POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

A multi-modal approach combining first person vision and surface EMG for the functional assessment of the upper extremities: a feasibility study

Tesi di Laurea Magistrale in
Biomedical Engineering - Ingegneria Biomedica

Giada Zecchin, 10602586

Abstract: Upper extremity (UE) function recovery is essential to improve the quality of life and it is a top priority for people with spinal cord injury (SCI) and stroke. Recent years have seen advancements in daily life assessment and rehabilitation using wearable technologies like first person vision (FPV), which has however some limitations. The aim of this thesis is to develop a multi-modal system that combines FPV and surface electromyography (sEMG) to record activities involving the UE and validate its performance on healthy adults during standardized tasks in a lab setting. The performance of sEMG in detecting functional interactions (i.e., manipulation an object for a functional purpose) was evaluated by training and testing four machine learning algorithms. A deep learning algorithm was utilized to detect hand-object interactions from the FPV frames. The resulting detections were analyzed to obtain the FPV classification performance. Three different combinations of FPV and sEMG were then tested. The first involved concatenating the interaction state from the FPV analysis with the most relevant features extracted from the sEMG signal. The second and third approaches used the AND and OR logical operators, respectively, to combine the interaction state predicted by the sEMG signal and FPV analysis. The performance of single-modal and multi-modal approaches were evaluated using accuracy, F1-score, precision, recall, and specificity metrics. Hand–object interactions were automatically detected with a median accuracy of 0.653 (0.044) for sEMG, 0.650 (0.067) for FPV, 0.716 (0.067) for the FPV and sEMG combination. Our results demonstrated that the multi-modal strategy outperformed the two single-modal approaches, as demonstrated by most of the evaluation metrics used in this study. This study suggested that the combination of FPV and sEMG is an effective way to capture both functional and non-functional hand-object interactions in healthy individuals. Future research will involve validating our findings in subjects with SCI or stroke, both in a clinical setting and in their homes.

Advisor:
Prof. Emilia Ambrosini

Co-advisors:
Andrea Bandini

Academic year:
2022-2023

Key-words: first person vision, surface electromyography, multi-modal approach, hand-object interaction, hand function

Contents

1. Introduction

1.1. Upper limb impairment

1.1.1 Stroke

Stroke is a potentially fatal illness that is the leading cause of disability and fatalities globally [1].
Stroke has two primary causes: ischemic, which accounts for 87% of occurrences and is caused by a blood clot that blocks blood flow, and hemorrhagic, which accounts for 13% of instances and is caused by a weak blood vessel supplying the brain that bursts [1]. Over the last three decades, the risk of developing a stroke has increased by 50% [2]. The 2022 Global Stroke Fact sheet indeed reports a substantial increase in cases and deaths due to stroke from 1990 to 2019 [2].
The main risk factors for stroke are depicted in Figure 1, taken form [2].

Figure 1: Stroke risk factors [2]

Stroke survivors have to deal with life-long problems and the severity of the disability is determined by several factors such as the duration of the lack of oxygen and the area of the brain involved [3].
Stroke may suffer a range of cognitive, psychological, and physical deficits, including difficulties with speech, memory, judgment, spatial awareness, attention, and somatic sensations like touch, warmth, pain, and proprioception [3][4]. Motor deficits include limb weakness or paralysis, with hemiparesis of the contralateral upper limb being the most common (77.4%) [5]. The American Stroke Association defines hemiparesis as a weakness or inability to move on one side of the body, which can lead to challenges in walking, balance, coordination, and physical fatigue [6]. As a result, daily activities may be significantly impacted. In fact, the majority of stroke survivors struggle to perform basic actions like reaching, picking up, and holding objects [4].
Automatic recovery mechanisms, such as neuroplasticity, take place after a stroke [7]. Especially during the first month after stroke, the brain reorganizes and rewires itself to make up for the damaged areas [8].
Rehabilitation is crucial for stroke survivors to regain lost function and promote brain plasticity for maximum recovery [9]. Its implementation should depend on the stage of neurological recovery, and timing and intensity are crucial to its success [9]. Improvement profiles might change over time and don't always follow a predictable pattern [10].
Recovery after a stroke usually follows a proportional recovery rule [11]. Patients with mild initial deficits tend to recover better and faster than those with severe deficits (Figure 2) [11]. Personalized approaches that consider individual differences in the nature and severity of the stroke, as well as the patient's age and overall health, are necessary [10].

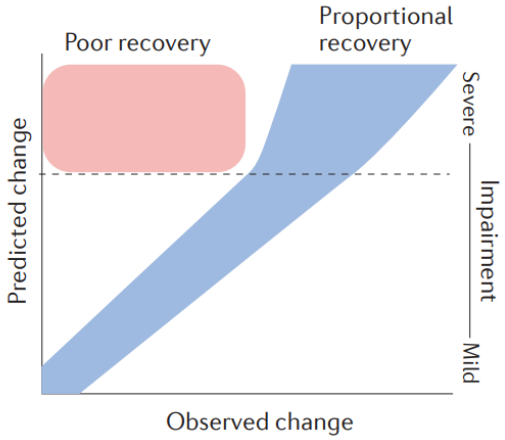

Figure 2: Proportional motor recovery in the upper limb taken from [11]: plot of the observed changed against the predicted changes 3 months post stroke. The recovery of patients in the blue area corresponds to the predicted outcomes, whereas those in the red area experience less favorable recovery than expected. Patients located above the dotted line initially exhibit severe impairment, with approximately equivalent numbers of individuals experiencing either good or poor (proportional) recovery.

1.1.2 Spinal cord injury

Another frequent reason for upper limb disability is spinal cord injury (SCI) [12]. The World Health Organization (WHO) reports that SCI occurs at a global rate of 40 to 80 cases per million individuals per year, with 90% of cases caused by traumatic incidents [12].
SCI is characterized by damage to the spinal cord that can lead to lasting or temporary changes in its typical operation [13]. When the spinal cord in the cervical region is injured, it can cause tetraplegia, which is the paralysis and loss of function of all four limbs and the trunk [14].
This type of injury can lead to other health issues such as deep vein thrombosis, urinary tract infections, muscle spasms, osteoporosis, pressure ulcers, chronic pain, and problems with bladder and bowel control [12][15].
The degree of neurological recovery in patients with traumatic SCI is influenced by various factors, such as the level and severity of the injury, as well as the age and general health of the patient [16]. Patients generally show some recovery of motor function in the weeks and months after the injury, with significant improvement by the nine-month mark [17]. The rate of recovery tends to slow down over time, typically reaching a plateau by 12 to 18 months [17]. However, some patients may continue to experience improvements in motor function for several years after injury [17]. They can achieve high level of function and self-reliance through appropriate therapy and rehabilitation [18].

1.2. Continuum of care

Recovery can vary significantly from person to person, depending on many factors (see section 1.1). The priorities identified by stroke survivors include improvement of upper limb function, recovery of mobility, reduction of spasticity, interventions for language and communication, and support for psychological and emotional problems [19]. Research studies ([20][21]), have also highlighted the importance of upper limb function recovery in improving the quality of life of people with SCI.
Assessment scales are frequently used to evaluate and monitor the recovery of individuals with stroke including the National Institutes of Health Stroke Scale (NIHSS) [22], Modified Rankin Scale (mRS) [23], Fugl-Meyer Assessment (FMA) [24], Action Research Arm Test (ARAT) [25], Motor Activity Log (MAL) [26], Stroke Impact Scale (SIS) [23] and Barthel Index [23]. These scales assess various aspects, such as motor function, cognition and language. Scales used to assess UE function in SCI patients include the International Standards for Neurological Classification of Spinal Cord Injury (ISNCSCI) [14][27], Upper Extremity Motor Score (UEMS) [28], the Spinal Cord Independence Measure (SCIM) [28], and Graded Redefined Assessment of Strength, Sensibility and Prehension (GRASSP) [28].
Although such scales are useful for measuring functional capacity, they may not accurately assess real-world performance due to their subjective nature and lack of specificity [22]. Furthermore, they are typically designed for use at specific time points for example at hospital admission or discharge, which may influence their accuracy and usefulness [23].
Recovering the use of UE can be challenging for several reasons [29][30]:

- Lack of outcome measures that take into account the performance (and not only capacity);
- Limited access to rehabilitation services, including issues with distance, transportation, cost, and inadequate insurance coverage;
- Patients may also face difficulties engaging in rehabilitation due to depression, fatigue, and anxiety;
- Specialized interventions may be necessary as standardized conventional therapies may not consider individual differences;
- Financial pressure on the healthcare system, which results in early discharge.

Monitoring the functional use of the upper extremities in real-life settings is crucial to tailor therapy and maximize rehabilitation outcomes [30].

1.3. Method of daily life assessment

Researchers have suggested integrating robotics, cameras, and wearable systems to overcome the limitations of clinical assessments and improve rehabilitation outcomes. These devices can provide more objective and reliable assessments in uncontrolled environments and offer effective exercises tailored to the patient's needs and progress [31][32].
There is growing interest in wearable technologies, such as accelerometers that measure the acceleration of motion, gyroscopes that measure orientation and rotation, and surface electromyography (sEMG) that measures electrical activity generated by skeletal muscles [33].

1.3.1 Accelerometer and IMU

Gomez-Arrunategui J. et al. (2022) presented the Arm Rehabilitation Monitor (ARM). This wrist-worn device uses an Inertial Measurement Unit (IMU) and machine learning techniques to detect reaching actions of stroke survivors during assessment tasks and Activities-of-Daily-Living (ADL) [34]. Similarly, Lum et al. (2020) also adopted a wrist-worn accelerometer to collect data, which was then analyzed using machine learning algorithms to measure upper limb movement directly of stroke patients [35]. Additionally, Biswas et al. (2014) proposed an algorithm for detecting arm movements during daily activities. They used tri-axial accelerometers strapped to the wrist to detect movement [36]. Nevertheless, these tools cannot tell whether a user has effectively gripped an object [30][37].

1.3.2 Force myography

The ability of force myography (FMG) to record hand actions like grabbing has also been examined [37][38]. FMG is a method of evaluating the condition of the underlying muscle-tendon complex by monitoring the pressure or force on the surface of the limb. The band detects arm movement before or after the wrist activity detection, and the device generates a unit called a hand count, which counts as one [37][38]. Sadarangani G. et al. (2017) demonstrated its capability to differentiate between grasping and non-grasping movements during daily activities. They proposed using force myography combined with machine learning methods to monitor the practical use of the hand in everyday activities [37]. Grasping was also analyzed by Lisa A.Simpson et al. (2019) [38] and Yang C. et al. (2021) [39] using a prototype version of the TENZR wristband (BioInteractive Technologies, Canada) to capture reach-to-grasp repetitions. The wristband featured proximity sensors, inertial measurement units, and force myography sensors. However, this device cannot differentiate between various grasp and object interactions when minimal wrist and finger movement occurs [33] [39].

1.3.3 Smart glove

A smart glove was presented by Dutta D. et al. (2022) to asses patients' grasping abilities. The device was equipped with multiple sensors, including bend and force sensors and an accelerometer, and it was coupled with a machine learning algorithm to quantify grasping abilities [40].
However, those with spasticity and a limited range of motion could find it difficult to use such devices. Moreover, patients who wear them may notice less palmar feeling [37].

