for each task setting and selected the top 100 most important features based on feature importance scores. Then, XGBoost regressors were trained on the selected feature subsets.

To ensure generalizability across participants, leave-one-subject-out cross-validation (LOSO-CV) was applied. In each fold, all data from one subject were excluded during training and used for testing, resulting in 66 folds when merging the two datasets. All features were normalized using standard scaling (i.e., subtracting the mean and dividing by the standard deviation) to reduce intersubject variability and ensure consistency across inputs. The hyperparameters of the model were tuned using RandomizedSearchCV. Model performance was evaluated using mean absolute error (MAE), coefficient of determination (R^2), root-mean-squared error (RMSE), and mean absolute percentage error (MAPE).

D. Features Importance Analysis

Feature importance analysis was performed using two complementary approaches: aggregated feature ranking across task groups and SHapley Additive exPlanations (SHAP) value computation for model interpretability.

[1] [Online]. Available: https://github.com/Qirtas/EstimatingFatigue

[2] [Online]. Available: https://zenodo.org/record/8415066

[3] [Online]. Available: https://zenodo.org/records/13773233

First, to identify the features most consistently relevant for fatigue estimation, we computed the aggregated importance of the features across different task groups. Separate models were trained for All Static, All Dynamic, All Internal Rotation, and All External Rotation tasks, as well as the fully merged datasets. To evaluate the overall impact of each feature, we first summed its raw importance scores across each task setting. This approach captured the total contribution of each feature to fatigue estimation, independent of any specific group of movement–resistance tasks. Then, we normalized the aggregated scores to have a direct comparison between features across different task groups.

In parallel, SHAP analysis was conducted on the XGBoost model trained on the fully merged Rotation and Flexion Datasets. SHAP values quantify the contribution of each feature to individual predictions to provide a sample-level interpretation of the behavior of the model.

E. Sensor Optimization Analysis

We performed a sensor optimization analysis on the merged dataset to investigate how a sensor counts influenced model accuracy.

First, feature importance scores were computed using the model trained on the merged Rotation and Flexion Datasets. Each feature was mapped to its corresponding sensor, and the importance of the sensor level was calculated by aggregating the importance scores of all the features associated with that sensor. This produced an overall ranking of the sensors based on their contribution to fatigue prediction across the entire merged dataset.

Based on this ranking, we created eight sensor subsets, starting with the most important sensor and incrementally adding sensors in order of their ranked contribution. For each subset, a new feature set was generated by retaining only the features associated with the selected sensors. Separate regression models were then trained and evaluated in each sensor subset using LOSO CV.

III. RESULTS

A. Prediction Accuracy

Table 1 provides the model performance in individual tasks for the Rotation and Flexion Datasets. In general, ER tasks and StSh tasks achieved the best prediction accuracy with lower MAE and RMSE values compared to dynamic tasks.

Table 1 also presents the performance of the model when the tasks were grouped within each dataset and when the datasets were fully combined. Performance was highest for ER tasks and static tasks. When merging both datasets, the combined model achieved an MAE of 2.37, R^2 of 0.36 and a MAPE of 18.78%.

B. Features Importance Analysis

The aggregated importance heatmap (see Fig. 1) shows the top features across all static, all dynamic, all internal movement, and all external movement tasks. The x-axis represents the task groupings, while the y-axis shows the most influential features extracted from sensors. Each cell in the heatmap shows the normalized importance score of a feature for a specific task group.

Several features appeared in the top-ranked cells, which indicate their broader relevance. Mean Z-axis acceleration of the shoulder and upper arm and deltoid EMG spectral features were among the top contributors across multiple task groupings.

The SHAP beeswarm plot for the fully merged model (see Fig. 2) further confirmed these findings. Deltoid spectral centroid, Z-axis shoulder acceleration mean, and Z-axis forearm acceleration mean emerged as the top three features by global impact. The top feature is the deltoid mean spectral centroid. Higher values of this feature were associated with lower predicted Borg RPE scores.

C. Sensor Optimization Analysis

A sensor importance analysis was conducted by ranking sensors based on average feature importance scores aggregated across all tasks. Eight sensor subsets were built by adding sensors in

Table 1. Model Performance for Individual Tasks Across Rotation and Flexion Datasets

Task	MAE	R^2	RMSE	MAPE
Rotation Dataset				
IR 30%-40%	2.11	0.39	2.64	16.49%
ER 30%-40%	1.66	0.58	2.06	13.24%
IR 40%-50%	1.79	0.48	2.27	13.60%
ER 40%-50%	1.51	0.54	1.92	11.52%
IR 50%-60%	1.88	0.37	2.32	13.90%
ER 50%-60%	1.32	0.46	1.62	12.05%
All IR	1.99	0.43	2.50	14.71%
All ER	1.60	0.57	2.06	12.10%
All Combined	2.02	0.41	2.52	14.93%
Flexion Dataset				
StSh 20%-30%	1.49	0.68	1.93	11.23%
StSh 40%-50%	1.69	0.41	2.35	12.76%
StEl 20%-30%	2.38	0.37	2.94	18.22%
StEl 40%-50%	1.93	0.38	2.46	13.53%
DySh 20%-30%	2.56	0.42	3.16	20.94%
DySh 40%-50%	2.08	0.52	2.60	16.86%
DyEl 20%-30%	2.58	0.30	3.19	21.02%
DyEl 40%-50%	2.33	0.40	2.87	19.05%
All St	2.02	0.44	2.59	15.23%
All Dy	2.52	0.38	3.07	20.60%
All Combined	2.39	0.40	2.96	18.96%
Merged Rotation + Flexion				
All Tasks Combined	2.37	0.36	2.91	18.78%

IR: internal rotation, ER: external rotation, StSh: static shoulder, StEl: static elbow, DySh: dynamic shoulder, and DyEl: dynamic elbow

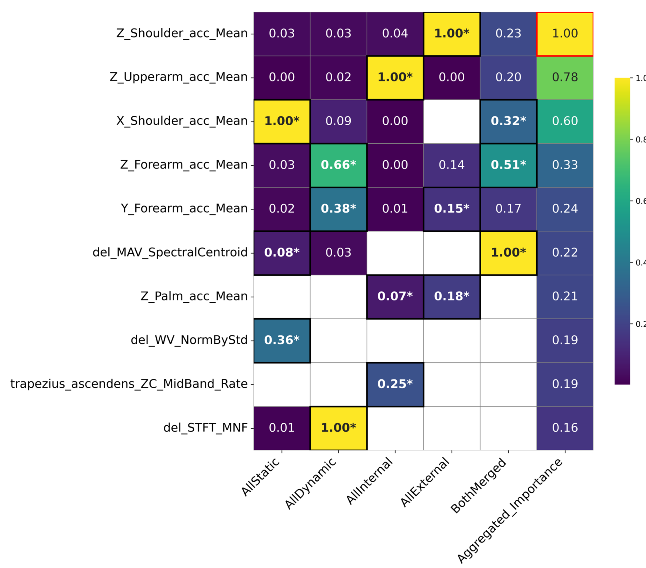


Fig. 1. Aggregated feature importance heatmap across grouped tasks. Lighter colors indicate higher normalized importance. Asterisks (*) mark the top-three features per group. STFT: short-time fourier transform; MAV: mean absolute value; WV: wavelet variability; ZC: zero crossing; MNF: mean frequency.

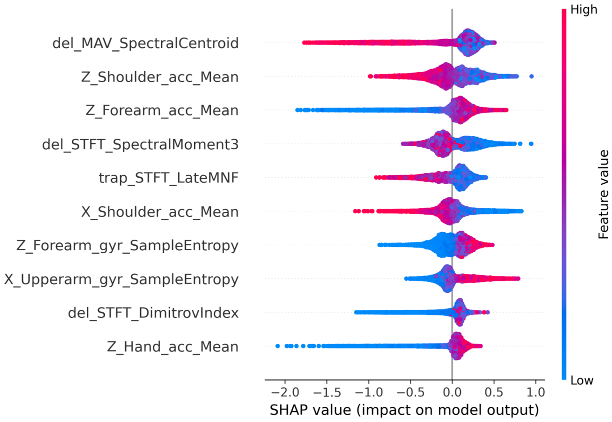


Fig. 2. SHAP plot for the top ten features contributing to fatigue prediction for the combined Rotation and Flexion Dataset model.

Table 2. Model Performance Across Sensor Subsets, Including a Baseline Model That Always Predicts the Mean RPE

Sensor Setup	MAE	R^2	RMSE	MAPE
Baseline (mean predictor)	3.12	–0.01	3.76	25.48%
Subset 1	2.67	0.20	3.26	21.51%
Subset 2	2.53	0.27	3.11	20.18%
Subset 3	2.45	0.31	3.03	19.42%
Subset 4	2.40	0.35	2.95	18.97%
Subset 5	2.35	0.37	2.89	18.60%
Subset 6	2.37	0.36	2.92	18.83%
Subset 7	2.38	0.36	2.92	18.82%
Subset 8	2.37	0.36	2.91	18.78%

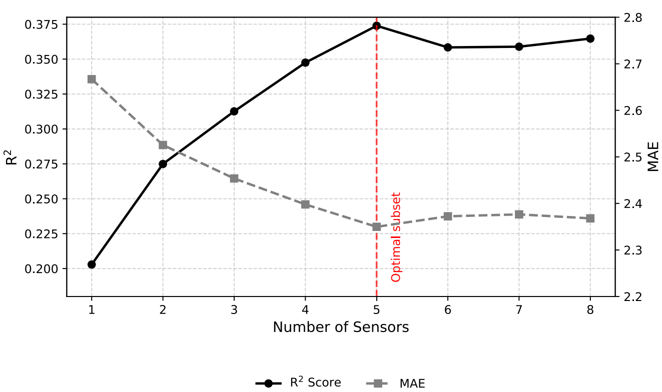


Fig. 3. Effect of sensor subset size on model performance.

descending order of importance (IMU_shoulder, IMU_upperarm, EMG_tricep, EMG_deltoid, IMU_forearm, IMU_hand, IMU_torso, and EMG_bicep). For each subset, a regression model was trained and evaluated using LOSO-CV.

Table 2 shows the performance metrics for each subset. As a reference, a naive baseline model that always predicts the mean Borg RPE has an MAE of 3.12. With just two sensors (shoulder and upper arm IMUs), the MAE drops to 2.53. The five-sensor setup further reduces this to 2.35 and provides the best tradeoff between accuracy and reducing the total sensor count by nearly 40%.

To further illustrate the tradeoff between model complexity and performance, Fig. 3 shows the R^2 score and MAE compared to the number of sensors used. R^2 steadily increased and MAE decreased as we continued to add sensors. Although R^2 values are modest, it is expected due to higher intersubject variability and noise in reporting subjective BORG RPE ratings, but it still consistently outperformed the baseline model with $R^2 < 0$. The optimal subset was identified in five sensors, which provided a good balance between predictive accuracy and practical deployment constraints.

IV. DISCUSSION AND CONCLUSION

This study shows that a five-sensor wearable setup can reliably predict perceived fatigue across 14 heterogeneous upper limb tasks drawn from two independent datasets. While prior work has linked fatigue to reduced movement speed and altered kinematics [9], our study extends this by showing that vertical acceleration features and EMG spectral shifts remain predictive across diverse tasks using a generalizable machine learning framework. The consistently high importance of vertical acceleration features from the upper arm and shoulder shows a key fatigue effect, which is that as muscles tire, people lose the ability to generate or control upward arm motion. Similarly, spectral compression in deltoid EMG signals aligned with existing findings [10], but here, we validate its predictive strength across diverse conditions. Beyond confirming known fatigue patterns, our results contribute by systematically identifying robust biomechanical features and achieving strong prediction accuracy with only five sensors to support practical fatigue monitoring deployments. Despite these strengths, the work has some limitations. Participants were healthy young adults tested in laboratory settings, so generalizability to older or clinical populations needs to be tested in future work.

ACKNOWLEDGMENT

This work was supported in part by the Research Ireland under Grant 12/RC/2289-P2, Grant 13/RC/2077, Grant 21/RC/10303_P2, and Grant 16/RC/3918 and in part by the European Regional Development Fund.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Clinical Research Ethics Committee (CREC) of the Cork Teaching Hospitals at University College Cork under Application No. ECM 4 (p) 6/7/2021, ECM 3 (ww) 09/08/2022, and ECM 4 (p) 6/7/2021, and performed in line with the Declaration of Helsinki.

REFERENCES

[1] R. M. Enoka and J. Duchateau, "Translating fatigue to human performance," *Med. Sci. Sports Exercise*, vol. 48, no. 11, 2016, Art. no. 2228.

[2] A. Garg, K. Hegmann, and J. Kapellusch, "Short-cycle overhead work and shoulder girdle muscle fatigue," *Int. J. Ind. Ergonom.*, vol. 36, no. 6, pp. 581–597, 2006.

[3] G. A. Borg, "Psychophysical bases of perceived exertion.," *Med. Sci. Sports Exercise*, vol. 14, no. 5, pp. 377–381, 1982.

[4] R. Merletti and P. J. Parker, *Electromyography: Physiology, Engineering, and Non-Invasive Applications*. Hoboken, NJ, USA: Wiley, 2004.

[5] M. N. Yasar, M. Sica, B. O'Flynn, S. Tedesco, and M. Menolotto, "A dataset for fatigue estimation during shoulder internal and external rotation movements using wearables," *Sci. Data*, vol. 11, no. 1, 2024, Art. no. 433.

[6] P. O'Sullivan, "Swifties (subject and wearables data for investigation of fatiguing tasks in extension/flexion of shoulder/elbow)," 2024. [Online]. Available: https://zenodo.org/records/13773233

[7] F. Crenna, G. B. Rossi, and M. Berardengo, "Filtering biomechanical signals in movement analysis," *Sensors*, vol. 21, no. 13, 2021, Art. no. 4580.

[8] S. McKinley and M. Levine, "Cubic spline interpolation," *College Redwoods*, vol. 45, no. 1, pp. 1049–1060, 1998.

[9] D. H. Gates and J. B. Dingwell, "The effects of muscle fatigue and movement height on movement stability and variability," *Exp. Brain Res.*, vol. 209, no. 4, pp. 525–536, 2011.

[10] B. Gerdle, B. Larsson, and S. Karlsson, "Criterion validation of surface EMG variables as fatigue indicators using peak torque: A study of repetitive maximum isokinetic knee extensions," *J. Electromyography Kinesiology*, vol. 10, no. 4, pp. 225–232, 2000.