Estimating Optimal Training Repetitions Using EMG-Based Muscle Fatigue Detection

Team: BioTeam7 **Course:** Biomedical Signal Processing — 2025/26/1

Team members: Dario Ranieri, Islam Muhammad Muqsit, Zsuzsanna Rohán

Abstract

This project focuses on the independent measurement and analysis of surface electromyography (EMG) signals to detect muscle fatigue in human arms during resistance training. By analyzing EMG signal characteristics, we aim to estimate the optimal number of repetitions per set for each individual, providing a personalized approach to training and rehabilitation.

1. Introduction

1.1. Background and Motivation

Since with this EMG-associated method we can quantify measures related to muscle fatigue, we hope to identify reproducible fatigue thresholds which can be used in personalized training plans, in rehabilitation techniques, such as FES (functional electrical stimulation) therapy: so based on analysis of recorded EMG signals we can determine the optimal number of stimulations based on the identified threshold. Moreover it also can be used to develop wearable monitoring devices for athletes.

The optimal number of repetitions can be used to gain a better picture of the patient's or athlete's performance, therefore helping in injury prevention and performance maximisation.

1.2. Problem Statement

Identifying optimal repetition number based on classification of recorded sets of repeated bicep curl time series data on different fatigue levels.

1.3. Research Objectives and Questions

Can a quantifiable, reproducible threshold in the RMS and/or MDF feature trends be identified that reliably signals the onset of significant muscle fatigue?

Null Hypothesis (H0): There is no significant change in the Median Frequency (MDF) or Root Mean Square (RMS) of the EMG signal as contraction time increases during a set of repetitions.

Alternative Hypothesis (H1): As muscle fatigue progresses, the Median Frequency (MDF) will significantly decrease due to slowing muscle fiber conduction velocity, while the RMS amplitude will increase due to motor unit recruitment to maintain force.

2. Theoretical background

2.1. Electromyography (EMG) Fundamentals

Electromiography (EMG) is a neurophysiological examination method that allows us to assess nerve and muscle function by recording and analyzing the bioelectrical activity of skeletal muscles. It is used in various areas such as in medical research to help detect neuromuscular abnormalities, in rehabilitation, ergonomics design and sport science.

2.1.1. Recording types

We can use surface electrodes to record EMG signals from muscles close to the surface. not covered by bones or other tissues, but to reach deeper muscles, invasive needle electrodes need to be inserted through the skin into the muscle.

2.1.2. Generation and characteristics of EMG signal

The functional unit of the muscle is the motor unit. It is a single spinal cord motoneuron (in the ventral horn) and all its innervated muscle fibers. EMG is able to measure the MUAP (Motor Unit Action Potential), which is the electrical signature of a single motor unit, formed by the synchronized firing of its fibers, detectable by electrodes.

As the muscle is contracted more forcefully, gradually more and more muscle fibers are activated (recruitment) in a specific order (smaller, low-threshold units first, then larger ones). In case of contraction it is difficult to measure individual action potentials, since each active motor unit produces a series of MUAPs, which, when superimposed, create the overall EMG signal, so we actually detect the interference pattern, which is the result of asynchronously activated motor units, whose potentials are overlapping. So the resulting EMG signal's shape, amplitude, and frequency are influenced by the number of active units, their firing patterns, fiber size, location, and tissue properties.

2.1.3. Resting EMG

At rest, EMG activity cannot be measured, since muscles do not normally produce electrical signals during rest. By measuring EMG activity during rest can be associated with neuromuscular diseases or can originate from measurement errors.

2.2. EMG Signal Characteristics During Exercise

Based on literature research [1], [2], we have chosen two types of measure to identify muscle fatigue:

2.2.1. Root Mean Square (RMS):

For an sEMG signal consisting of a sequence of n discrete samples, denoted as $x_1, x_2, \ldots, x_n$, the RMS value is calculated using the following formula:

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2}$$

Where
- x_i is the value of the sEMG signal at sample i
- n is the total number of samples within the analyzed time window.

This value represents the amplitude of the EMG signal, indicating the muscle's power or force, and is calculated over a specific time window. There is a strong positive correlation between muscle force and RMS value during isometric contractions, meaning larger forces correspond to larger EMG signals and higher RMS values (proportional, often linear, relationship).
Fatigue detection: It shows a significant increase shortly after the initiation of movement, peaking at exhaustion, then it will drop, so it is an ideal measure to use to determine the onset of the fatigue. During sustained contraction, the RMS amplitude typically increases over time as the muscle fatigues due to the nervous system recruits more motor units to maintain force (see: Fig.1.).

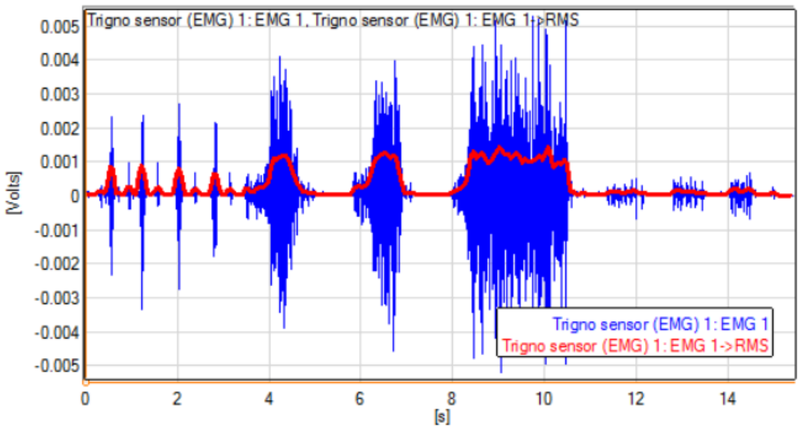


Fig.1. Illustrates the increasing RMS during muscle fatigue. (Source: https://delsys.com/amplitude-analysis-root-mean-square-emg-envelope/)

2.2.2. Median Frequency (MDF)

The Median Frequency indicates the frequency value that divides the spectrum into two equal parts.
Fatigue detection: A greater decrease in median frequency is observed at higher force levels of sustained contraction (see: Fig.2. and Fig.3.).
This gradual decrease is observed due to changes in the muscle's signal frequency, which shifts to lower frequencies due to slowing muscle fiber conduction velocity caused by metabolic changes during sustained contractions.

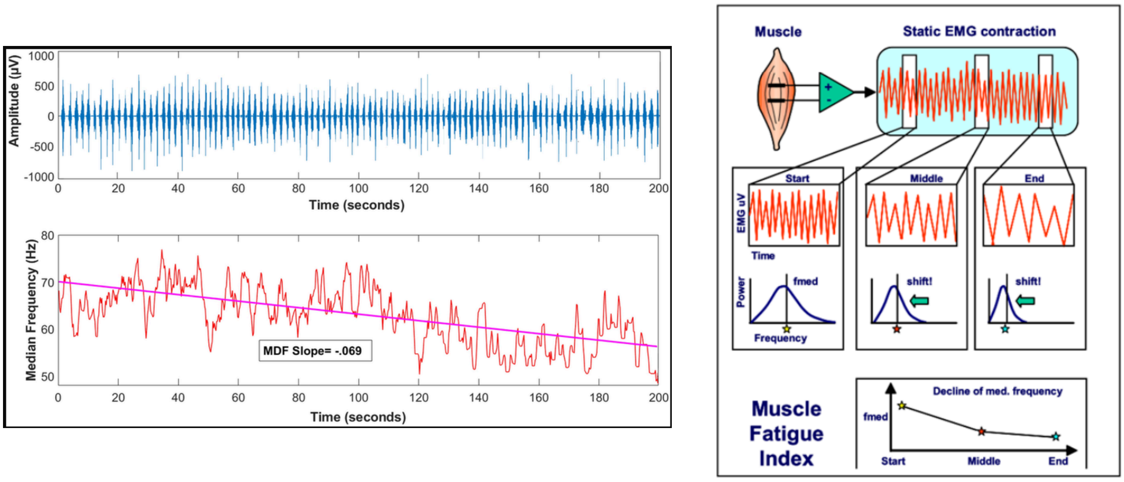


Fig.2. Relationship Between EMG and Muscle Fatigue: Illustrates the decreasing MDF trend during muscle fatigue. (Source: Aghaie Ataabadi, Peyman & Abbasi, Ali & Shojaatian, Mohsen & Letafatkar, Amir & Svoboda, Zdenek & Rossettini, Giacomo. (2022). The effects of facilitatory and inhibitory kinesiotaping of Vastus Medialis on the activation and fatigue of superficial quadriceps muscles. Scientific Reports. 12. 10.1038/s41598-022-17849-x.) (Source: The ABC of EMG A Practical Introduction to Kinesiological Electromyography Peter Konrad Version 1.0 April 2005 Current Training Methodologies)

By using the intersection or thresholds of these different feature trends (see: Fig.4.) we can suggest an optimal number of repetitions.

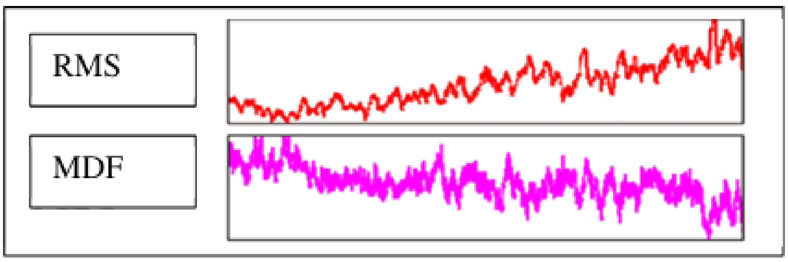


Fig.4. The oppositely changing trends of RMS and MDF as the muscle fatigues. (Source: Yousif, Hayder A., et al. "Assessment of muscles fatigue based on surface EMG signals using machine learning and statistical approaches: A review." *IOP conference series: materials science and engineering*. Vol. 705. No. 1. IOP Publishing, 2019.)

2.2.3. Connection with fitness level

This [3] study investigated muscle fatigue characteristics during isometric contraction using surface electromyography (sEMG) in athletes and non-athletes. They experienced that athletes showed a significantly slower decline in median frequency (MF) and the decrease in root mean square (RMS) amplitude was also lower compared to non-athletes, indicating superior fatigue resistance, meaning having a higher capacity to maintain performance levels for longer periods compared to an average person, and having greater ability to resist the onset of fatigue. Athletes also have better neuromuscular efficiency, meaning the ability of the nervous system to effectively and synchronously recruit the appropriate muscles to produce force and execute coordinated movements with minimal energy waste

3. Methodology

3.1. Experimental Equipment and Setup

- **EMG recording**: For the recording of surface EMG signals we used a Cometa MiniWave wireless EMG device connected to two 3M Red Dot ECG surface electrodes provided by the Data Analytics and Sports Laboratory here at the Faculty. The EMG recording device has a 2000 Hz sampling frequency and 40 meters indoor range. The recorded signals were saved as a c3d file format with the help of the Cometa device software. During the measurement, we had visual feedback from the recorded signal.
- **Resistance training**: using dumbbells of varying sizes between 2-9.5 kg.
- **Subjective Feedback Collection**
    - **Record Perceived Soreness and Log Feedback:** Immediately after each set, we asked the subject to provide feedback on "what the user might think was the point he/she felt sore" or first noticed significant fatigue. We recorded this subjective feedback and associated it with the corresponding set number. This is used to "validate the method against subjective exertion", so it is set as the ground truth for the classification algorithm (see: Fig.7.).

3.2. Participants

- **Number of Users:** A minimum of 5-10 subjects is recommended to ensure variability in the data and validate the method across different individuals. We were able to record data of 12 people. The volunteer test subjects were healthy university students, aged between 22-24 years, with varying previous exercise experience.
- **Desired Fitness Level:** Participants should ideally have experience with resistance training to accurately determine their 1-Repetition Maximum (1RM). The protocol has been standardized to each individual's fitness level by using a load corresponding to **60-70% of their 1RM**, as this load range is suitable for hypertrophy training.
  In the case that the participant was unaware of his/her RM, we can consider the **load** for a **4x8** exercise on the muscle.
  Before involving the participants, a questionnaire was proposed to assess the fitness level of each user.

3.3. Exercise Protocol

0. Calibration

Some inexperienced users went through a calibration phase, where we assessed their optimal weight shortly before the measurement.

Fig.5. Planned execution position and warmup procedure.

Subject Preparation

1. **Warm-up:** to avoid injuries, participants planned to warm up with an elastic band for a couple of minutes (see: Fig,5,). Due to inability to access resistance bands throughout the measurements a 1kg dumbbell was used for warm-up.
2. **Techniques:** First, we needed to establish each subject's ability to perform the desired exercise. This was assessed with a light/moderate load, so that the optimal isolation of the muscle can be checked. Our team includes two people with good expertise in these kinds of workouts who were able to help participants to maintain a proper form during the exercise.
3. **Electrode Placement:** We placed two surface EMG electrodes on the target muscle (see: Fig.6.). We tried to ensure proper skin preparation by cleaning the area under the electrode with alcohol. The optimal distance between the electrodes was approximately 2-3 centimeters (based on testing and recommendation from the expert in the lab).
4. **Signal Check:** We connected the electrodes to the biopotential amplifiers and checked the signal quality based on visual feedback: in case of optimal measurement setups we had a flat line in the resting period.
5. **Execution**: users started with a weight they thought could be lifted with a moderate effort, then for the next series, the weight was adjusted in order to find one that made the user exhausted after 10/12 repetitions.

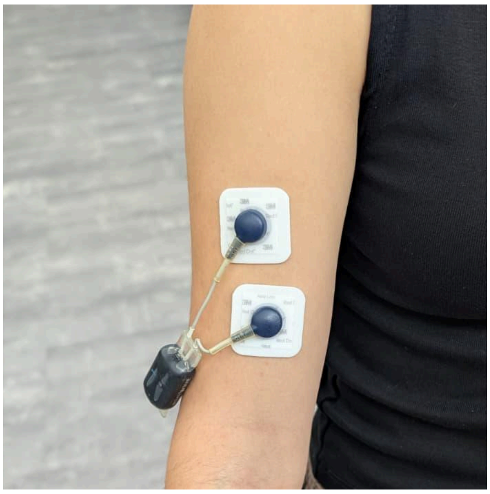


Fig. 6. The correct electrode placement on the biceps brachii

3.4. Data Acquisition

Data Acquisition Procedure

1. **Perform Exercise Sets:** The subject performed multiple sets of the exercise (e.g., biceps curls) using their calculated 60-70% 1RM load.
2. **Repetitions:** Each set was performed until **task failure** (the point where good form can no longer be maintained). This is crucial for capturing the full fatigue trend and validating against "total repetitions to failure".
3. **Record EMG:** We continuously recorded the raw EMG signal during the entire set.
4. **Log Set Number:** We associated each measurement with the "series number" (set number).
5. **Rest Period:** The subject needed to rest for **1 minutes and 30 seconds** between each set.
6. **Repeat Measurements:** The users repeated the procedure for a total of 3-5 sets to gather sufficient data.

Measurement ID	Name	Number of repetitions	Subjective max	Used dumbbell (kg)
1	Dario	12	11	9.5
2	Dario	12	10	9.5
3	Dario	12	9	9.5
4	Dario	12	8	9.5

Fig.7. Subjective results acquired during measurements.

3.5. Data Processing and Analysis

3.5.1. Filtering: We processed the Raw EMG with a Bandpass Filter (20–450 Hz) and a Notch filter (50 Hz) to remove powerline interference (see: Fig.8.).

1. We used a Butterworth filter (bandpass_filter) to remove unwanted low-frequency noise (e.g., movement artifacts) and high-frequency noise outside the typical EMG range (default 20 Hz to 450 Hz).
2. We used the filtfilt() function from scipy.signal, which is a zero-phase digital filter, meaning it applies the filter forwards and then backwards across the data.
3. For the notch filter implementation we used a second-order IIR band-stop filter with a narrow bandwidth (iirnotch() from scipy.signal).

3.5.2. Envelope: Rectify the signal and create a Low-pass Envelope (Smoothing).

1. We take the absolute value (full-wave rectification) and then smooth it with a low-pass filter to create an "envelope" that represents the intensity or amplitude of the muscle activity over time, rather than the raw oscillating signal.

3.5.3. Segmentation: Segment the processed signal to isolate each repetition (2 seconds).

1. Analyzes the signal envelope to automatically identify individual "repetitions" by finding peaks in the activity level (see: Fig.9.).
2. We ensure that detected peaks are separated by at least a certain time interval (e.g., 0.5 seconds), preventing multiple detections for a single, sustained contraction.
3. For each detected peak, it defines a time window around that peak (default of 0.6 seconds before and after the peak) to isolate the data segment corresponding to that specific repetition.

3.5.4. Feature Extraction: Calculate the fatigue-related features (e.g., **RMS** and **MDF**) for each segmented repetition (see: Fig.10. and Fig.11.).

1. **Time-Domain Analysis (i.e.: Amplitude Changes)**
    1.1. Implementing the mathematical definition of RMS: calculating the square of every segment, finding the mean, then taking the square root of it,
2. **Frequency-Domain Analysis (i.e.: Median Frequency Shift)**
    2.1. We use Welch's (periodogram) method (scipy.signal.welch) to estimate the Power Spectral Density (PSD), showing which frequencies contribute most to the signal's power.
    2.2. We calculate the cumulative sum of the PSD, which shows what percentage of the total power is accumulated below a given frequency.
    2.3. Then we find the Median Frequency by identifying the point where the cumulative sum reaches half of the total performance.

3. **Trend Analysis and Optimal Repetition Estimation:** Analyze the feature trends across the repetitions (expecting RMS to increase and MDF to decrease) to identify the onset of fatigue.

3.1. Scoring system

3.1.1. We normalize both RMS and MDF values between 0 and 1

3.1.2. We create a composite score. We use the assumption that as a person exercises, RMS tends to stay high or increase, while MDF (fatigue indicator) tends to decrease. We calculate the score as rms_norm - reversed(mdf_norm). A higher score indicates a repetition where the muscle is working hard despite high fatigue.

3.2. Fatigue Threshold Identification

3.2.1. We define a threshold as 40% of the maximum score achieved.

3.2.2. We are looking for the earliest repetition index that meets this threshold, ensuring the detection happens after a minimum lookback period (default 2 reps) to avoid catching early stabilization points.

3.2.3. If no rep meets the threshold, the code simply returns the repetition with the single highest score.

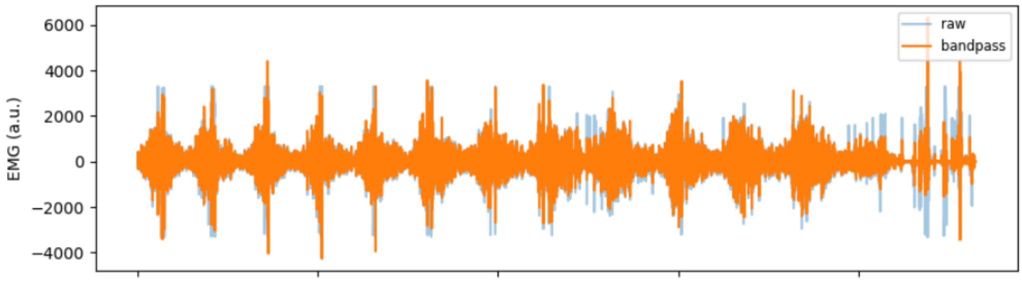


Fig.8. Example of Bandpass filtering

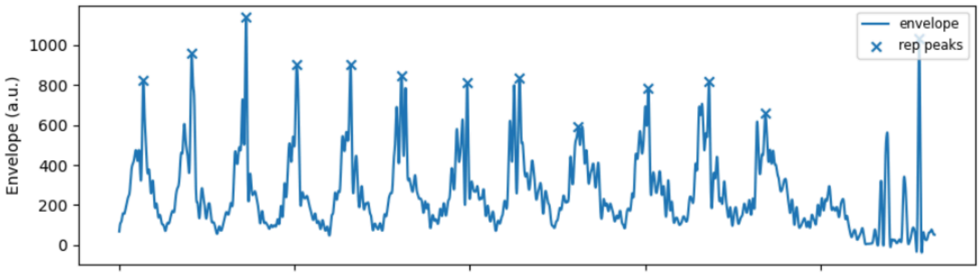


Fig.9. Peak detection on a 2-second interval.

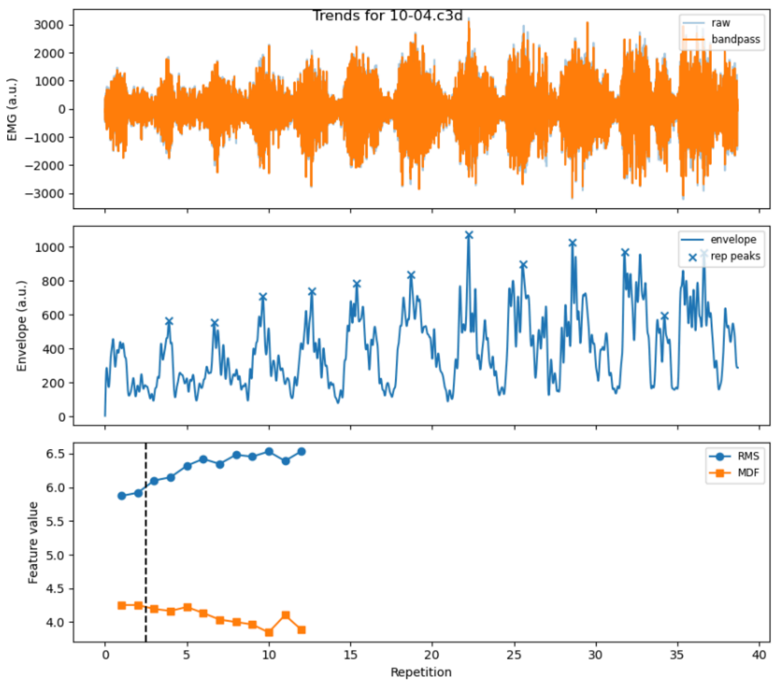


Fig.10. An example of the output of the pipeline, including the result of filtering (1), peak detection (2) and feature extraction (3).

rep	start	end	peak_idx	peak_time	rms	mdf	env_peak	rep_duration	is_fatigued
1	0	7940	4794	2.397	1665.8860429980355	78.125	2059.3015491997767	7940	0
2	7940	14204	11087	5.5435	1467.8097450233447	78.125	2080.604610318887	6264	0
3	14204	20793	17322	8.661	1667.5398359326957	62.5	2504.2398261577323	6589	0
4	20793	28804	24264	12.132	1440.8733899159174	97.65625	1949.559487636669	8011	0
5	28804	36300	33345	16.6725	1454.339704567291	111.328125	2128.166069680865	7496	0
6	36300	42884	39255	19.6275	1477.4981732312035	101.5625	2276.562506686403	6584	0
7	42884	49739	46514	23.257	1468.6929470439052	111.328125	2247.4948595174037	6855	0
8	49739	56952	52964	26.482	1487.5789159947687	101.5625	2402.116668292869	7213	0
9	56952	63458	60940	30.47	1545.0572341794596	97.65625	2365.1100570378057	6506	0
10	63458	69338	65977	32.9885	1528.1996924575426	83.984375	2272.1228895380914	5880	0
11	69338	75964	72700	36.35	1342.7956928276917	146.484375	1875.514074397626	6626	1
12	75964	88400	79229	39.6145	1360.7729941743094	150.390625	2657.2553411173526	12436	1

Fig.11. An example of the output data of the feature extraction pipeline.

3.5.5. Validation: Compare the algorithm's detected "optimal repetition" against the user's subjective feedback on soreness.

3.5.1. Fatigue classification algorithm

1. **Training of the classifier**

1.1. The input of the algorithm is the extracted RMS and MDF features for all the detected repetitions and the label of the fatigue onset per sets.
1.2. We organized the data into a feature matrix (X), target labels (y), and grouping identifiers (groups, patient and session IDs to prevent data leakage during training).
1.3. We created a machine learning model using build_model(),
  1.3.1. First we standardized the features using StandardScaler (making the mean zero and standard deviation one).
  1.3.2. We used Logistic Regression for the classification using class_weight="balanced" argument to automatically adjust weights inversely proportional to class frequencies in the input data during training, which helps mitigate issues with imbalanced datasets.
1.4. We trained the model with GroupKFold cross-validation to split the data into the desired number (n_splits) of folds. We are using Out-Of-Fold (OOF) prediction, which instead of training one final model, trains a separate model for each fold and generates predictions for the data points that were left out (the "out-of-fold" data).
1.5. We analyzed the OOF probabilities to find the optimal classification threshold that maximizes a performance metric.
  1.5.1. We sweep across a defined range of thresholds (e.g., 0.05 to 0.95). For each threshold, we calculate the Balanced Accuracy Score, which is a good metric when dealing with imbalanced data. We find the threshold that yields the highest balanced accuracy score.

2. Triggering/Onset Detection Functions

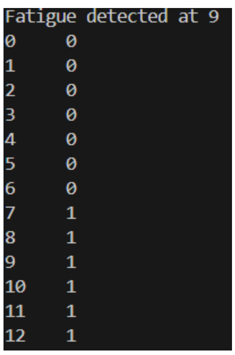
2.1. We find the actual first index in a group where the ground truth label (is_fatigued) becomes 1 (the true onset of fatigue).
2.2. We scan the predicted probabilities for a group and "triggers" (returns the index) the first time the probability exceeds a certain threshold (thr) for M consecutive steps.
2.3. Another approach: It triggers when M out of the last N probability values (within a sliding window) are above the threshold. This makes the trigger less sensitive to single-point noise.

3. Evaluation of the classifier

Finally we calculated various performance metrics (accuracy, precision, recall, confusion matrix, etc.) using the best threshold.

The prediction will come up in form, where repetitions under fatigue is labelled as **1**, while the repetitions before fatigue labeled as **0** (see: Fig.12.).

[Fig.12.Output of classifier]



4. Results

Classification report (OOF):

- Not fatigued (0): precision **0.939**, recall **0.818**, F1 **0.874** (support 264)
- Fatigued (1): precision **0.711**, recall **0.894**, F1 **0.792** (support 132)
- Overall accuracy: **0.843** (396 repetitions)

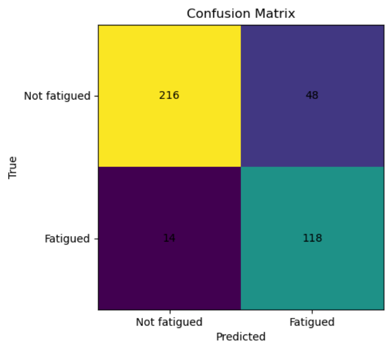


Fig.11. Confusion matrix (repetition-level): Based on 396 repetitions

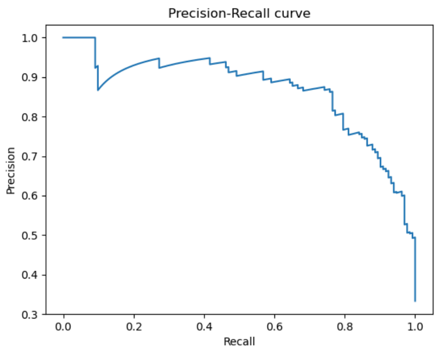


Fig.12. Precision-Recall curve: AUC: 0.868

Cross-validated (OOF) classification performance

Threshold selection was performed on out-of-fold (OOF) probabilities by maximizing balanced accuracy.

- OOF best threshold: 0.385
- OOF Balanced Accuracy: 0.856
- OOF ROC AUC: 0.933
- OOF PR AUC: 0.868

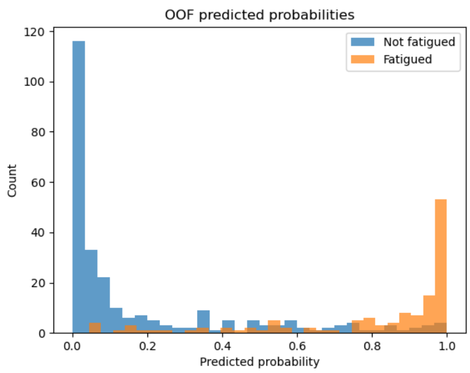


Fig.13. Out-of-Field (OOF) predicted probabilities

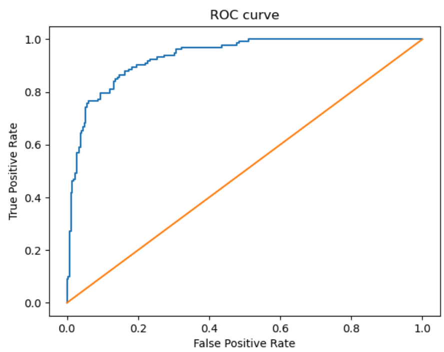
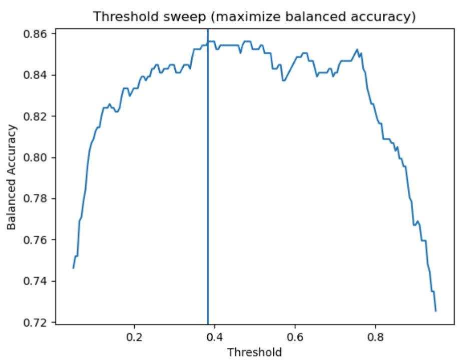


Fig.14. ROC curve

Fig.15. Threshold sweep



5. Discussion

5.1.Performance summary

Trigger evaluation measures how closely the first trigger aligns with labeled onset (when available). Errors are reported in repetitions.

Name	Value	Description
files_with_onset_and_trigger	**26**	**Sessions where a labeled fatigue onset exists and the trigger fired at least once.**
never_triggered	**2**	**Sessions where the trigger never fired — these represent file-level false negatives.**
mean_delta_reps	**0,308**	**On average, the trigger fires about 0.3 reps late**
median_delta_reps	**0,0**	**The median indicates it fires exactly on time in half of the sessions.**
mae_reps	**1,231**	**Mean absolute timing error: the trigger is typically within about 1 rep of the true onset.**
pct_within_0_reps	**0,269**	**The trigger fires exactly on the onset rep in ~26.9% of cases.**
pct_within_1_rep	**0,692**	**About 69.2% of triggers occur within ±1 rep of the true fatigue onset.**
pct_within_2_reps	**0,885**	**Roughly 88.5% of triggers occur within ±2 reps of onset.**
early_rate	**0,308**	**Among sessions where the trigger fires: Early in 30.8% of cases**
late_rate	**0,423**	**Among sessions where the trigger fires: Late in 42.3% of cases**

This reflects a system slightly biased toward late detection, which reduces false alarms but may delay early-warning functionality.

5.2. Limitations of the study

During the acquisition, some differences in the signal shapes of the different participants were experienced. Initially, the electrodes were positioned at the center of the biceps with a distance of approximately 4 cm. After some testing, we have seen that better results were achieved by moving the electrodes slightly to the outer part of the biceps and having them 2-3 cm apart. We suppose that there is some correlation between poor-quality signal and hair on the arm. Indeed, signals acquired from female participants were slightly better, both raw and after preprocessing, than male signals.

Another point to discuss should be the real knowledge of the participants about their capabilities. Some users were aware of their inexperience and, for this reason, they went through a calibration phase to know their real capabilities. Some other users weren't able to arrive at an exhausted point through the first sets, so we had to increase the weight during the acquisition. This changed the protocol of the acquisitions. The study could be improved by testing a greater number of volunteers and evaluating their results in stratified groups based on their gender, and previous fitness experience.

A further factor in getting reliable results is the proper execution form of the biceps curls. The curls were executed in a standing position without any assistance (not like in the case of Fig.5.), which in case of inexperienced users could have led to an improper form, where they used momentum and shoulder muscles to lift the weight, instead of clearly using only their biceps muscle.

The main problem remains the quality of the acquisition, which is probably strongly correlated to hair in the area of acquisition, since it increases the electrode impedance, resulting in low-quality signal. This artefact could be improved by placing conductive gel under the electrodes.

6. Conclusion

The project successfully identifies trends where RMS increases and MDF decreases as the set progresses toward failure. The algorithm's estimated "optimal repetition" is compared against the user's recorded subjective soreness point (Ground Truth) to validate accuracy has been proved to be an efficient way for fatigue detection.

7. Future Work

The project could be improved by acquiring greater and more precise dataset to train the model and trying different classifiers can further increase the efficiency of the implemented algorithm.

References

[1] Oliveira, Hugo Barcelo, Vinicius Abrão da Silva Marques, and Luciane Fernanda Rodrigues Martinho Fernandes. "Analysis of the force–time curve and median frequency of surface electromyographic signals during isometric hand grip test for estimation of a temporal pattern for muscle strengthening." *Research on Biomedical Engineering* 39.1 (2023): 179-187.

[2] Muceli, Silvia, and Roberto Merletti. "Tutorial. Frequency analysis of the surface EMG signal: Best practices." *Journal of Electromyography and Kinesiology* 79 (2024): 102937.

[3] Parimala, Shyam Prasad, Pranoti P. Shinde, and Sumalatha Naitham. "Evaluation of Muscle Fatigue Using Surface Electromyography during Isometric Contractions in Athletes and Non-Athletes." *European Journal of Cardiovascular Medicine* 15 (2025): 682-685.

Appendix

All of the codes, data and documentation connected to the project can be accessed on GitHub: https://github.com/muqsitamir/EMG_fatigue_detection.git

We implemented both the planned signal-processing pipeline (filtering, envelope, repetition segmentation, RMS/MDF extraction, trend analysis) and the planned ML workflow (standardization, cross-validated logistic regression, threshold-based onset detection, and evaluation). We successfully observed the expected fatigue trends (RMS increasing and MDF decreasing) and validated detection against recorded subjective fatigue/onset labels. Regarding the original data collection goal, we recorded 12 participants as planned but due to time limitations with the EMG device, there were some issues in recording some of the sets i.e., some recordings were excluded from analysis and model training, leaving only the usable sessions in the final dataset. We also went beyond our original plan by adding a deployable ML-trigger mode (trained classifier + tuned probability threshold and onset-timing evaluation) not just the initial signal-trend heuristic approach.

Task	**Team member**
Data acquisition (experiment)	Dario, Muqsit, Zsuzsanna
Measurement protocol	Dario
Signal processing pipeline	Dario
ML Classification Pipeline (training + inference + utilities)	Muqsit
Codebase Structuring	Muqsit
Report, documentation	Zsuzsanna
Measurement administration tasks	Zsuzsanna