1.3.4 Surface electromyography

sEMG is a non-invasive technique that involves placing small electrodes on the surface of the skin in order to detect the electrical signals produced by the muscle contractions [33][41].
Tang W. et al. (2018) investigated neuromuscular changes in proximal and distal muscles post-stroke [42]. Jones C. et al. (2018) measured involuntary coupling activity in the thumb and finger during hand movements, a compensatory mechanism to generate sufficient force during hand movements [43]. Furthermore, Hesam-Shariati

N. et al. (2017) explored the correlation between improvements in motor function resulting from therapy and changes in muscle activation patterns, highlighting the significance of utilizing objective measures like EMG signals to track progress [44]. Lee S. et al. (2011) suggested a method of sEMG pattern classification to identify the muscle activation patterns of stroke survivors by placing electrodes on the hand and the forearm [45].
Wearable technology that incorporates sEMG sensors enables non-intrusive, continuous data collection while patients move about and go about their everyday activities, providing direct feedback on muscle activation [33]. One example of an sEMG-embedded armband is the Myo armband (MYB), whose design has made it a well-liked option among researchers [46]. It has been used, for example, for hand gesture recognition, among other applications [46][47][48].
Xie1 T. et al. (2020) examined the changes in muscle activity in stroke patients with spasticity using high-density EMG [49]. With HD-sEMG, EMG signals from numerous channels are simultaneously recorded across a particular body region [50]. This is achieved using an array of electrodes displayed in a specific arrangement to capture muscle activity patterns. HD-sEMG can provide information on the activity and coordination of specific muscles during movement by identifying relevant patterns [50]. However, most of the HD-EMG recording systems currently available like the Refa (TMSi, The Netherlands) are typically stationary, they require a main power source and are therefore too bulky and heavy to be used on wearable devices [50]. There have been proposals such as those described in [51], [52] and [50] address the need for more portable and flexible HD-EMG recording systems that can be worn comfortably for extended periods.
sEMG can be combined with various devices to enhance their effectiveness in rehabilitation and training. For example, it can be used with exoskeletons to stimulate the muscles and improve the range of motion in the joints. Studies such as those of Lu Z. et al. (2017) [53] and Qian Q. et al. (2017) [54] investigate the effectiveness of (EMG)-controlled robotic arms. Through the use of surface electrodes, the patient can control the movements of the robotic hand. Mulas M et al. (2005), for example, proposed an exoskeleton comprised of two support structures: one for the hand and the other for the wrist, interconnected by a set of actuators and sensors controlled by surface EMG. However, even though the device has demonstrated validity in clinical trials, additional testing is needed to determine its effectiveness in home-based rehabilitation scenarios [55].

1.3.5 First person vision

The development of telemedicine and remote monitoring may completely alter how post-acute stroke care is provided by introducing more convenient and affordable solutions [29].
Egocentric vision (or First-Person Vision - FPV) is a potential tool for monitoring hand use in natural environments. Through FPV it is possible to capture the user's point of view and, if the camera is worn on the head, to focus on hand movements and manipulation, minimizing the presence of self-occlusions [21][56]. An example is depicted in Figure 3.

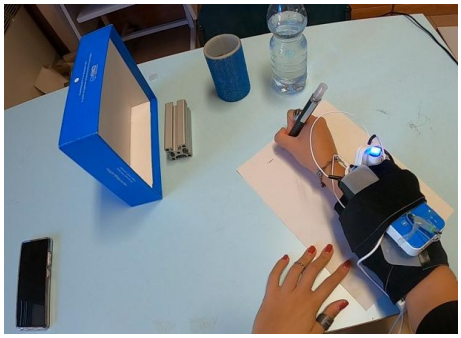

Figure 3: A sample frame captured with a wearable camera worn on the head, during writing tasks.

FPV devices have been developed since 1997 [57], mainly for research purposes. With recent technical developments, several businesses have displayed an inclination toward such devices [57]. Through their survey, Betancourt et al. (2015) explored the evolution of FPV methods and the advancement of commercial FPV devices like Google Glass (Google, USA) and GoPro (GoPro Inc., California) [57].
Unlike the previously mentioned wearable sensors, FPV can provide information on functional hand-obect interaction and grasp types, offering valuable insights into functional hand use: it enables the automatic detection of hands and objects and the specific actions and activities taking place [30].
Third person vision (TPV), which refers to the use of visual feedback tools such as mirrors or video recordings to view upper limb movements from an external perspective, can also be used as a rehabilitation technique [58].
Several articles developed FPV approaches to detect and analyze hand movements [20] and to use in various contexts, such as rehabilitation for people with from SCI [30][59] or upper limb impairment [56], as well as

for stroke recovery [60]. Tsai et al. (2023) offered the first proof that hand-use ratios determined from the egocentric video can be used to assess hand function after stroke [61].
Bandini et al. (2020) [21] summarized the existing hand-based FPV techniques used to examine egocentric data. These approaches can be divided into three primary categories: localization, interpretation, and application. The first step is localization, which involves identifying the position of the hands within the camera's frame. Hand detection, segmentation, and identification are subareas of hand localization (Figure 4). Hand detection involves locating the hands in an image or video frame, usually by fitting a bounding box around the detected hand, while hand segmentation implicates separating the outline of the hand from the background in an image or video frame based on color, texture, and shape. The identification of hands distinguished between, for example, the left and right hands.

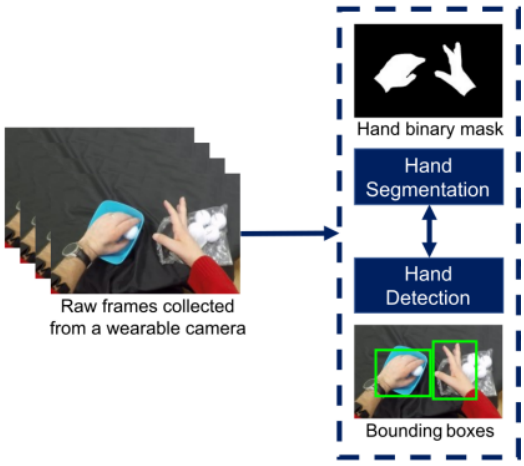

Figure 4: Hand detection and segmentation from [21]: A schematic diagram outlining the process of hand localization and segmentation in FPV.

The interpretation category provides a way to capture more detailed information about what the hands are doing beyond just their physical location, for instance, if hand-object interaction occurs. According to Tsai M. et (2021), a functional hand-object interaction is manipulating an object for a functional purpose. Detecting hand-object interaction can serve as a valuable metric to assess the independent use of the hand in everyday tasks at home [60]. Application is the final category in the taxonomy of hand-based methods in FPV. It involves taking the information gained from the previous categories and applying them to create useful real-world applications to solve specific problems. For instance, one of the primary application areas is healthcare, where this technology can be used to monitor and improve patient rehabilitation and motor skills [21].

1.4. Objectives

Although FPV is a helpful tool for recording activities involving the upper extremities, it fails to discriminate interactions in presence of "non-standard" grasps and compensatory strategies [61]. The research hypothesis is that such limitation can be overcome by combining FPV with sEMG. However, sEMG alone does not provide contextual information about how hand-use activity relates to daily life.
Therefore, this thesis aims to develop a multi-modal system that combines FPV and sEMG to record activities that involve the use of the upper extremities. The goal is to validate this system's performance in detecting functional and non-functional interactions among healthy adults during standardized tasks in a lab setting. In this study, we defined functional interaction as a manipulation of the object, any contact with a fixed or portable object, while non-functional interaction referred to no contact between the hand and an object, any self-contact and any contact with another person. Once the system's effectiveness is verified, it will be utilized on SCI and stroke patients to provide more precise and detailed information about their hand-use activity. This will make it possible for physicians to create individualized rehabilitation plans that are tailored to each patient's needs.

2. Materials and Methods

2.1. Participants

A group of 10 healthy adults was recruited for this experiment at the Biorobotics Institute, Scuola Superiore Sant'Anna, Pisa (Italy). The Joint Ethical Committee of Scuola Superiore Sant'Anna and Scuola Superiore Normale (Nr. 3/2023) approved this study and all subjects provided their written informed consent before starting the acquisition. The average age of the group was 27, comprising six males and four females, all right-handed.
Inclusion Criteria

- Adult men or women >18 and <65 years;
- To be able to use their hands during regular ADLs at home, based on self-report;
- To be able to turn off the camera on their own.

Exclusion Criteria

- Presence of wrist or hand deformities or injuries to the UEs;
- Presence of UE impairments that prevent them from using their hands in any ADLs.

2.2. Experimental set up

The study consisted of a series of predetermined tasks that required participants to use ULs while wearing a sleeve equipped with sEMG sensors and a wearable camera mounted on their heads. The instrumentation used in the experiment included:

2.2.1 OTBioelettronica Sessantaquattro

OTBioelettronica Sessantaquattro hardware is comprised of an EMG acquisition system, with a portable and compact design. sEMG signals were acquired with the OTBiolab + software (OTBioelettronica, Turin, Italy) [62]. The software offers a range of features, including real-time visualization of signals, which can provide direct feedback for quantitative assessments during acquisition [62]. Additionally, it offers signal filtering, calibration, and the option to save and export data for further analysis [62]. It comprises a surface electromyographic amplifier equipped with Wi-Fi communication, creating a wireless access point capable of supporting up to 64 channels. During the acquisition process, only 32 channels were utilized and no filtering was applied, leading to a raw signal sampled at 2k Hz. Additionally, it was possible to convert OTB+ files directly to Matlab files, including multiple files at once.

2.2.2 Synchronization circuit

The device included two additional auxiliary channels (AUX), one of which was directly linked to a synchronization circuit. This circuit was composed of a button and two LEDs. When the button was pressed, power was supplied to the LEDs, and 3.3mV-pulses were transmitted to the two AUX. This LED-activated switch provided synchronization cues to the camera. The other AUX connected to the reference electrode, which was a wet electrode.

2.2.3 MXene electrodes

The experiment utilized MXene dry electrodes (Figure 5), made of a silicon substrate and a two-dimensional transition metal carbide nanomaterial [63]. Dry electrodes do not require conductive paste or gel to make contact with the skin, unlike wet electrodes [64]. Due to their excellent electrical properties, MXene electrodes are susceptible and almost as effective as wet electrodes [63]. They are also reusable, making them more cost-effective than disposable wet electrodes [63]. Moreover, the flexibility to place electrodes in any desired pattern enables researchers to achieve optimal configurations. As a result, the electrodes can be positioned in the most effective locations for measuring the electrical signals of interest.

Figure 5: Maxine electrodes used for sEMG acquisitions: it is composed of 32 dry electrodes

2.2.4 Camera

The camera was a GoPro Hero 8 (GoPro Inc., California), which was strapped to a headband to allow for hands-free acquisitions. The camera was set to record video footage at a resolution of 1080p, which produced images with a 1920 x 1080 pixel width and height. The footage was recorded at a rate of 30 frames per second, which meant that 30 still images were captured every second.

2.3. Experimental protocol

During the single session that lasted about 60 minutues, sEMG signals indicating muscle activation of the forearm and video of activities performed were recorded. The experimental session consisted of three phases:

1. setup of the EMG system;
2. setup of the wearable camera;
3. data collection during task execution

2.3.1 Step 1

Step 1 involved setting up the sEMG system on the subject's forearm, which took about 5 minutes.
The length of the ulna was first measured for each participant from the olecranon (tip of the elbow) to the styloid process (head of the ulna) [65]. Also, the forearm's maximum and minimum circumferences (i.e., its proximal and distal halves, respectively) were measured. These measurements will allow researchers to develop customized sEMG sensors based on anatomical dimensions in the near future.
Before positioning the electrodes, the skin of the forearm was cleaned with an alcohol solution and then hydrated with a damp cloth. The sEMG array was wrapped around the proximal part of the participant's forearm. The crest of the ulna was used as the starting position for the wrapping direction, moving from the crest of the ulna to the dorsal side and finally to the ventral side. The reference electrode was placed on the styloid process, an electrically unaffected but nearby area. The OTBioelettronica Sessantaquattro hardware and connectors were placed on the forearm and kept in place with elastic bands and medical tape. The recordings were started and stopped from a dedicated laptop by the research team.

2.3.2 Step 2

Step 2 involved setting up the camera, which took around 2 minutes to be completed. The GoPro camera was mounted on the headband to conduct egocentric recordings of the tasks. The camera tilt angle was adjusted so that the hands and synchronization led were always within the frames. The recordings were also started and stopped from a smartphone by the research team using the GoPro app.

2.3.3 Step 3

The third step concerned the recording of task execution. This step lasted approximately 20 minutes and included several tasks. It was essential to check that the data streaming was operating correctly and that the electrodes were positioned accurately on the skin before starting the acquisitions. Once this was verified, the video recording and sEMG were initiated. The synchronization switch was clicked twice and three times at the beginning and end of each task, respectively. The following tasks were repeated:

1. Maximum voluntary contraction (MVC);
2. Box and block test;
3. Pretend to pour a full water bottle into a glass and pretend to drink;
4. Write a sentence on a sheet of paper;
5. Use a smartphone to type a message and scroll web pages;
6. Non-functional hand movements (opening-closing fist, raising arm, etc.);
7. Walking;
8. Rest position.

Finally, the multimodal data collection was terminated by stopping the video and sEMG recordings. The data collection was repeated twice to capture data from both ULs. Examples of frames captured with the GoPro camera during the task executions are illustrated in Figure 6.

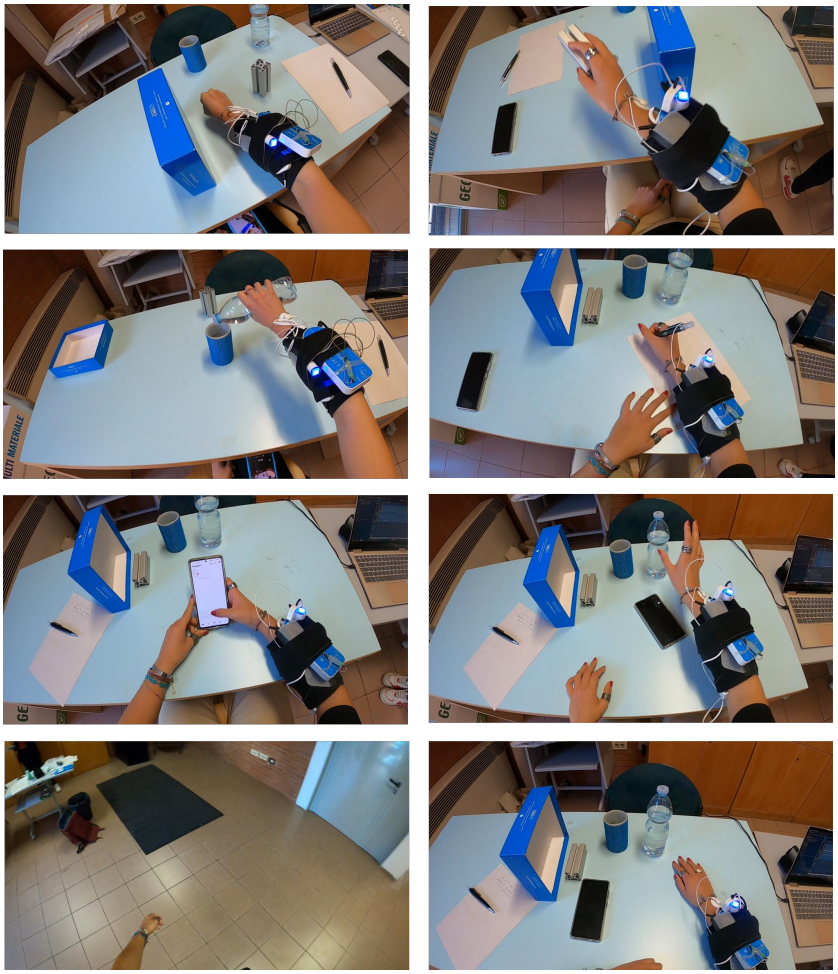

Figure 6: Examples of egocentric frames collected form the study during the task executions.

2.4. Data analysis

2.4.1 sEMG data analysis

Labelling An extensive manual labeling process was performed to obtain the ground truth of the hand-object interaction state (presence or absence of interactions). Each video was watched in slow motion to observe the hand-object interactions carefully. Whenever a contact was detected, the video was paused and rewound to mark the frames corresponding to the start and end of the interactions. All frames in between the start and end frames of each movement were labeled as functional interactions (label = 1), while the periods between task repetitions or hand/arm movements were labeled as non-functional (label = 0). To synchronize the ground truth frames with the sEMG samples, a process of alignment was performed. A Matlab script was created to automatically align the spike from the trigger channel with the corresponding frame that detected the LED.

sEMG processing Matlab R2022b was used for all sEMG processing. The sEMG acquisitions obtained using the OTBlab+ software were converted into .mat files and subsequently imported into Matlab for analysis. First, inter-task pauses (i.e., intervals when the participants prepared themselves for the next task) were removed. This was achieved by utilizing the trigger channels, specifically, the 33rd channel containing square waves resulting from the taps on the synchronization switch. To remove such intervals, spikes were derived from the square waves. The samples between the conclusion of one task (three spikes) and the beginning of the following task (two spikes) were then eliminated. In some of the sEMG recordings, spikes were not detected because some subjects failed to press the LED buttons accurately. In such cases, it was necessary to identify the beginning and end of the task manually.
In Figure 7, all the start and end spikes are depicted, and red rectangles correspond to the deleted samples.

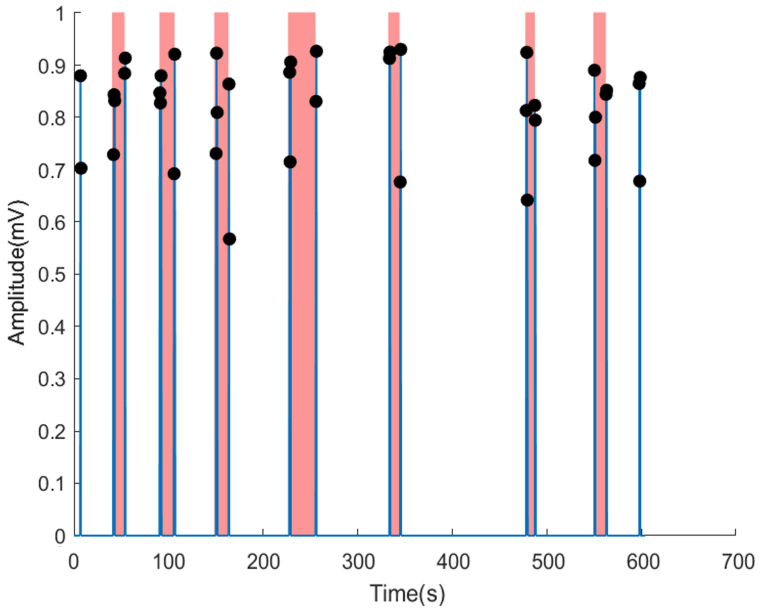

Figure 7: Trigger spikes: 2 consecutive spikes represent the start of the task, while 3 spikes depict the end. The red rectangles correspond to the deleted samples.

After performing a spectral analysis with a Fourier Transform on the signal, a power line interference at 50 Hz was detected, causing distortion of the signal. To minimize this effect on sEMG recordings, a Notch filter (2nd-order infinite impulse response (IIR) bandstop filter) with cut-off frequency was at 49 Hz and 51 Hz was applied.
The bandwidth of interest for the sEMG signal is typically below 500 Hz . However, there may be motion artifacts below 5 Hz that alter the signals [66], For this reason, the signal was filtered with a Butterworth bandpass filter between 10-250Hz.
The next step was signal normalization. The maximum amplitude value was obtained from the first portion of the signal of each of the 32 channels, corresponding to the Maximum Voluntary Contraction (MVC). Firstly, this portion of the signal was rectified to capture both positive and negative components equally. The maximum value was calculated over a 100ms moving average window. A 100ms window was chosen to level out the peaks without distorting the signal (see Figure 8).

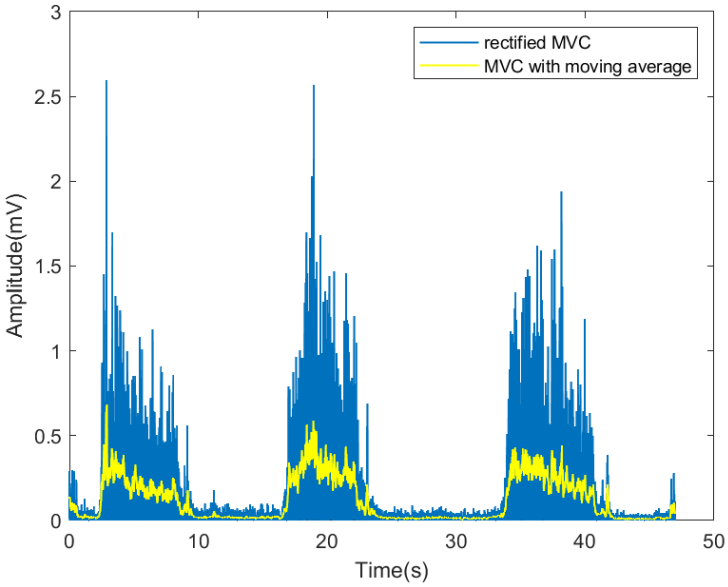

Figure 8: Maximum Voluntary Contraction: the blue signal is the rectified MVC while the yellow one is the MVC calculated over a 100ms moving average window.

After this procedure, the portion relative to MVC was removed from each channels. All of the remaining recordings were then normalized with the previously determined maximum value.
Afterward, all silent channels were eliminated. Due to damaged electrodes or incorrect electrode placement, the actual number of usable channels was, on average, less than 28.
A method for minimizing the dimension is Principal Component Analysis (PCA) [67]. A dataset benefits from having fewer variables while maintaining as much data as possible. A dataset's first principal component captures the most variability in the data and is the direction in which the data varies the most [67].
PCA was applied and the signal was projected onto the first 11 principal components. The decision to retain only the first 11 principal components was based on two main factors: (1) these 11 components accounted for more than 90% of the total variance present in the data, making them the most important for accurately representing the sEMG signal channels (2) there are approximately 11 detectable superficial forearm muscles, and therefore retaining the first 11 principal components may allow for better characterization and identification of the underlying muscle activations. The twenty forearm muscles (twelve superficial and eight deep muscles, Figure 9) play an essential role in controlling the grasping and mobility of the hands as they regulate the wrist and fingers' flexion, extension, abduction, and adduction by inserting directly into the wrist and hand [68].

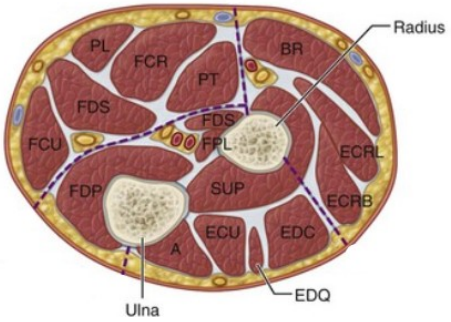

Figure 9: Cross section of the forearm. Brachioradialis (BR), Extensor carpi radialis longus (ECRL), Extensor digiti quinti (EDQ), Extensor carpi radialis brevis (ECRB), Extensor digitorum communis (EDC), Extensor carpi ulnaris (ECU), Flexor carpi ulnaris (FCU), Flexor carpi radialis (FCR), Flexor digitorum sublimis (FDS), Flexor pollicis longus (FPL), Flexor digitorum profundus (FDP), Palmaris longus (PL), Pronator teres (PT), Supinator (SUP) [69].

Feature extraction The filtered sEMG signals were windowed. Windows of 250ms, 500ms, 750ms, and 1s with 50% overlap were chosen to segment the signal. With the use of this approach, it was possible to analyze

how the outcomes changed depending on the window length. Twenty-one features were extracted from each window, twelve in the time domain (TD) and nine frequency domain (FD).
Six time-domain features were generated based on the rectified signal amplitude, while the other six required the signal to be enveloped before feature extraction. Enveloping was carried out using a moving average with a window of 100 ms, i.e. 200 samples. This window length (the same one used for normalization), was chosen as it allowed to smooth the peaks without distorting the signal.
The considered TD features were [70] [71] [72] [73]:

- The Integrated EMG (IEMG), which represents the total muscle activation within a certain period, is frequently used as a pre-activation signal for muscular activity. It is the area under the curve of the rectified EMG signal, so it's calculated with the following formula.

$$IEMG = \sum_{i=1}^{N} |x_i| \tag{1}$$

  with N being the signal length in samples and x the data set.
- Average Amplitude Change (AAC) tracks the overall change in signal amplitude over a specified time period. It is determined by averaging the absolute differences between adjacent samples taken within a specific time range.

$$AAC = \frac{1}{N} \sum_{i=1}^{N} (|x_{i+1} - x_i|) \tag{2}$$

- Mean Absolute Value (MAV) calculates the average amplitude of the rectified signal during a specified time period. Because MAV is measured by averaging the absolute value of the EMG signal, it is an easy way to measure the force of muscle contractions.

$$MAV = \frac{1}{N} \sum_{i=1}^{N} |x_i| \tag{3}$$

- Mean Absolute Deviation (MAD) measures the average difference between each data point and the mean of the entire dataset.

$$MAD = \frac{1}{N} \sum_{i=1}^{N} |x_i - \bar{x}| \tag{4}$$

- Waveform Length (WL) measures the signal waveform's overall length over a time period. It is calculated as the sum of the absolute differences between adjacent samples

$$WL = \sum_{i=1}^{N} |x_i - x_{i-1}| \tag{5}$$

- Log Detector (LD) provides an estimate of the muscle contraction force. It is measured by calculating the logarithm of the rectified EMG signal, therefore is based on the logarithmic connection between muscle force and EMG amplitude.

$$LD = \exp^{(} \frac{1}{N} \sum_{i=1}^{N} \log |x_i|) \tag{6}$$

- Root Mean Square (RMS) is obtained by taking the square root of the mean of the squared values of the EMG signal over a time window. It gives an indication of the force and contraction of the muscles.

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2} \tag{7}$$

- Variance (VAR) calculates the amount of variability in the amplitude of the EMG signal over a specific time period. It is calculated by taking the average of squared value of each sample.

$$Var = \frac{1}{N-1} \sum_{i=1}^{N} x_i^2 \tag{8}$$

- Simple Square Integral (SSI) reflects the overall strength of muscular contraction over time. It is calculated by squaring the EMG signal at each time point, then summing the resulting values.

$$SSI = \sum_{i=1}^{N} x_i^2 \tag{9}$$

- The coefficient of variation (COV) expresses how variable a set of data is. It is calculated as the ratio between the standard deviation and the mean value of the data.

$$COV = \frac{\text{std(x)}}{\text{mean(x)}} \tag{10}$$

- Kurtosis (KURT) defines the pattern of the EMG amplitude distribution across a specific time interval. It is a measure of how prone a distribution is to outliers. A more peaked distribution is indicated by a greater kurtosis value, whereas a more flat or dispersed distribution is indicated by a lower kurtosis value. The formula for calculating kurtosis is:

$$Kurtosis = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2\right)^2} \tag{11}$$

- Skewness (SKEW) is a measure of the asymmetry of the data around the sample mean. The data spread more to the left of the mean than to the right if skewness is negative. The distribution of the data moves closer to the right if skewness is positive. The normal distribution has zero skewness. The skewness is defined as:

$$Skewness = \frac{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^3}{\left(\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2\right)^{3/2}} \tag{12}$$

The Fast Fourier transform was used to transform the EMG signal from the time domain to the frequency domain, producing therefore the power spectral density (PSD). The following FD features, often employed in estimating muscular force and exhaustion, were extracted from the PSD [71].

- Mean frequency (MNF) is determined by multiplying the frequency by the total sum of the power spectra divided by the EMG power spectrum.

$$MNF = \frac{\sum_{j=1}^{M} f_j p_j}{\sum_{j=1}^{M} p_j} \tag{13}$$

where M is the frequency bin's length, p_j is the EMG signal's power spectrum at frequency bin j, and f_j is the frequency variable at that frequency bin.

- Median Frequency (MDF) is a frequency point that separates the power spectrum of EMG signal into two regions with equal amplitude. It is defined as half of the total power.

$$MDF = \frac{1}{2} \sum_{j=1}^{M} p_j \tag{14}$$

- Total power (TP) indicates the total power of the EMG power spectrum. It is also known as the energy and the zero spectral moments. (SM0). It is calculated with the sum of power across all frequencies in the signal.

$$TP = \sum_{j=1}^{M} p_j \tag{15}$$

- Mean Power (MP) denotes the average power of the EMG power spectrum. It is calculated by dividing the signal's total power by the number of frequency components in the signal.

$$MP = \frac{\sum_{j=1}^{M} p_j}{M} \tag{16}$$

- Peak frequency is the frequency at which maximum power is obtained in the EMG power spectrum.

$$PKF = \max\ (p_j) \tag{17}$$

with j=1,..M.

- The First, Second and Third Spectral Moments (SM1, SM2, SM3) are calculated by multiplying the EMG power frequency spectrum by the frequency raised to the power of k, which acts as a weighting function. The degree to which the frequency is raised determines the current order. The area under the resulting spectral curve represents the value of the spectral moment of order k. They describe the overall distribution, symmetry, and concentration of the frequency content of the signal.

$$SM1 = \sum_{j=1}^{M} p_j f_j \tag{18}$$

$$SM2 = \sum_{j=1}^{M} p_j f_j^2 \tag{19}$$

$$SM3 = \sum_{j=1}^{M} p_j f_j^3 \tag{20}$$

- Variance of central frequency (VCF) displays the variation in the power spectrum's center frequency over a given time interval. The frequency where the power spectrum is at its peak is known as the central frequency. It is derived using the EMG signal's spectral moments (SM).

$$VCF = \frac{SM2}{SM0} - \left(\frac{SM1}{SM0}\right)^2 \tag{21}$$

The array of ones and zeros representing the interaction state ground truth was also divided into windows. Within each window, the value of 1 was assigned if the majority of samples were ones, and 0 otherwise.
All of the data resulting from the sEMG signal preprocessing were displayed in table format to facilitate further analysis. The table contained all the samples in the rows and all the features in the columns. With each of the 11 PC windowed and all 21 TD and FD features extracted, the total number of features amounted to 11*21, meaning 231. Furthermore, the first column represented the participant's ID, while the last column contained the interaction state ground truth. Therefore, the total number of columns in the table was 233.
In addition, by using four different window lengths, four separate datasets were produced.

2.4.2 First person vision data analysis

Labelling Extensive manual labeling was required beforehand to obtain several ground truths. The ground truth regarding the interaction state, previously described in section 2.4.1, was obtained by manually identifying the hand-object interaction by reviewing the videos.
Another type of ground truth was obtained to analyze the performance of the hand localization algorithm. Specifically, for each of the 20 recordings, 20 frames were randomly chosen and subjected to manual analysis. We extracted the following information from each of the 400 frames: the subject, frame number, bounding box coordinates, and the identified side.
In addition, the device's LED, which turned on with a tap on the switch, was used to manually mark the beginning and end frames of each task for the FPV recordings.

FPV processing First, the video clips were split into frames to perform data preparation. All frames were reduced from 1920x1080 original resolution to 848x480 pixels to aid future processing.
Shan et al.'s deep learning model (Figure 10), which uses a Faster-RCNN (Region-based Convolutional Neural Network) evaluated on the PASCAL VOC 2007 detection benchmark [74], was used to locate the hands, determine their interaction state, and identify the object they are in contact with [75]. The main advantage of the Faster R-CNN is that it merged in a single network the Region Proposal Network (RPN) [76][75] with Fast R-CNN [77], which reduced the time required for object detection.

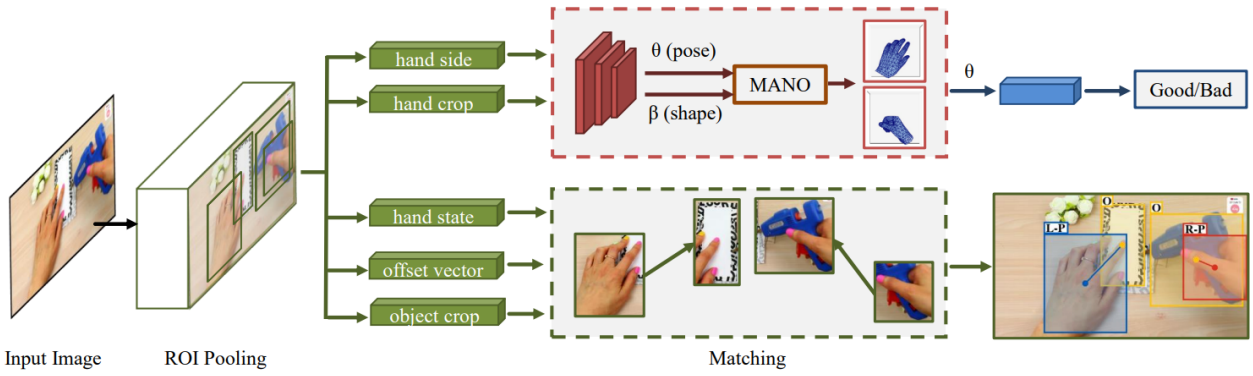

Figure 10: Faster-RCNN taken from [75]: A schematic diagram outlining the process of hand localization and hand-object contact detection in FPV.

The detection results obtained with the Faster R-CNN proposed by Shan et al. [75] were saved on a CSV file (one file per video) composed of 12 columns. Specifically: column 1 contained the frame's name; column 2 indicated whether the row referred to a detected hand or object; columns 3 to 6 contained the four bounding box coordinates (i.e., the x and y coordinates of the top-left and bottom-right corners); column 7 contained the prediction's confidence score, which is a value between 0 and 1 that represents how confident the network was in detecting the hand or the object; column 8 reported the interaction state, which indicated whether the hand physically touched another object or hand (possible contact state values were 0 = no contact, 1 = self contact, 2 = contact with another person, 3 = contact with a portable object, 4 = contact with a fixed object); columns 9 to 11 reported the offset vector, which was defined as the offset between the hand and the object; and column 12 indicated the side of a detected hand (0 = left, 1 = right).
A visual representation of the bounding box, hand and object identification, and contact state is reported in Figure 11.

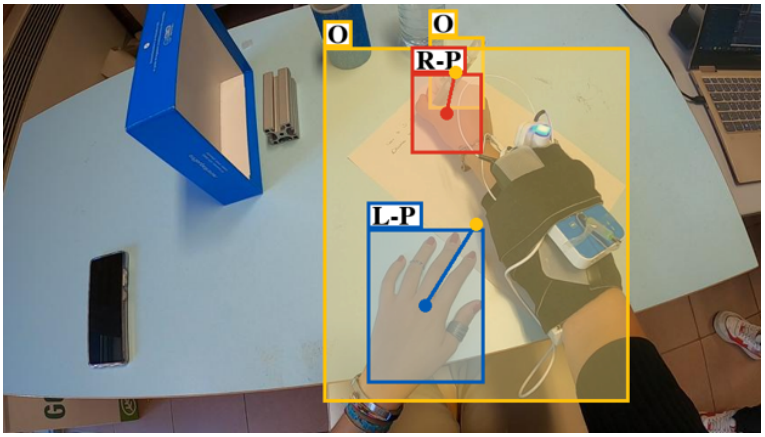

Figure 11: Hand-object interaction. The figure depicts the correct left and right-hand identification.

Just like the sEMG signal processing (see section 2.4.1), the inter-task pauses and the MVC portion were eliminated using the manually marked start-end frames of each task.
During the data collection, we ensured not to interfere with the subjects' task execution so that only the participant's hands were present within the camera view.
Due to the presence of at least two hands (participants' right and left hands) and objects, each frame resulted in more than one detection. A series of processing steps were required to obtain a maximum of two detections per frame (i.e., one for each hand). The first step involved removing the object detections, to focus only on detections relative to hands. The second step was filling in the missing frames, corresponding to moments when the hand was not detected. Except for the interaction state, which was given a value of "0" (i.e., non-functional interaction), all 12 columns were filled with "NaN" values. Filling in the missing frames with appropriate values made the resulting data set complete, with no temporal gaps, and ready for further processing.
Afterwards, hand detections that displayed a confidence value lower than 0.5 were removed.

Considering that each frame was supposed to have a maximum of one hand per side, strategies to remove duplicated hands (i.e., two or more hands with the same side, see Figure 12) had to be devised. A side correction step was implemented to solve this issue.

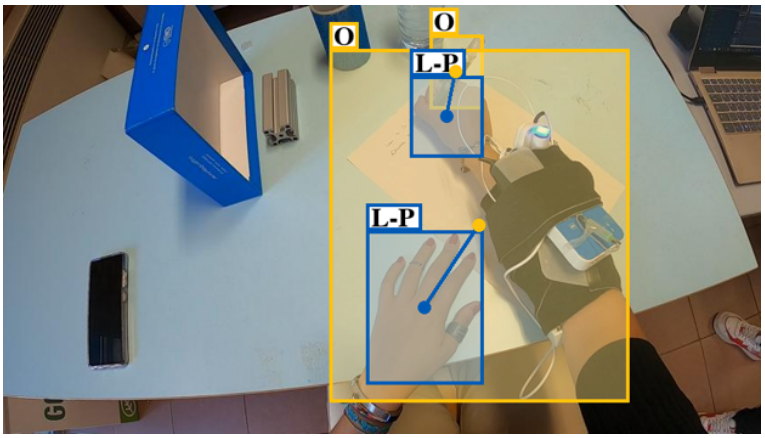

Figure 12: The figure is an example of duplicate detections.

Four criteria were developed and tested to choose the best approach for the side correction. The first three were:

- Criterion 1: In case of multiple hand detections with the same side, only the one with the highest confidence was retained;
- Criterion 2: Given that the participant's left (right) hand is more likely to appear more often in the left (right) half of the image, the retained left (right) hand was the one with the most-left (most-right) coordinates.
- Criterion 3: Among multiple detections for the same side, only the detected hand with the highest centroid was retained.

The best criterion was chosen by comparing the correct side with the ground truth side manually obtained on the 400 randomly selected frames. A different strategy (i.e., Criterion 4) was also explored in which several machine learning algorithms were trained to recognize the correct side based on input bounding boxes. Data from 70% of the patients were used for training and validating multiple machine learning algorithms, with the remaining 30% serving as a test-set. Specifically, we used: Support Vector Machines (SVMs), Random Forest (RF) and k-Nearest Neighbors (kNN)(see paragraph 2.4.3). Each was run twice using the ground truth relative to the side detection as the response variable: once using all four bounding box coordinates as predictors and once using just the x-coordinate as predictors.

Afterward, we focused on binarizing the interaction states following the strategy from Bandini et al. (2022) [30]. We selected as functional interactions (label = 1) frames were the predicted interaction state was either 3 or 4 (i.e., contact with a movable and fixed object, respectively), whereas frames with interaction state < 3 belonged to the non-functional interaction class (label = 0).

The interaction state output and the relative ground truth (ones and zeros arrays, see section 2.4.2) were processed similarly to the EMG signal (see section 2.4.1), using consecutive windows of 250 ms, 500 ms, 750 ms, and 1 s with a 50% overlap. The most frequent interaction state value was assigned to each frame of the window.

The results of the processing of FPV were expressed in a table format with the frames in the rows and the predicted interaction state and the relative ground truth in the columns.

Like EMG signal processing, four distinct datasets were collected, each corresponding to a different window length.

2.4.3 Detection of hand-object interactions

To obtain a balanced dataset with an equivalent amount of functional and non-functional interactions, the sEMG signal activity and egocentric video detections corresponding to non-functional tasks were removed (i.e, non-functional hand movements, walking and resting). The sEMG signal and FPV detections corresponding to the functional tasks already consisted of active and inactive periods.
We randomly chose seven participants from the dataset (representing 70% of the total data) for the training and validation set, and three subjects (30%) were set aside for the test set.
Once sEMG and FPV processing were concluded, we evaluated the classification performances in automatically identifying the presence of functional interactions from: (1) sEMG features alone; (2) FPV; and (3) sEMG and FPV approaches combined together.

Interaction detection from sEMG To identify the most important features and exclude any that were irrelevant or redundant ones, a feature selection was carried out on the training set to reduce the dimensionality of the problem. The Maximum Relevance Minimum Redundancy (MRMR) algorithm was considered. It ranks features sequentially to select a subset of highly relevant features to the target variable while minimizing redundancy among the selected features [78]. The algorithm examines the pairwise mutual information between all potential pairs of characteristics to determine redundancy [78]. If two features share a lot of information and have a high pairwise mutual information, they may be redundant. The algorithm considers the mutual information between each feature and the response variable to determine relevance, it measures how much information that feature provides about the response variable. This algorithm ranks feature significance using the mutual information quotient (MIQ) metric. The formula of MIQ is the following:

$$MIQ(x) = \frac{I(x,y)}{\frac{1}{n}\sum_{z\in\mathcal{S},z\neq x} I(x,z)}$$

where I(x, y) is the mutual information between feature x and the response variable y, I(x, z) is the mutual information between feature x and all other features z in the dataset, and n is the total number of features in the dataset S.
A feature is highly relevant and offers valuable information for predicting the response variable if it has high mutual information with the response variable (i.e., the interaction state ground truth) [78].
We considered the top 6, 12, and 23 most predictive features, representing 2.5%, 5%, and 10% of the total number of features.
Both training and testing sets were standardized using Z-score. Its formula is [79]:

$$Z = \frac{x - M}{SD}$$

Four classifiers were compared for detecting the functional interactions from sEMG features. These included support vector machines (SVM) with linear and RBF kernels [80], random forest (RF) [80] and K- Nearest Neighbor (kNN) [80][81].

1. SVM identifies the hyperplane that separates the data points into two classes while maximizing the distance between the hyperplane and the closest data points of both classes [81].
   It uses a mathematical function to transform data into a feature space with a high number of dimensions, allowing for the classification of data points that may not be linearly separable in their original form [82].
   (a) The RBF kernel is suitable when the data cannot be separated linearly and the relation between predictors and response value is nonlinear [83]. Its main hyperparameters are C and gamma, C is a parameter that controls the cost of misclassification in a model, while gamma affects the shape of the RBF kernel [84].
   (b) The linear kernel is the most basic kernel employed when the data can be partitioned by line, plane, or hyperplane [83]. Its main hyperparameter is C.
2. RF constructs multiple decision trees in parallel, trained on different subsets of the training data and each tree predicts a given input. The final prediction of the RF model is then obtained by combining the predictions of all the individual trees [80]. The settings that can be modified to fine-tune the performance of the model are the quantity of trees, the maximum number of characteristics to consider when dividing a node, and the number of tiers in each decision tree [85].
3. kNN uses existing data points in a dataset to predict new data points. The algorithm first creates a multidimensional space in which it stores all instances corresponding to the training data [81]. It then uses these instances to classify new data points according to their similarity to the training data using similarity indices like the Euclidean distance function. The value of 'k' in kNN represents the nearest neighbors considered when predicting a new data point [80][81]. K and the distance metric are the model's main hyperparameters.

The selected features of sEMG served as the predictors, while the ground truth of the interaction as the response variable.
Hyperparameter tuning was done by defining a list of values for the specific hyperparameters required by each machine learning algorithm. The model was then trained and evaluated for each combination of hyperparameters. The leave-one-subject-out cross-validation (LOSO-CV) method, was applied to the training validation set. It implies that one subject is excluded from each fold while the other subjects are used to train the model. We used metrics of accuracy and F1-score to determine the best-performing algorithm.
Once the hyperparameters that allowed for achieving the best performance for the training and validation had been obtained, the model with those hyperparameters was applied to the test set. Table 1 displayed all values of the main parameters used for each classification model.

Classifier	Parameter1	Parameter2	Parameter3
SVM	kernel: RBF, linear	C=[0.01, 0.1, 1, 10, 100]	gamma=[0.0001, 0.01, 0.1, 1, 10]
RF	trees=[10,20,50,100,200,500,1000]		
kNN	k=[3,5,7,9]	distance= Euclidean	

Table 1: Hyperparameters of the classifiers used in this work.

The sEMG classification performance was evaluated using the following metrics: accuracy, F1-score, precision, recall, and specificity [86].

- Accuracy: measures how often, out of all predictions produced, the model accurately predicts the correct class (i.e., the true positive and true negative predictions).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{22}$$

- F1-score: measures the overall performance of a classification model, taking into account both precision and recall.

$$F1score = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN} \tag{23}$$

- Precision: measures how often, out of all positive predictions, the model accurately predicts the positive class.

$$Precision = \frac{TP}{TP + FP} \tag{24}$$

- Recall: measures how often, out of all the actual positive cases in the data, the model correctly predicts the positive class.

$$Recall = \frac{TP}{TP + FN} \tag{25}$$

- Specificity: measures how well the model can correctly identify the negative class.

$$Specificity = \frac{TN}{TN + FP} \tag{26}$$

where TP stands for true positive, TN for true negative, FP for false positive and TN for false negative.

Interaction detection from FPV By comparing the interaction state of the test set with the manually identified ground truth, we were able to obtain the performance of the FPV in detecting hand-object interactions, evaluated with accuracy, F1-score, precision, recall, and specificity metrics.

Interaction detection by combining sEMG and FPV After obtaining the individual performances of the EMG and FPV modalities, we calculated the performance of the multimodal approach combining the two modalities.
Specifically, we tested three different combinations:

1. FPV + EMG: we concatenated the interaction state from the FPV analysis with the most significant features extracted from the sEMG signal (which served as predicotrs), creating a new dataset. This dataset was split into training and test sets, following the same procedure as the individual modalities. We trained the model that performed best in the single modalities. This approach aimed to explore the possibility that the two modalities could complement each other and provide better overall performance.
2. AND logical operator: The AND operator was used to combine the interaction state predicted by the sEMG signal and that derived from the FPV analysis. The AND operator returns true only if both operands are true, meaning that it only predicted a functional interaction when both modalities indicated that there was one (Table 2).

EMG	Egocentric	AND
1	1	1
1	0	0
0	1	0
0	0	0

Table 2: AND logical operator

3. OR logical operator: The OR operator was also tested, which returns true if at least one of the operands is true. In this case, if either the EMG or FPV modality predicted a functional interaction, the OR operator predicted that there was one (Table 3).

EMG	Egocentric	OR
1	1	1
1	0	1
0	1	1
0	0	0

Table 3: OR logical operator

The interaction state ground truth used to calculate the performance for all three combinations. The performance of each combination was evaluated with accuracy, F1-score, precision, recall, and specificity.

Performance comparison (single mode vs multimodal) The analysis described in section 2.4 was repeated 12 times, each time with a different dataset split. This allowed us to evaluate the robustness of each approach to variations in the data across subjects.
The non-parametric Kruskal-Wallis test, which is the non-parametric version of the one-way ANOVA test for comparing multiple groups [87], was used to compare the classification performance obtained with single and multimodal techniques. If the obtained p-value was lower than 0.05, implying that group medians were significantly different, a post-hoc analysis was performed. In particular, the multiple comparison test (i.e., Dunn's test) with the Bonferroni correction was carried out to identify which approaches were significantly better than others. The statistical analysis was conducted twice:

1. In the first analysis, three techniques for combining EMG and FPV (FPV + EMG, FPV AND EMG, FPV OR EMG) were compared to determine whether there was a significant difference between them, and if so, to identify the best performing one.
2. The second analysis compared the most accurate combination method identified in the first analysis with the two single modal approaches. The objective of this comparison was to determine whether there were any significant differences between the approaches and to identify the one with the best performance.

For every analysis conducted, five distinct results were generated, with each result corresponding to one of the following five evaluation metrics: accuracy, F1-score, precision, recall, and specificity.

3. Results

3.1. sEMG Results

3.1.1 PCA

PCA was used to reduce the dimensionality of the sEMG data. Only the first 11 PCs were considered, as, on average, they covered over 90% of the variance, and because there are approximately 11 detectable superficial forearm muscles [68]. We determined the percentage of variance explained by the first 11 components by calculating and averaging the results across ten subject acquisitions for both the right and left hand. The average percentage of variance explained by these components was 92.2%. Figure 13 depicts the percentage of variance covered by the first 11 PC for one subject.

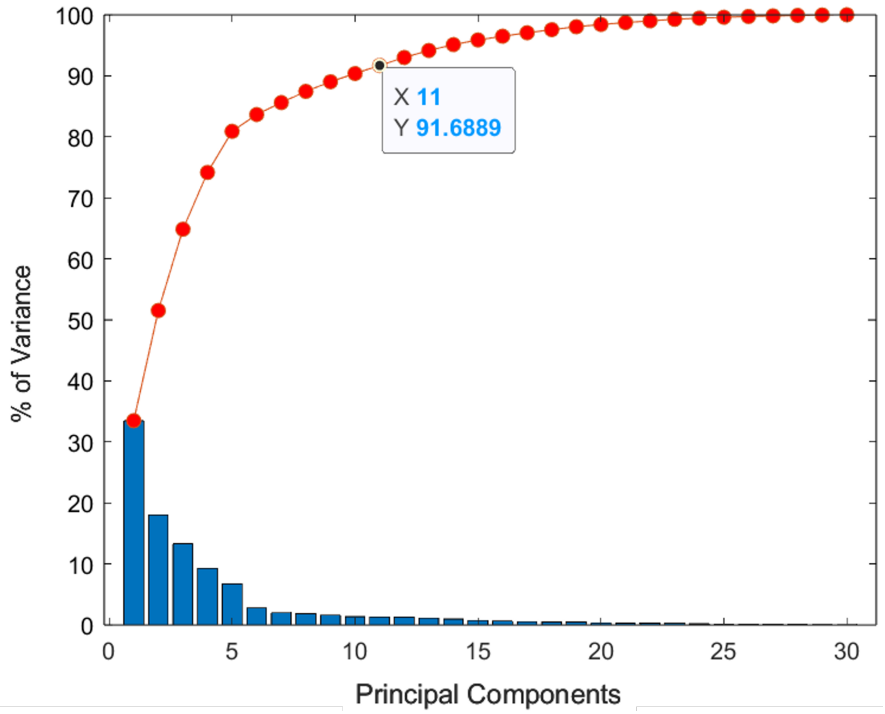

Figure 13: Barplot that depicts the percentage of variance: on average 92% of the variance was covered by the first 11 PCA. The bars represent the percentage of variance for each PC and are arranged in decreasing order. The red line represents the cumulative distribution of the PC.

3.1.2 Feature selection

To identify the key sEMG features while removing any irrelevant or redundant ones, we conducted a feature selection. The MRMR algorithm was a reliable option for feature selection because it ranked a subset of characteristics pertinent to the target variable and uncorrelated with one another.
Figure 14 represents the first 23 selected features for window lengths of 1 second.

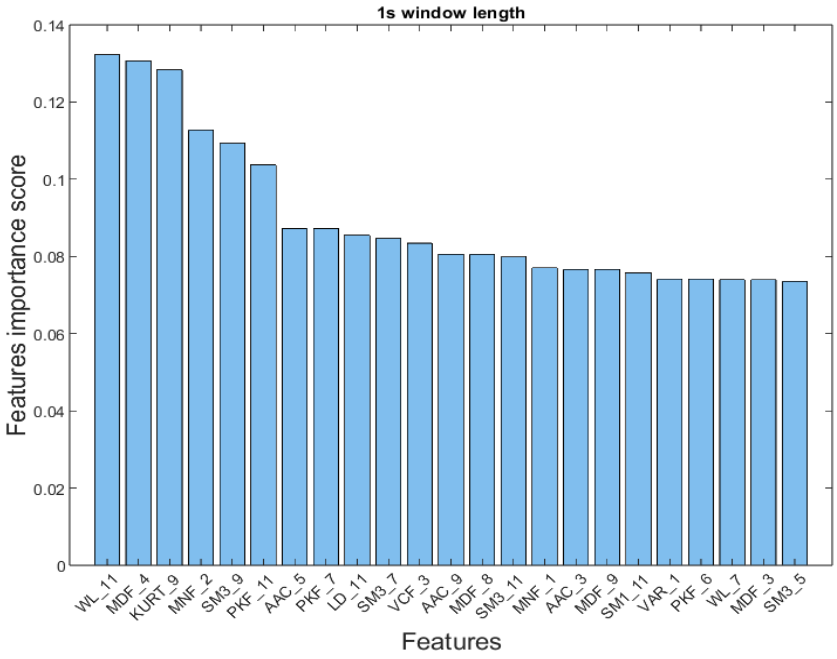

Figure 14: The 23 most relevant features are displayed for window length of 1 second. The vertical axis shows the predictor's important score, specifically the mutual information quotient (MIQ). The horizontal axis lists the selected features from the most to the least significant. Overall, the time domain and frequency attributes selected across all four window lengths were: Average amplitude change (AAC), Log detector (LD), Simple Square Integral (SSI), Variance (VAR), Waveform Length (WL), Coefficient of Variation (COV), Kurtosis (KURT), Skewness (SKEW), Mean frequency (MNF), Median frequency (MDF), Variance of central frequency (VCF), Peak frequency (PKF), First Spectral Moment (SM1), Second Spectral Moment (SM2), Third Spectral Moment (SM3). The number after the acronym of the features represents the PC number (1 to 11).

3.1.3 sEMG performance

Training and validation The classification results are shown in figure 15. We concluded that the support vector machine with the RBF kernel and with the top six sEMG features (ranked according to MRMR) produced the best performance across all four window lengths.
Notably, Figure 15 demonstrates that increasing the window size did not significantly improve model performance.

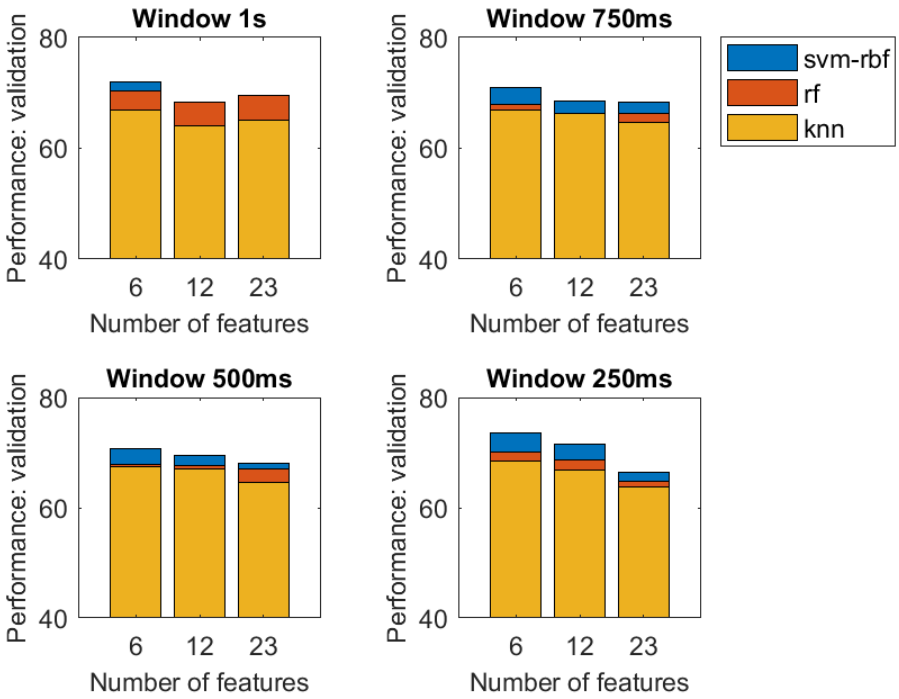

Figure 15: Performance EMG with 6, 12, and 23 features for each window size: the blue rectangles represent the performance of the SVM model with RBF kernel, the orange one the RF and the yellow one the kNN.

Additionally, we conducted training and validation on the SVM model with a linear kernel using a window length of 1 second. However, the model exhibited poor performance and required a considerable amount of run time. As a result, we decided against using this model for the other window lengths.
Table 4 depicts classifier and the relative training and validation performance. The SVM classifier with linear kernel clearly underperformed.

Features	Classifier	Accuracy	F1score
6	**SVM (RBF kernel)**	**0.685**	**0.720**
6	SVM (linear kernel)	0.616	0.587
6	RF	0.679	0.703
6	kNN	0.650	0.669
12	SVM (RBF kernel)	0.666	0.683
12	SVM (linear kernel)	0.601	0.584
12	RF	0.661	0.683
12	kNN	0.618	0.640
23	SVM (RBF kernel)	0.667	0.681
23	SVM (linear kernel)	0.608	0.590
23	RF	0.676	0.696
23	kNN	0.625	0.650

Table 4: EMG training and validation performance on window length of 1 second. The best classification performances are highlighted in bold.

Testing The five metrics that were used to assess sEMG classification performance of the test set on the best model (SVM with RBF kernel, 6 features and 1-second window length) are listed in Table 5.

Accuracy	F1score	Precision	Recall	Specificity
0.660	0.708	0.626	0.814	0.501

Table 5: sEMG classification performance

The overall performance of the approach was promising, with favorable results across various metrics. Notably, the F1-score and recall were particularly strong. However, the approach demonstrated lower specificity compared to other metrics.

We created a confusion matrix, illustrated in Figure 16, to assess the sEMG's performance further. The matrix revealed that the system identified many true positives and true negatives, but also produced several false positives. Its limitation, however, was its potential to incorrectly classify positive cases as negative (FN).

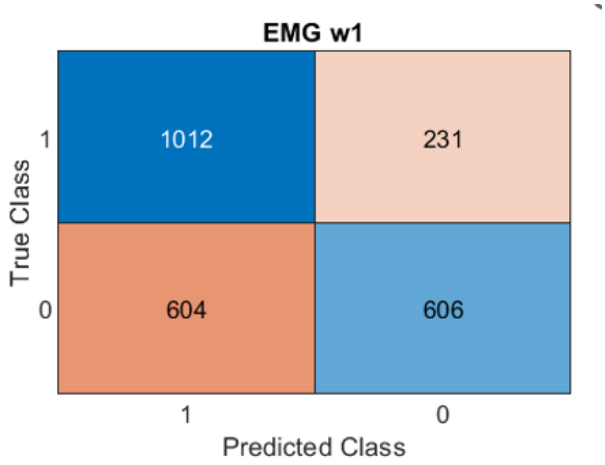

Figure 16: sEMG confusion matrix obtained with the 1-second window length. Label = 1 represents a functional interaction while label = 0 represents a non-functional one.

3.2. FPV results

3.2.1 Side correction

Each participant's video was 8 minutes and 30 seconds long on average, and roughly 15,300 total frames were recorded.
One fundamental step in processing the FPV recording was the filtering of duplicate hands (i.e., with the same side, see section 2.4.2). Table 6 depicts the accuracy of each criterion described in section 2.4.2.

Criterion	Features	Classifier	Accuracy
Criterion 1	/	/	0.833
Criterion 2	/	/	**0.907**
Criterion 3	/	/	**0.923**
Criterion 4	2	SVM (RBF kernel)	0.722
Criterion 4	2	SVM (linear kernel)	0.707
Criterion 4	2	RF	0.707
Criterion 4	2	kNN	0.707
Criterion 4	4	SVM (RBF kernel)	0.692
Criterion 4	4	SVM (linear kernel)	0.714
Criterion 4	4	RF	0.669
Criterion 4	4	kNN	0.654

Table 6: Comparison between the criteria used for hand side correction. Criterion 1 = retaining the duplicated detection with the highest confidence score; Criterion 2= assigning "right" ("left") to the right-most (left-most) detection, Criterion 3 = retaining the duplicated detection with the uppermost coordinates and Criterion 4 = implementing machine learning algorithms that were trained to recognize the correct side based on the coordinates of the hand bounding boxes. The best results are highlighted in bold.

Criterion 2 (i.e., assigning "right" ("left") to the right-most (left-most) detection) and 3 (i.e., retaining the duplicated detection with the uppermost coordinates) performed the best, with Criterion 3 having the highest accuracy but being less generalizable. Our acquisitions were recorded following a standardized protocol in which participants were asked to use and record the hand under analysis, keeping the other one lower in the frame. However, subjects use both hands during at-home recording without following such instructions. Therefore, we opted for criterion 2, which had an accuracy value of 0.91 and was more suitable for daily life activity recording.

Criterion 4 involved training and testing different classification algorithms. However, the accuracy value of each model was lower than that obtained using the CNN algorithm (0.750). Therefore, we proceeded with the side correction using Criterion 2.

3.2.2 FPV performance

Table 7 summarizes all five metrics that describe FPV performance in detecting hand-object interactions.

Window length	Accuracy	F1score	Precision	Recall	Specificity
1 s	0.528	0.682	0.518	0.997	0.047
750 ms	0.529	0.680	0.517	0.995	0.057
500 ms	0.534	0.681	0.519	0.991	0.074
250 ms	0.537	0.680	0.519	0.987	0.086

Table 7: FPV performance in detecting hand-object interactions

Similar to the sEMG performance (see section 3.1.3), based on the F1-score, the one that performed the best was the 1-second window length.
Based on the data presented in Table 7, the FPV exhibited high recall values, but had a significant limitation in terms of specificity, resulting in a large number of false positives and few true negatives. Furthermore, the precision values were also relatively low.
The confusion matrix displayed in Figure 17 provides further insight into the performance of the FPV approach.

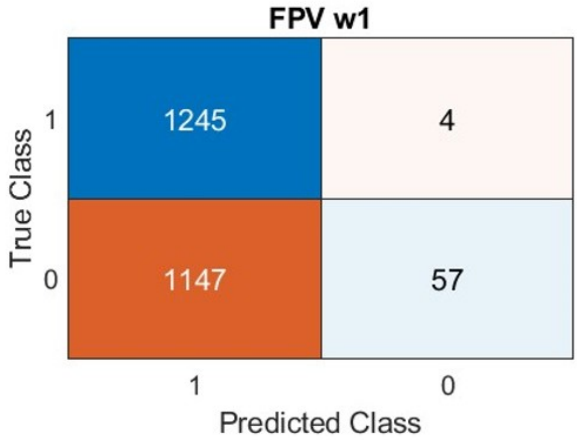

Figure 17: FPV confusion matrix obtained with the 1-second window length. Label = 1 represents a functional interaction while Label = 0 represents a non-functional one.

Indeed, while there are very few FN, the number of false positives FP is considerably high, and the count of true negatives TN is notably low.

3.3. Combination results

This section presents the results of the two separate statistical analyses conducted on all 12 data set split combinations. The first analysis compared the combination techniques, while the second analysis compared the performance of the best multi-modal approach with that of the single modalities.

3.3.1 Comparison of combination techniques

The median values, interquartile range, Kruskal-Wallis test p-value and post-hoc test p-value for each evaluation metric are depicted in Table 8.
In terms of accuracy, precision, and specificity, significant differences were observed across the three multi-modal approaches. Specifically, the "FPV OR EMG" combination performed significantly worse (post-hoc test p-value < 0.01) than both "FPV + EMG" and "FPV AND EMG", while "FPV + EMG" and "FPV AND EMG" had similar results (post-hoc test p-value > 0.05). However, in terms of recall, "FPV OR EMG" outperformed "FPV + EMG" and "FPV AND EMG", which once again showed no significant difference. There were no significant differences among the three groups for the F1-score metric.

Metric	Median (interquartile range)			Kruskal-Wallis p-value	Post-hoc p-value		
	FPV + EMG	FPV AND EMG	FPV OR EMG		A	B	C
Accuracy	0.716 (0.067)	0.719 (0.080)	0.598 (0.043)	<0.01	1	<0.01	<0.01
F1score	0.743 (0.061)	0.723 (0.092)	0.717 (0.030)	0.055	/	/	/
Precision	0.712 (0.075)	0.722 (0.059)	0.560 (0.036)	<0.01	1	0.01	<0.01
Recall	0.849 (0.118)	0.771 (0.124)	0.998 (0.004)	<0.01	0.525	0.01	<0.01
Specificity	0.631 (0.143)	0.665 (0.133)	0.165 (0.069)	<0.01	0.969	<0.01	<0.01

Table 8: Statistical analysis on three combination options. A= FPV + EMG compared to FPV AND EMG; B= FPV + EMG compared to B=FPV OR EMG; C= FPV AND EMG compared to FPV OR EMG. Significant difference p-value=0.05.

3.3.2 Single-modal vs multimodal approach

For our second statistical analysis, we focused on comparing the "FPV + EMG" combination to the single modality options. We chose this combination as it showed a lower interquartile range for both accuracy and F1 score, indicating less variability in the data. Table 9 presents the results.
Based on the Kruskal-Wallis test, we saw that there were significant differences between the three approaches for all five metrics.
The analysis found that the combined approach of using both FPV and EMG data (FPV+EMG) performed significantly better than using only FPV or EMG data in predicting functional and non-functional hand-object interactions. This was demonstrated by the higher scores in accuracy, precision, and specificity metrics. For recall, the FPV approach displayed high scores, which was anticipated in the findings in section 7. The FPV performed better than the other two approaches for this particular metric.
In terms of F1-score, both FPV and FPV+EMG had very similar scores, and no significant difference was found between these approaches. On the other hand, the F1-score value for the multi-modal approach was significantly higher than that of the EMG approach.

Metric	Median (interquartile range)			Kruskal-Wallis p-value	Post-hoc p-value		
	FPV	EMG	FPV+EMG		D	E	F
Accuracy	0.650 (0.067)	0.653 (0.044)	0.716 (0.067)	0.003	1	0.005	0.016
F1score	0.745 (0.037)	0.698 (0.068)	0.743 (0.061)	0.017	0.082	1	0.023
Precision	0.598 (0.048)	0.635(0.036)	0.712 (0.075)	<0.01	0.114	<0.01	0.173
Recall	0.990 (0.007)	0.776 (0.125)	0.849 (0.118)	<0.01	<0.01	<0.01	1
Specificity	0.297 (0.146)	0.520 (0.081)	0.631 (0.143)	<0.01	0.002	<0.01	0.913

Table 9: Statistical analysis single-modal vs multi-modal approach: D=FPV compare to EMG; E=FPV compare to the multimodal approach FPV+EMG; F=EMG compare to the multimodal approach FPV+EMG. Significant difference p-value=0.05

Overall, the multi-modal strategy consistently outperformed the two single-modal approaches in detecting functional and non-functional interactions between the hand and object, as demonstrated by the majority of the evaluation metrics.

4. Discussion

We introduced a novel multi-modal approach that integrates FPV and sEMG to automatically detect hand-object interactions. This approach has the potential to be used for conducting assessments of the upper extremities in individuals with neurological impairments affecting the upper limbs.
Our primary objective was to validate the system by testing it on healthy participants during standardized tasks in a controlled laboratory setting.
Firstly, we analyzed the performance of the sEMG approach. As stated in section 2.4.1, its performance was evaluated using four different classifiers: sSVM with both linear and RBF kernels, RF, and kNN. The results showed that the SVM with the RBF kernel and the first six significant features (WL, MDF, KURT, MNF, SM3, PFK) best predicted functional and non-functional interactions.
The selection of these specific features is attributed to their ability to capture different aspects of the EMG signal. For example, WL, MDF, MNF and PKF can provide information about the frequency content of the EMG signal, which can be used to identify changes in muscle activation and endurance during different hand-object interactions [71]. KURT and SM3 can provide information about muscle co-contraction and asymmetry in the EMG signal, which may be relevant for distinguishing between functional and non-functional hand-object interactions [71].
Because the SVM model with a linear kernel can only separate the data points using a straight line or a hyperplane [83], it may not be able to adequately reflect the complexity of the data, which may explain why it did poorly. The SVM model with the RBF kernel, on the other hand, can capture the non-linear correlations between the data points with more intricate decision limits, potentially improving performance with the available data [83].
Among four different window lengths, the 1-second window length achieved the highest accuracy and F1-score. sEMG alone showed promising results in predicting functional interactions, as shown by the high recall and precision values. However, the system had lower specificity, which indicates its lower ability to identify non-functional interactions correctly.
These findings suggest that the sEMG has the potential to be a helpful tool for identifying functional and non-functional interactions, even though there is room for improvement in reducing FPs and FNs.
The second single approach we analyzed was the FPV. One crucial aspect of FPV processing involved side correction, which aimed to accurately identify the correct side (i.e. left or right hand) that the CNN algorithm had misclassified. Four criteria were tested: Criterion 2 (i.e., assigning "right" to the one with the coordinates furthest to the right and "left" to the one with the coordinates furthest to the left) and Criterion 3 (i.e, retaining the detection with the highest centroid) were found to have the best performance. Ultimately, Criterion 2, which assigned "right" to the hand with coordinates furthest to the right and "left" to the hand with coordinates furthest to the left, was selected due to its high accuracy, as well as its suitability for daily life activity recording.
Using different window lengths, the FPV approach was tested to recognize functional interactions. By comparing the interaction state of the test set with the manually identified ground truth, we were able to assess the accuracy, F1-score, recall, precision, and specificity of FPV. More specifically we found that the 1-second time frame was the most effective in terms of F1-score, similar to how sEMG performed. FPV had a high recall score, indicating that it recognized most functional interactions, but it had very low specificity values, meaning it struggled to identify non-functional interactions, resulting in many false positives and few true negatives.
Overall, these findings imply that while FPV is capable of accurately identifying functional interactions, however, additional advancements are required to increase accuracy, precision, and specificity.
Three different options for combinations of FPV and EMG were tested. The first (FPV + EMG) consisted of simply concatenating the most significant features from the EMG analysis and the contact state from the FPV one. The second approach involved merging the interaction states predicted by the EMG and FPV analysis using the AND logical operator (FPV AND EMG), while the third approach utilized the OR logical operator (FPV OR EMG).
Only the dataset related to the window length of one second was used as it delivered the best performance in both single-modal approaches. We were interested in detecting hand-object interactions. These occur over a span of a few seconds rather than milliseconds, which may explain why the longer window length of 1 second performed better.
To compare the five approaches and determine whether there were any statistically significant differences we utilized the non-parametric Kruskal-Wallis test. To determine which showed significant differences, we next performed a post-hoc analysis. First, we analyzed the three techniques for combining EMG and FPV. Second, we compared the most accurate technique from the first analysis with the two single-modal approaches.
From the first analysis, it was observed that the "FPV + EMG" and "FPV AND EMG" combinations had promising and comparable performance. Even though the "FPV OR EMG" combination was successful in identifying almost all hand-object interactions, as demonstrated by the very high recall, it had the poorest performance across every other metric.

In the second statistical analysis, the single-modal approaches were compared to the "FPV + EMG" combination, selected based on its lower interquartile range. The results showed significant differences between the three approaches for all five metrics. The "FPV+EMG" combination outperformed the two single-modal approaches in terms of accuracy, precision, and specificity, implying its effectiveness in accurately predicting functional and non-functional hand-object interactions. In terms of recall the FPV performed the best, as demonstrated by the very low number of FN. However, the multi-modal approach also exhibited a high recall, demonstrating once again its ability in identifying functional interaction.
In summary, this analysis found that different evaluation metrics showed varying performances for each approach. Overall, these findings suggest that the combination of FPV technology and sEMG analysis is an effective approach for capturing functional and non-functional interactions, minimizing errors related to interaction identification.
Previous studies ([30] [59] [60]) have investigated the feasibility and validity of using FPV technology to measure hand use and function in individuals with SCI and stroke survivors. For example, Likitlersuang J. et al. (2019) [59] and Bandini et al. (2022) [30] found FPV technology to be a feasible and valid tool for capturing hand use and function in individuals with SCI. Similarly, a study by Tsai and colleagues (2021) investigated the validity of using FPV technology to record hand use and function in stroke survivors [60]). All three studies suggest that the technology accurately captured hand movements and identified functional hand-object interactions, and the use of FPV technology combined with computer algorithms is an effective method for analyzing hand use and hand roles in individuals with these conditions during daily activities. However, they also mentioned several limitations including small sample sizes, limited diversity, lack of control groups, and a focus on a limited range of hand function tasks. Additionally, the computer algorithm may not accurately classify hand use and hand roles in certain situations: FPV fails to discriminate interactions in presence of "non-standard" grasps and compensatory strategies. Nevertheless, FPV was found to be more effective than TPV for upper limb (UL) rehabilitation. This is because FPV provides a more immersive and engaging experience, allowing the user to feel as if they are actually performing the movements themselves [58].
sEMG is a widely used technology for assessing muscle activity and designing rehabilitation interventions for patients with upper-limb impairments [42][43][44][49][53][54][55]. For instance, Lee et al.(2011) developed a classification system to determine the intended manual task based on the sEMG data of stroke survivors [45]. The system was able to differentiate between open-hand and grip tasks, however, accuracy was lower for similar grip tasks, particularly for severely impaired subjects.
Wearable sensors, such as IMUs, can also provide valuable information on upper-limb kinematics and movement patterns after stroke as proven by [35][34][36][39]. Overall, the application of IMUs in rehabilitation demonstrated promising results in terms of delivering precise and objective measurements, customizing therapies, and enhancing patient outcomes [33]. However, these methods might fail to capture complex hand and finger movements [35]. Unlike our proposed multimodal approach, IMUs are limited to measuring kinematics and do not provide information on other important aspects of upper-limb function, such as muscle activity or hand-object interactions[33][37].
On the other hand, force myography has been explored as a means to distinguish between grasping and non-grasping movements during daily activities [38]. For instance, Sadarangani et al. (2017) demonstrated that FMG achieved a high average grasp detection accuracy across all tasks for both stroke patients and healthy participants [37]. However, FMG has several limitations, as it cannot capture dynamic changes in force during movement or the force generated by individual muscles. In addition, it is only capable of measuring the force generated during gross motor movements and not the fine ones [38].

4.1. Limitations and future work

The study was limited by a small sample size (10 healthy individuals) which reduced the generalizability of the results. Additionally, it only tested the multi-modal approach on healthy subjects, and it still needs to be determined how well it performs with individuals with hand impairments, for example, stroke survivors and individuals with SCI. The acquisitions were conducted in a laboratory setting, which may only partially reflect the real-world experience of individuals with hand impairments. Moreover, the study focused on analyzing hand use during specific and standard task execution, which may only capture part of the range of hand use in individuals. In addition, the placement of the sEMG array of the subject hand required, in some cases, several attempts. This issue prolonged the acquisition time, which was already limited due to the short battery life of both the OTB Sessantaquattro device and GoPro Hero 8. Finally, a significant amount of manual labeling was required to establish the ground truth regarding the contact state, hand side, and bounding box. It involved scrutinizing each video frame, which was time-consuming and mentally taxing.
Although this thesis offers insightful information, more investigation is required to address some of its limitations. For instance, expanding the sample size would provide access to a population that is more representative and diverse, potentially exposing new insights. In addition, further research should consider a wider range of

upper limb movements, focusing especially on everyday activities. To further validate this approach, testing it on individuals who have experienced a stroke or SCI could provide valuable information about its potential clinical applications. Additionally, evaluating the approach's performance outside of the clinical setting's tightly controlled environment could entail examining hand use during routine activities at home. Such validation could help overcome the limited access to rehabilitation services that complicate the accurate tracking of neurological and functional recovery. It could provide more objective and reliable and offer effective exercises tailored to the patient's needs and progress. To increase the duration of the recording sessions, battery limits and usability need also be addressed. To increase usability, it is essential to optimize the design of the wearable sleeve by making it more comfortable, appropriate, and user-friendly. An excellent example is the Myo armband, which has an adjustable and compact design that can contribute to a more user-friendly acquisition process. Finally, the distribution of the labeling task among a large number of researchers may help speed up the process of ground truth extraction.

5. Conclusion

We demonstrated that the combination of FPV and sEMG is an effective method for automatically capturing both functional and non-functional hand-object interactions in healthy individuals. This is important because understanding how the hand interacts with objects provides valuable insights into functional hand use. From a clinical perspective, monitoring the functional use of upper extremities in real-life settings is crucial to tailor therapy and maximize treatment outcomes.

Specifically, we have established that the multi-modal approach consistently outperformed the two single-modal approaches. The sEMG single-modal approach shows promising results for both functional and non-functional interactions, although there is room for improvement in reducing false positives and negatives. Our results demonstrated that while the FPV single-modal approach accurately identifies most functional interactions, it does not perform as well in detecting non-functional interactions, as indicated by the low specificity. Furthermore, our study highlights that the choice of FPV and EMG combination significantly impacts the model's performance. The "FPV + EMG" and "FPV AND EMG" combinations showed comparable outcomes, while the "FPV OR EMG" combination performed worse. Overall, the combination of FPV and EMG resulted in positive outcomes for all five evaluation parameters.

Future research will involve validating our findings in subjects with SCI or stroke, both in a clinical setting and in their homes.

6. Bibliography and citations

References

[1] WHO. World Stroke Day 2022. https://www.who.int/srilanka/news/detail/29-10-2022-world-stroke-day-2022. Accessed: 07/04/2023.

[2] Valery L. Feigin, Michael Brainin, Bo Norrving, Sheila Martins, Ralph L. Sacco, Werner Hacke, Marc Fisher, Jeyaraj Pandian, and Patrice Lindsay. World Stroke Organization (WSO): Global Stroke Fact Sheet 2022, 1 2022.

[3] NHS. Stroke Recovery. https://www.nhs.uk/conditions/stroke/recovery/. Accessed: 07/04/2023.

[4] Samar M. Hatem, Geoffroy Saussez, Margaux della Faille, Vincent Prist, Xue Zhang, Delphine Dispa, and Yannick Bleyenheuft. Rehabilitation of motor function after stroke: A multiple systematic review focused on techniques to stimulate upper extremity recovery. *Frontiers in Human Neuroscience*, 10(SEP2016), 9 2016.

[5] Enas S Lawrence, Catherine Coshall, Ruth Dundas, Judy Stewart, Anthony G Rudd, Robin Howard, ; Charles, and D A Wolfe. Estimates of the Prevalence of Acute Stroke Impairments and Disability in a Multiethnic Population. Technical report, 2001.

[6] American Stroke Association. https://www.stroke.org/en/. Accessed: 29/03/2023.

[7] Randolph J. Nudo. Postinfarct cortical plasticity and behavioral recovery. In *Stroke*, volume 38, pages 840–845, 2 2007.

[8] Kevin B. Wilkins, Meriel Owen, Carson Ingo, Carolina Carmona, Julius P.A. Dewald, and Jun Yao. Neural plasticity in moderate to severe chronic stroke following a device-assisted task-specific arm/hand intervention. *Frontiers in Neurology*, 8(JUN), 6 2017.

[9] Steven C. Cramer. Repairing the human brain after stroke: I. Mechanisms of spontaneous recovery, 3 2008.

[10] Christian Grefkes and Gereon R. Fink. Recovery from stroke: current concepts and future perspectives. *Neurological Research and Practice*, 2(1), 12 2020.

[11] