Data in Brief 69 (2026) 113312

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

sEMG dataset of simulated pain-related behaviours for robotic orthopaedic shoulder rehabilitation

Giulio Besi[a,*], Alessandra Fava[a], Haoxiang Wu[b], Sara Garavini[c], Lorenzo Sabattini[a], Federica Ferraguti[a]

[a] Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, via Giovanni Amendola, 2, Reggio Emilia, RE, 42122, Italy
[b] Department of Engineering for Innovation Medicine, University of Verona, Strada le Grazie, 15, Verona, VR, 37134, Italy
[c] Studio Canali: Physiotherapy, Osteopathy and Personal Training, Via Bologna, 6, Reggio Emilia, RE, 42123, Italy

ARTICLE INFO

Article history:
Received 25 July 2026
Revised 28 September 2026
Accepted 30 September 2026
Available online 5 October 2026

Dataset link: Simulated Pain in Robotic Orthopaedic Shoulder Rehabilitation (Original data)

Keywords:
Robotic manipulator
Orthopaedic rehabilitation
Machine learning for rehabilitation
muscular activity analysis

ABSTRACT

This dataset contains surface electromyography (sEMG) recordings from 25 healthy participants aged 21–41 years performing shoulder abduction, shoulder flexion, and scapular retraction ("rower") with a robotic manipulator. Upon oral cues, participants simulated three pain-related behaviours observed during shoulder rehabilitation and selected in consultation with an expert physiotherapist. Signals were recorded from the anterior and rear deltoids and the upper and lower trapezius using a Trigno Centro system and four wireless, single-channel Avanti sensors (Delsys Europe, Greater Manchester, UK).

To the best of the authors' knowledge, frequently used pain-detection datasets contain signals recorded during externally induced pain. However, during real orthopaedic rehabilitation sessions, pain is never voluntarily and externally induced. Pain during movement can alter muscle activation and motor capacity, while measures of muscle activity, such as the nociceptive flexion reflex (NFR), are used in pain assessment and are normally measured using sEMG signals. This motivated the use of sEMG alone. The aim of this dataset is to provide

* Corresponding author.
E-mail address: giulio.besi@unimore.it (G. Besi).
Social media: @lorenzo_s83 (L. Sabattini)

https://doi.org/10.1016/j.dib.2026.113312
2352-3409/© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

a simple yet efficient source for designing machine learning algorithms for pain detection. Moreover, to the best of the authors' knowledge, no currently publicly available dataset contains signals recorded during rehabilitation exercises performed using a robot.

© 2026 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Specifications Table

Subject	Health Sciences, Medical Sciences & Pharmacology
Specific subject area	Human-robot interaction, Physiological signal analysis, Robotic upper-limb rehabilitation, Orthopaedic shoulder rehabilitation
Type of data	Data format: Raw, Filtered, Processed Type of data: EMG signal, Pain Labels, EMG reference values
Data collection	sEMG signals were recorded in a controlled laboratory from healthy participants performing three orthopaedic shoulder rehabilitation exercises. Participants simulated pain-related behaviours when prompted. Signals were acquired using a Trigno Centro system and four wireless, single-channel Avanti EMG sensors (Delsys Europe), positioned over the anterior and rear deltoids and the upper and lower trapezius. A physiotherapy expert defined the exercises, sensor positions, and simulated behaviours. A custom Python application was used to coordinate data collection. Informed consent was obtained, and subjects with musculoskeletal disorders were excluded.
Data Source Location	ARSControl Lab, Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, Piazzale Europa 1, 42,124, Reggio Emilia, RE, Italy
Data accessibility	Repository name: Simulated Pain in Robotic Orthopaedic Shoulder Rehabilitation Data identification number: 10.5281/zenodo.21573324 Direct URL to data: https://zenodo.org/records/21573324 Instructions for accessing these data: visit the URL of the dataset, download the .zip file and extract the content.
Related research article	None

1. Value of the Data

- **First dataset for detecting shoulder pain during movement**: the data support the development of algorithms for detecting pain-related behaviours during orthopaedic shoulder exercises performed with a robotic manipulator. To the best of the authors' knowledge, this is the first publicly available dataset designed for this specific setting.
- **Robotic orthopaedic rehabilitation vs neurorehabilitation**: research on robotic upper-limb rehabilitation has largely focused on neurorehabilitation, such as post-stroke recovery, while orthopaedic rehabilitation has received less attention. The absence of a dataset addressing pain-related behaviours in robotic orthopaedic shoulder rehabilitation motivated this work.
- **Rehabilitation exercise specificity**: recordings were made during shoulder abduction, flexion, and rower exercises while participants interacted with a robotic manipulator. They therefore capture muscle activity during exercises relevant to robot-assisted rehabilitation, rather than relying solely on externally induced pain.
- **Data subjectivity reduction**: sEMG is widely used for biofeedback in rehabilitation [1] and records muscle activity relevant to pain-assessment measures such as the NFR. Focusing on sEMG limits reliance on other physiological measures, including EDA and HRV, which may be consciously modulated [2,3]. The recorded behaviours were nevertheless voluntarily simulated and do not represent true nociceptive responses. Nonetheless, by focusing exclusively on sEMG signals, the dataset aims to reduce subjectivity associated with biosignal selection while acknowledging the voluntary component of the simulated behaviours.

- **Machine learning readiness**: the dataset includes recordings at rest and during three exercises, each performed under two conditions: 10 repetitions without a time limit and 30 s without a repetition limit. These conditions provide different recording contexts for developing and evaluating automated detection algorithms.
- **Easy setup and data scalability**: the use of sEMG signals alone facilitates the replication of the experimental setup by reducing the total number of sensors required. It also supports straightforward data scalability, as additional muscles or signal types can be incorporated into the dataset by simply replicating the data-acquisition protocol.

2. Background

Robots show promise for repetitive, labour-intensive upper-limb rehabilitation tasks [4,5]. Robotic manipulators have attracted interest because their control capabilities offer greater versatility than exoskeletons [6–8]. However, their integration into rehabilitation centres has focused mainly on neurorehabilitation, such as post-stroke rehabilitation, while orthopaedic rehabilitation remains underexplored [9]. Addressing this shortage and integrating pain status into robotic control in this scenario is therefore an important research direction for improving patient safety.

Existing pain-detection datasets have supported the development of machine-learning algorithms [10–14]. However, to the best of the authors' knowledge, none includes pain-related movements or behaviours recorded during robotic orthopaedic rehabilitation. Most datasets rely on artificially induced pain, which does not reflect realistic orthopaedic scenarios [15,16]. Only [12] reports naturally occurring pain during movement, but its sparse and heterogeneous observer labels limit its suitability for training automated detection models; moreover, its recordings were not obtained during robot-assisted rehabilitation. Therefore, the present dataset was created to provide an initial source of simulated pain-related behaviours in robotic orthopaedic upper-limb rehabilitation.

3. Data Description

The dataset contains electromyography (EMG) recordings acquired while participants performed three exercises for orthopaedic shoulder rehabilitation using a robotic manipulator and simulated pain-related behaviours when prompted. In total, 25 healthy participants without reported motor impairments contributed to the dataset. The data for each participant were anonymised so that individual participants cannot be identified.

Each participant completed a single experimental session. Oral instructions regarding the exercise modalities were provided to the participants, while a researcher used a custom Python interface to monitor the muscular activity and label the signals. Each exercise was performed under two different protocols. In the first protocol, participants performed 10 repetitions without any time constraint. In the second, they performed the exercise for 30 s without any constraint on the number of repetitions. Before recording the exercises, maximum voluntary contraction (MVC) procedures were performed for all four monitored muscles.

Each MVC cycle consisted of a 30-s rest period followed by a 5-s maximum isometric contraction. This cycle was repeated three times, resulting in a total recording duration of 105 s. The rest periods, including the first one, were included to reduce the effects of muscle fatigue. The final MVC value for each muscle was calculated as the mean of the root mean square (RMS) values obtained from the three contractions. For each 5-s contraction, the RMS was calculated using a 500-ms moving window with a one-sample step. The moving RMS was defined as:

$$X_{RMS}[i] = \sqrt{\frac{1}{N}\sum_{k=i-N+1}^{i} x^2[k]} \tag{1}$$

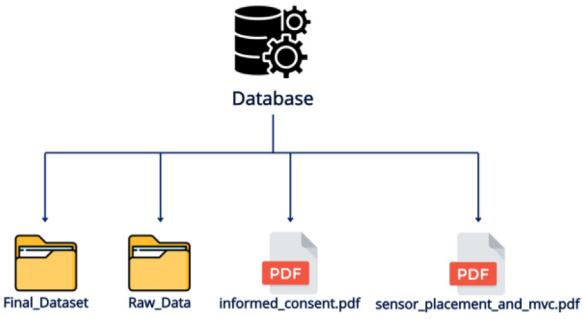


Fig. 1. Main folder structure.

where x denotes the signal within the moving window, i is the window position, and N is the number of samples per window. Eq. (1) was applied separately to each 5-s contraction, and only the maximum RMS value was retained for each contraction, denoted as X_{RMS}^{max}. The final MVC RMS value was then calculated as:

$$MVC_{RMS} = \frac{1}{3}\sum_{j=1}^{3} X_{RMS_j}^{max} \tag{2}$$

A detailed description and illustration of the isometric contractions are provided in the EXPERIMENTAL DESIGN, MATERIALS AND METHODS section.

The dataset is organized into a main directory containing two PDF documents and two subdirectories, called *Raw_Data* and *Final_Dataset* (Fig. 1). The PDF documents are:

- *sensor_placement_and_mvc.pdf*, which describes the precise steps for correctly placing the sensors and performing the MVC procedures – the document's content and description can be found in section EXPERIMENTAL DESIGN, MATERIALS AND METHODS and Fig. 11.
- *informed_consent.pdf*, which contains the template for the informed consent form signed by participants to authorize data collection.

The contents and structure of the *Raw_Data* and *Final_Dataset* folders are described in the following subsections.

3.1. Raw_Data folder

Raw_Data contains one folder for each participant, identified by the prefix *Subject_* followed by a unique identification number (e.g. *Subject_07*). Each participant folder contains 18 CSV files divided into three categories: data (7 files), key (7 files), and MVC (4 files). The hierarchical structure of each folder can be seen in Fig. 2.

Excluding the MVC data, the data and key files are named *Subj_<subject_ID>_<exercise_ID>_data.csv* and *Subj_<subject_ID>_<exercise_ID>_key.csv*, respectively. Here, *exercise_ID* is a number indicating the recording phase stored in the file. The phases are numbered as follows:

1. Rest
2. Abduction – 10 repetitions
3. Abduction – 30 s
4. Flexion – 10 repetitions
5. Flexion – 30 s
6. Rower – 10 repetitions
7. Rower – 30 s

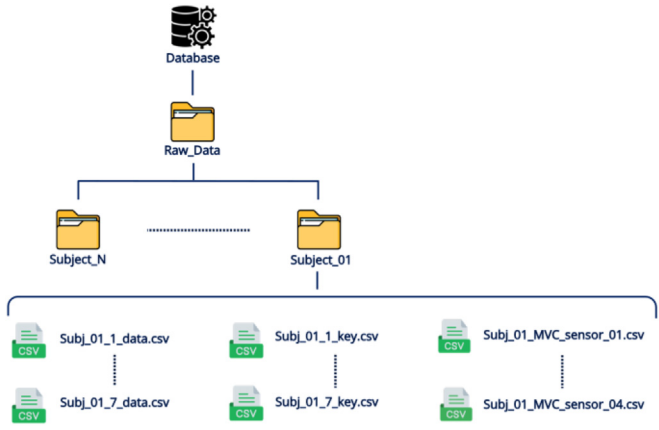


Fig. 2. Raw_Data folder structure.

For example, *Subj_03_2_data.csv* and *Subj_03_2_key.csv* contain the data and key files, respectively, for subject 03 and recording phase 2. Further details on the individual phases are provided in the EXPERIMENTAL DESIGN, MATERIALS AND METHODS section. Each csv file contains:

1. the recording start time (first line);
2. a mapping between sensor IDs (i.e. sensor numbers) and muscles, which depends on the sensor placement (e.g., Sensor 01 → anterior deltoid, Sensor 02 → rear deltoid, Sensor 03 → upper trapezius, Sensor 04 → lower trapezius);
3. the recorded EMG signals, stored in three columns:

   muscle_name, indicating the corresponding muscle;
   time_point, containing the timestamp of each EMG data point;
   value, containing the one-dimensional EMG signal value.

Each key file contains the labels for the corresponding data file, identified by the same suffix, and is used to annotate the beginning and end of the simulated pain-related behaviour. Each key file is organised into two columns:

- *press_time*, indicating the time at which the pain simulation button was pressed;
- *release_time*, indicating the time at which the button was released.

Similarly, MVC files follow the naming convention *Subj_<subject_ID>_MVC_sensor_<sensor_ID>.csv* (e.g. Subj_01_MVC_sensor_01.csv). Each *MVC* file contains:

1. the recording start time (first line);
2. the sensor-to-muscle mapping for the sensor indicated in the filename and used during the recording (e.g., Sensor 01 → anterior deltoid);
3. the recorded MVC signal, stored in three columns with the same structure as the data files (i.e. *muscle_name, time_point*, and *value*).

Specific information on the MVC recording procedure is provided in the EXPERIMENTAL DESIGN, MATERIALS AND METHODS section.

3.2. Final_Dataset folder

As for the *Raw_Data* folder, the subdirectories corresponding to all recorded participants are contained within *Final_Dataset* and are uniquely identified by the subject ID. Each subject directory contains four folders: one containing data recorded during the rest phase and the remaining three containing data recorded during the exercise phases, namely abduction, flexion,

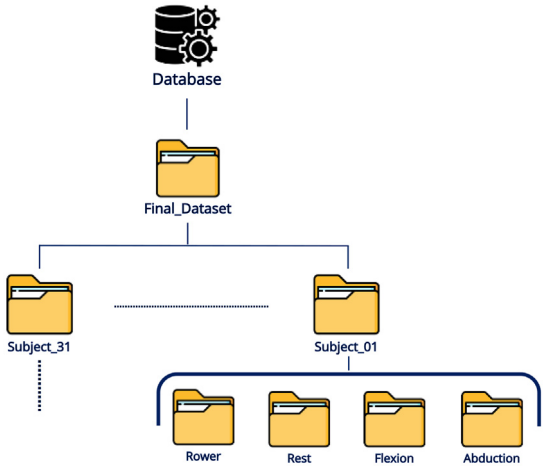


Fig. 3. Final_Dataset folder structure (down to exercises level).

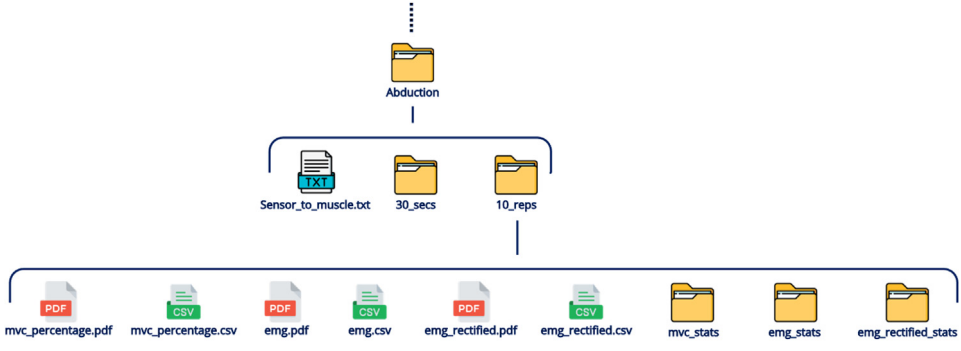


Fig. 4. Example structure of an exercise folder (in this case, abduction).

and rower. These directories are created by processing the data from the corresponding subject folder in *Raw_Data*, as shown in Fig. 3.

As shown in Fig. 4, each exercise folder is further divided into:

- *10_reps*, containing data recorded during the 10-repetition protocol
- *30_secs*, containing data recorded during the 30-second protocol
- *Sensor_to_muscle.txt*, containing the mapping between the Avanti sensor numbers and the muscles on which they were placed.

Both the *10_reps* and *30_secs* folders contain the corresponding participant's data, reorganised and processed into different formats, as shown in Fig. 4. Specifically, each protocol folder contains the following:

- *emg.csv*: a CSV file containing the raw data from *Raw_Data*, reorganised into six columns: one timestamp column (*time_point*), four columns for the recorded muscles (*Anterior_Deltoid, Lower_Trapezius, Rear_Deltoid*, and *Upper_Trapezius*), and one column for the binary pain labels (*Label*). Timestamp, muscle, and label values are expressed in seconds, millivolts, and scalar values, respectively.
- *emg.pdf*: a PDF file showing the temporal evolution of the signals contained in the corresponding CSV file, together with visual information about the assigned labels, represented by red areas in the plot.

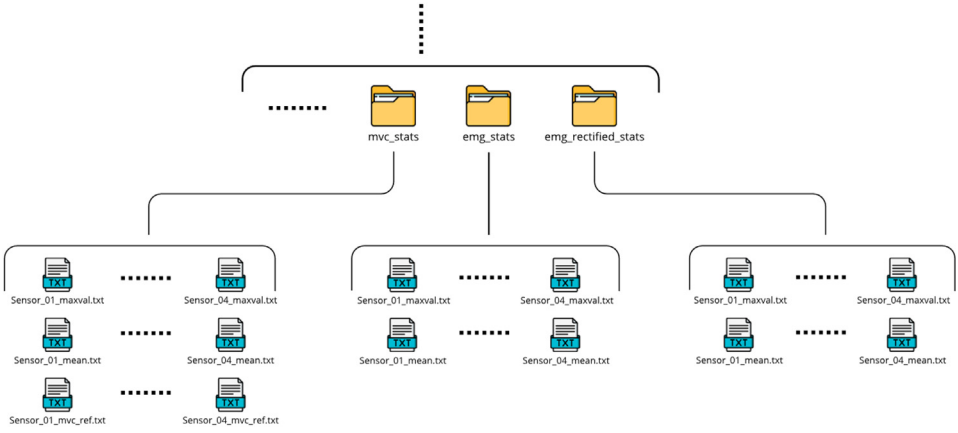


Fig. 5. "_stats" folders structure.

- *emg_rectified.csv*: csv file with same data and format of *emg.csv*, but with rectified EMG values.
- *emg_rectified.pdf*: a PDF file showing the rectified EMG signals and the assigned labels, following the same format as *emg.pdf*.
- *mvc_percentage.csv*: a CSV file with the same columns as *emg.csv*, containing muscle values normalised using the MVC RMS values and expressed as percentages.
- *mvc_percentage.pdf*: a PDF file following the same format as the other PDF files.
- *emg_rectified_stats*: a folder containing text files with statistical values calculated from *emg_rectified.csv*.
- *emg_stats*: a folder containing text files with statistical values calculated from *emg.csv*.
- *mvc_stats*: a folder containing text files with statistical values calculated from the MVC signal files in *Raw_Data*.

As shown in Fig. 5, each *_stats* folder contains two text files for each Avanti sensor, providing information about the corresponding recording phase:

- *Sensor_<sensor_ID>_maxval.txt*, containing the maximum EMG value of the recorded signal in millivolts;
- *Sensor_<sensor_ID>_mean.txt*, containing the mean value of the recorded EMG signal.

In addition to these two files, the *mvc_stats* folder contains a third text file for each sensor, namely *Sensor_<sensor_ID>_mvc_ref.txt*. This file contains the RMS value of the MVC signal, calculated as the mean of the maximum moving RMS values obtained from the three 5-second contraction windows. These values are used to normalise the signals in *mvc_percentage.csv* and are displayed in the legends of the plots contained in *mvc_percentage.pdf*.

The *Rest* folder follows the same directory structure as the *10_reps* folders. In addition, each exercise folder contains a *Sensor_to_muscle.txt* file alongside the *10_reps* and *30_secs* folders. This file specifies the mapping between each Avanti sensor and the muscle on which it was placed.

The data structure has been designed to enable users to navigate the dataset easily, isolate specific recordings, and reproduce the experimental procedure. The supporting documents facilitate replication of the sensor setup and provide the ethical framework for data collection. Moreover, the structure allows additional sensor data to be integrated, thereby supporting straightforward dataset expansion. Fig. 6 and Table 1 provide the mean and standard deviation of the sEMG signals and the label distributions, respectively, for each exercise.

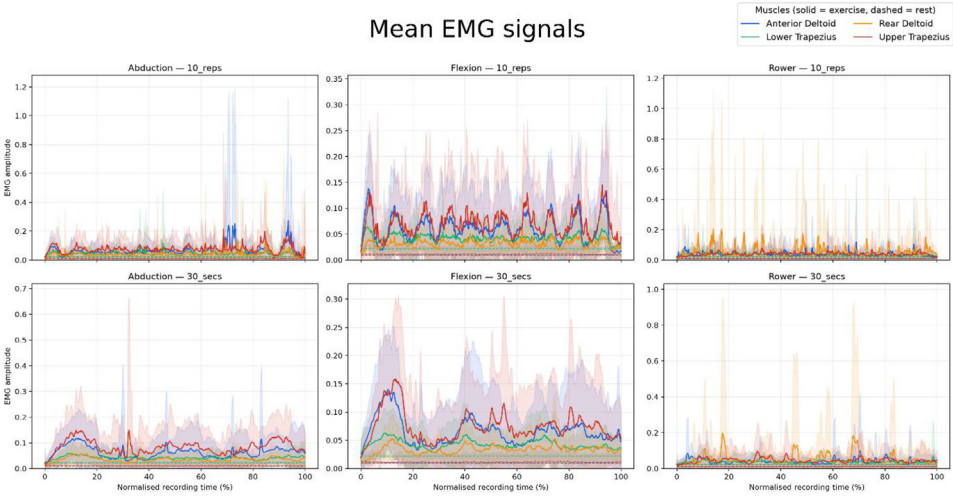


Fig. 6. Mean sEMG signals recorded during the three exercises under both recording protocols (10 repetitions and 30 s). Solid lines represent the mean sEMG amplitude for each muscle across all 25 subjects, with shaded areas of the corresponding colour indicating ±1 standard deviation. Dashed lines represent the mean sEMG amplitude recorded at rest across all 25 subjects. The colours used for anterior deltoid, rear deltoid, lower trapezius and upper trapezius are blue, orange, green and red, respectively.

Table 1
label distribution across exercises.

Exercise	# recordings	Label 0 total time [s] (%)	Label 1 total time [s] (%)
Abduction	50	3744 (91.79)	335 (8.21)
Flexion	50	3409 (91.41)	320 (8.59)
Rower	50	2434 (89.87)	274 (10.13)

4. Experimental Design, Materials and Methods

4.1. Participants

A total of 25 healthy adult volunteers (15 males and 10 females), aged 21–41 years (mean age = 28.0 ±3.67 years), were recruited for this study. No requirement was imposed regarding the participants' dominant hand, although all participants were right-handed. None reported a history of musculoskeletal disorders or presented any upper-limb impairment that could affect task execution. To ensure homogeneous data acquisition, participants were asked not to perform any physical activity on the day of the recording session. This measure was taken to minimise sources of muscle fatigue or reduced arousal that could interfere with muscular activity. Participants with considerable hair around the shoulder or beneath the scapula were advised to arrive with these regions shaved or groomed on the day of the experiment to maximise the recording capability of the EMG sensors. Although this was not mandatory, all participants complied with the request. Data were acquired in a controlled laboratory environment.

Before the recording session began, each participant was informed in detail about the experimental procedure and the type of data to be collected. Written informed consent was obtained from all participants, who also received a copy of the signed document. The study was conducted in accordance with the Declaration of Helsinki, and the experimental procedure was reviewed and approved by the Ethics Committee of the University of Modena and Reggio Emilia (Reggio Emilia, Emilia-Romagna, Italy) under protocol number 321,787/2025.

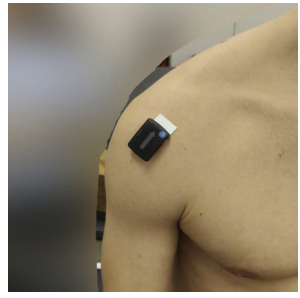


Fig. 7. Sensor placement - anterior deltoid.

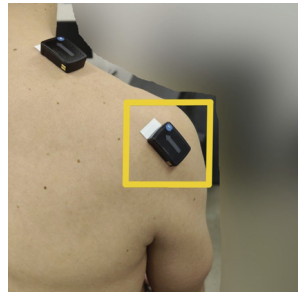


Fig. 8. Sensor placement - rear deltoid.

4.2. Instrumentation and sensor placement

Electromyographic activity was recorded using a Trigno Centro recording system (Delsys Europe) while participants performed shoulder-rehabilitation exercises by moving a UR10e robot (Universal Robots Inc., Michigan, USA). The Trigno Centro system comprises a radio station and four Avanti sensors, which were placed on the participants' skin. Each Avanti sensor is a wireless module containing a single-channel EMG recording unit. The Avanti sensor setup requires users to select preset values for the sensor parameters. For EMG acquisition, the following settings were selected:

- Sampling frequency: 4370 Hz
- Voltage amplitude reading range: ±5.5 mV
- Butterworth band-pass filter: 10–850 Hz

It should be noted that the band-pass filter could not be deactivated because the Trigno Centro system does not provide this option. Therefore, the data described in Section DATA DESCRIPTION were hardware-filtered using the specified band-pass filter.

Each sensor was placed over one of the muscles commonly targeted and monitored during orthopaedic shoulder-rehabilitation exercises: the anterior deltoid, rear deltoid, upper trapezius, and lower trapezius. Before placement, the participants' skin was cleaned to improve signal quality, facilitate sensor placement, and ensure that the sensors remained securely attached during limb movement. The sensors were then placed as shown in Figs. 7–10, all by following the instructions provided in *sensor_placement_and_mvc.pdf* (Fig. 11).

Finally, before data acquisition began, signal quality was assessed visually using a custom Python interface developed specifically for this purpose and built on the official Delsys Python API (version 2.8.0). The application was run on a computer using Windows 11 and was used to:

- configure the sensor parameters



Fig. 9. Sensor placement - upper trapezius.

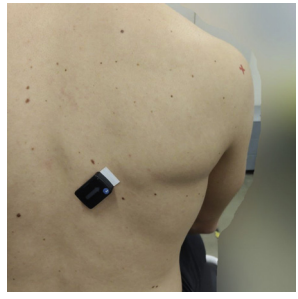


Fig. 10. Sensor placement - lower trapezius.

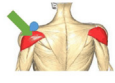
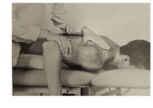
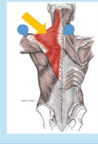
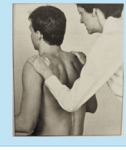
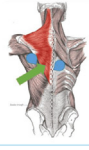
Muscles	Position	Image	MVC	Image
Anterior deltoid	1) Locate the acromion (anterior angle) 2)Position the sensor 3.5 cm below		Shoulder flexion: Subject seated, shoulder bent to 90°. Resistance applied to the humerus, downward.	
Posterior deltoid	1) Locate the acromion (posterior angle) 2)Position the sensor 2 cm below		Shoulder extension: Subject prone, shoulder abducted to 90°, elbow bent to 90°, resistance applied to the humerus downward.	
Upper trapezius	1)Locate C7 (head down) 2) Locate the acromion 3) Find midpoint 4) Place sensor 2 cm aside the midpoint towards the shoulder		Scapular elevation: Subject seated, scapular elevation with the arms close to the body. Resistance applied downward on the shoulders.	
Lower trapezius	1) Locate the inferior angle of the scapula 2) Locate T7 in line with the angle 3)Position the sensor halfway between		Scapular abduction and lowering: Subject prone with the arm abducted to 120°, scapula adducted and lowered, resistance applied downward to the forearm	

Fig. 11. Contents of sensor_placement_and_mvc.pdf. The Position column provides instructions for locating the target muscles and placing the EMG sensors, while the MVC column describes the procedures for recording the MVC of each muscle. The corresponding images illustrate sensor placement and MVC acquisition, respectively.



Fig. 12. MVC recording - rear deltoid.

- navigate through the experimental procedure
- trigger signal acquisition
- label the acquired signals

At the end of the recording session, the custom application was also used to organise and store the acquired data.

4.3. Experimental procedure

After the quality of the EMG signals had been assessed, the setup was complete and the experimental session began. The recording sequence was as follows:

1. Signal at rest
2. Shoulder abduction – 10 repetitions
3. Shoulder abduction – 30 s
4. Shoulder flexion – 10 repetitions
5. Shoulder flexion – 30 s
6. Rower – 10 repetitions
7. Rower – 30 s

During phase 1, the robot was not used, and EMG signals were acquired while participants were seated upright at rest.

Between phases 1 and 2, an MVC signal was recorded for each muscle in the following sequence:

1) Rear deltoid
2) Lower trapezius
3) Anterior deltoid
4) Upper trapezius

The procedures used to elicit maximum force from each muscle are described in *sensor_placement_and_mvc.pdf* (Fig. 11) and were established in consultation with an expert physiotherapist. An example of MVC recording during a single experimental session is shown in Figs. 12–15.

After MVC recording and before phase 2, the path constraining the robot during shoulder abduction and flexion was adjusted using the control method described in [17]. The path was based on each participant's upper-limb length and functional movements. With the elbow extended, the distal end of the limb follows a circular path during shoulder abduction and flexion [18,19]. A circle was therefore used to constrain the robot's motion. To define it, the Cartesian positions of the robot end effector were recorded in the following sequence (Figs. 16–18):



Fig. 13. MVC recording - lower trapezius.



Fig. 14. MVC recording - anterior deltoid.



Fig. 15. MVC recording - upper trapezius.

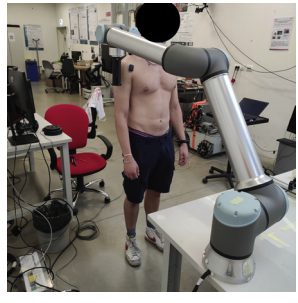


Fig. 16. Circle path setup - step 1.

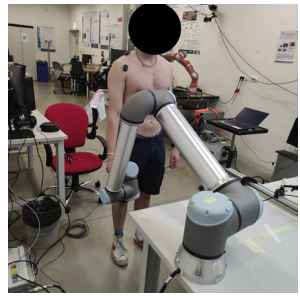


Fig. 17. Circle path setup - step 2.

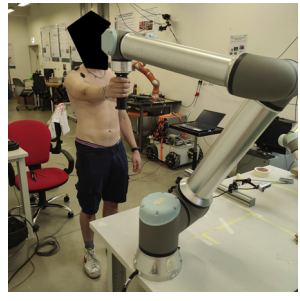


Fig. 18. Circle path setup - step 3.

1. The handle was placed next to the participant's shoulder; this point was defined as the centre of the work circle (**C**).
2. The participant held the handle with the arm and wrist in neutral positions; this point defined the circle radius ***r*** and was used to compute the initial unit vector **a**.
3. The participant held the handle at 90° in front of the body along the sagittal plane; this point was used to obtain a second unit vector, **v1**.

After the 3D coordinates had been recorded, vector **b**, orthogonal to **a** and lying in the same plane, was computed using a double vector product:

$$a \times v1 = v2$$

$$v2 \times a = b$$

Finally, a fixed number of points on circle **y** were obtained by interpolating its parametric equation, defined as:

$$y = C + rcos(\theta)a + rsin(\theta)b$$

The robot controller and reference-path generation were implemented using the ROS framework on a computer running Ubuntu 20.04, separate from the computer running the custom Python application.

Finally, between phases 5 and 6, the reference path used to constrain the robot during the rower exercise was recorded as follows:

1. The participant held the handle with the upper arm and forearm at a 90° angle in the sagittal plane (Fig. 19).
2. The first spline point was recorded with the handle in an "over-reach" position, at the same height above the ground as the starting position. The arm was fully extended forward, and the torso was slightly rotated to increase reach (Fig. 20).

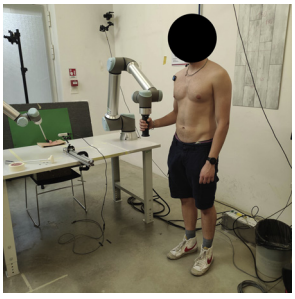


Fig. 19. Row path setup - step 1.

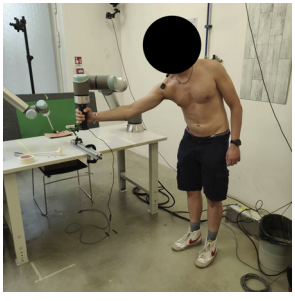


Fig. 20. Row path setup - step 2.



Fig. 21. Row path setup - step 3.

3. The second and final spline point was recorded with the handle in an exaggerated scapular-retraction position at the same height above the ground as the starting position (Fig. 21).

This procedure ensured that the rower movement did not reach the endpoints of the curve recorded in steps 2 and 3 above. Figs. 22–25 show examples of one participant during the exercise phases.

4.4. PAIN behaviours simulation AND labelling process

After the work-circle path had been tuned and before phase 2 began, participants were instructed to reproduce three behaviours simulating pain occurrences. These behaviours were selected based on [12] and in consultation with an expert physiotherapist:

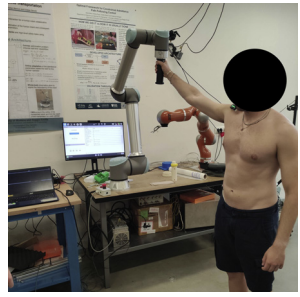


Fig. 22. Shoulder abduction exercise.

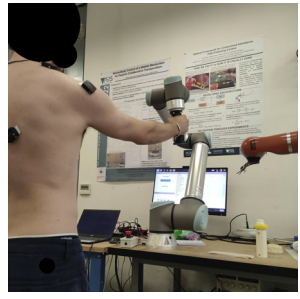


Fig. 23. Shoulder flexion exercise.

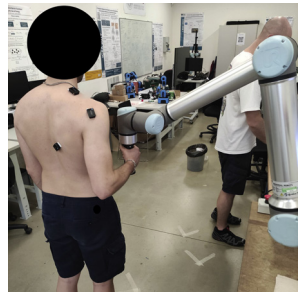


Fig. 24. Row exercise (elbow flexion).

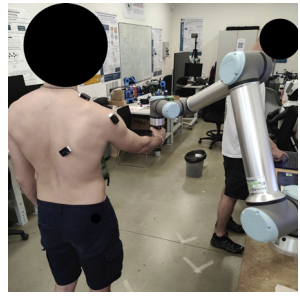


Fig. 25. Row exercise (elbow extension),.

- **Upper trapezius compensation**: incorrect activation of the upper trapezius on the active-limb side by raising the shoulder towards the ear. This muscle is generally monitored to assess compensatory activity because a healthy participant would not normally activate it in this compensatory manner during these exercises.
- **Limb sudden stiffness**: the active limb was suddenly stiffened during movement; in the EMG signal, this behaviour resembles a sudden pain occurrence.
- **Hesitation**: interruption of limb movement, followed by a slow resumption until the normal movement velocity was restored. This behaviour represents a situation in which a patient anticipates approaching a painful configuration; anticipation of pain is generally assessed in such instances.

Participants simulated these behaviours while performing the three orthopaedic rehabilitation exercises described above: shoulder abduction, shoulder flexion, and rower. All movements were performed using the robot. During each exercise, participants waited for an oral cue from the researcher. Upon receiving the cue, they simulated one of the pain-related behaviours. No constraint was placed on which behaviour to reproduce; participants selected one freely each time the cue was given. Simultaneously, the researcher used the custom Python application to label the signal in real time.

It should be noted that these labels do not correspond exactly to the start and end times of the simulated behaviour. This discrepancy resulted directly from the labelling procedure: the positive (i.e. pain-related) label began when the researcher pressed and held a button while giving the participant an oral cue, and ended when the button was released. Consequently, the entire simulation period was labelled as positive, although some uncertainty remained at the beginning and end of the labelled signal interval.

At the end of the recording session, the acquired data were stored in the participant's folder within *Raw_Data*. To generate the data in the corresponding participant folder within *Final_Dataset*, the raw data were automatically processed using the custom Python application.

Limitations

The dataset was collected from a homogeneous population of Italian adults aged 21–41 years; therefore, the recordings may not represent the general population, particularly younger and older individuals. Moreover, data were acquired from healthy participants who simulated pain-related movements; consequently, the dataset may lack the consistency and reliability required for clinical research. The experimental design included only three specific movements, which do not cover all shoulder-rehabilitation exercises and may limit generalisability to other shoulder movements. In addition, the specific simulated pain-related behaviour was not labelled, resulting in a purely binary labelling strategy. Muscle fatigue may also be more prevalent in the rower EMG signals because this was the final exercise performed. Finally, the data were recorded under controlled laboratory conditions and therefore do not capture the variability present in real clinical environments.

Ethics Statement

The study was conducted in accordance with the principles of the Declaration of Helsinki. All participants were fully informed about the experimental procedure and provided written informed consent before participation. The experimental procedure was reviewed and approved by the Research Ethics Committee of the University of Modena and Reggio Emilia (Reggio Emilia, Emilia-Romagna, Italy) under protocol number 321,787/2025. A copy of the informed-consent form is provided with the submission.

Credit Author Statement

Giulio Besi: Conceptualization, Methodology, Validation, Investigation, Supervision, Resources, Data curation, Formal analysis, Software, Writing – Original Draft, Writing – Review & Editing; **Alessandra Fava:** Resources, Data curation, Writing – Original Draft; **Haoxiang Wu:** Software, Resources; **Sara Garavini:** Methodology, Resources; **Lorenzo Sabattini:** Supervision, Writing – Review & Editing; **Federica Ferraguti:** Supervision, Project administration, Writing – Review & Editing.

Declaration of Generative AI and AI-Assisted Technologies in the Manuscript Preparation Process

During the preparation of this work, the author(s) used ChatGPT (OpenAI) for shortening the text length and improving readability. The author(s) reviewed and edited the output as needed and take full responsibility for the content of the published article.

Data Availability

Simulated Pain in Robotic Orthopaedic Shoulder Rehabilitation (Original data) (Zenodo).

Acknowledgements

This work was supported by the Italian Ministry of University and Research, under the complementary actions to the NRRP “Fit4MedRob - Fit for Medical Robotics” Grant (# PNC0000007).
This work was also partially supported by De Luca Foundation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] O.M. Giggins, U.M. Persson, B. Caulfield, Biofeedback in rehabilitation, J. Neuro. Eng. Rehabil. 10 (2013) 60, doi:10.1186/1743-0003-10-60.
[2] Y. Ikeda, H. Hirai, Voluntary control of electrodermal activity in relation to imagery and internal perception scores, Psychophysiology 13 (4) (1976) 330–333 PMID: 951474, doi:10.1111/j.1469-8986.1976.tb03085.x.
[3] S. Laborde, M.S. Allen, U. Borges, F. Dosseville, T.J. Hosang, M. Iskra, E. Mosley, C. Salvotti, L. Spolverato, N. Zammit, F. Javelle, Effects of voluntary slow breathing on heart rate and heart rate variability: a systematic review and a meta-analysis, Neurosci. Biobehav. Rev. 138 (2022) 104711 Epub 2022 May 24. PMID: 35623448, doi:10.1016/j.neubiorev.2022.104711.
[4] F. Zanatta, A. Giardini, A. Pierobon, M. D'Addario, P. Steca, A systematic review on the usability of robotic and virtual reality devices in neuromotor rehabilitation: patients' and healthcare professionals' perspective, BMC Health v Res. 22 (2022) 523, doi:10.1186/s12913-022-07821-w.
[5] V. Santoriello, G. Cesarelli, F. Ficuciello, C. Giugliano, F. Angelone, M. Russo, A.M. Ponsiglione, M. Romano, Assessing the relevance of biosignal-controlled robotic rehabilitation technologies: a systematic review, Prog. Biomed. Eng. 8 (2026) 032005, doi:10.1088/2516-1091/ae7862.
[6] Y. Yang, H. Hong Teo, Y. Jin King, Toward intelligent Human–Robot interaction for upper limb rehabilitation: a review of emerging modalities and strategies, IEEe Access. 13 (2025) 185513–185532, doi:10.1109/ACCESS.2025.3625220.
[7] L. Pezeshki, H. Sadeghian, M. Keshmiri, X. Chen, S. Haddadin, Cooperative assist-as-needed control for robotic rehabilitation: a two-player game approach, IEEE Robot. Autom. Lett. 8 (5) (2023) 2852–2859, doi:10.1109/LRA.2023.3261750.

[8] L. Zhang, S. Guo, Q. Sun, Development and assist-As-needed control of an end-effector upper limb rehabilitation robot, Appl. Sci. 10 (19) (2020) 6684, doi:10.3390/app10196684.
[9] K. Moulaei, K. Bahaadinbeigy, A.A. Haghdoostd, M.S. Nezhad, A. Sheikhtaheri, Overview of the role of robots in upper limb disabilities rehabilitation: a scoping review, Arch. Public Health. 81 (1) (2023) 84 PMID: 37158979; PMCID: PMC10169358, doi:10.1186/s13690-023-01100-8.
[10] P. Gouverneur, A. Badura, F. Li, M. Bieńkowska, L. Luebke, W.M. Adamczyk, T.M. Szikszay, A. Myśliwiec, K. Luedtke, M. Grzegorzek, E. Piętka, An experimental and clinical physiological signal dataset for automated pain recognition, Sci. Data 11 (2024) 1051, doi:10.1038/s41597-024-03878-w.
[11] S. Walter, S. Gruss, H. Ehleiter, J. Tan, H.C. Traue, P. Werner, A. Al-Hamadi, S. Crawcour, A.O. Andrade, G.M. da Silva, The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system, in: 2013 IEEE International Conference on Cybernetics (CYBCO), Lausanne, Switzerland, 2013, pp. 128–131, doi:10.1109/CYBConf.2013.6617456.
[12] M.S.H. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, M. Valstar, H. Meng, A. Kemp, M. Shafizadeh, A.C. Elkins, N. Kanakam, A. de Rothschild, N. Tyler, P.J. Watson, A.C.C. de Williams, M. Pantic, N. Bianchi-Berthouze, The automatic detection of chronic pain-related expression: requirements, challenges and the multimodal EmoPain dataset, IEEE Trans. Affect. Comput. 7 (4) (2016) 435–451, doi:10.1109/TAFFC.2015.2462830.
[13] R. Fang, E. Hosseini, R. Zhang, C. Fang, S. Rafatirad, H. Homayoun, Survey on pain detection using machine learning models: narrative review, JMIR AI 4 (2025) e53026, doi:10.2196/53026.
[14] C. Camacho-Navas, L. Li, K. Poply, V. Mehta, P. Kyriacou, Pain assessment using physiological responses/markers in different types of pain: a scoping review, npj Digit. Med 9 (2026) 64, doi:10.1038/s41746-025-02241-6.
[15] S.L. Merkle, K.A. Sluka, L.A. Frey-Law, The interaction between pain and movement, J. Hand. Ther. 33 (1) (2020) 60–66 Epub 2018 Jul 17. PMID: 30025839; PMCID: PMC6335190, doi:10.1016/j.jht.2018.05.001.
[16] V. Skljarevski, N.M. Ramadan, The nociceptive flexion reflex in humans – review article, Pain 96 (1-2) (2002) 3–8 PMID: 11932055, doi:10.1016/s0304-3959(02)00018-0.
[17] G. Besi, A. Pupa, C. Secchi, F. Ferraguti, Optimal framework for constrained admittance path-following control, in: 2025 IEEE International Conference on Robotics and Automation (ICRA), Atlanta, GA, USA, 2025, pp. 16391–16397, doi:10.1109/ICRA55743.2025.11127361.
[18] R. Lugo, P. Kung, C.B. Ma, Shoulder biomechanics, Eur. J. Radiol. 68 (1) (2008) 16–24 Epub 2008 Jun 3, PMID: 18511227, doi:10.1016/j.ejrad.2008.02.051.
[19] J. McMullen, T.L. Uhl, A kinetic chain approach for shoulder rehabilitation, J. Athl. Train. 35 (3) (2000 Jul) 329–337 PMID: 16558646; PMCID: PMC1323395.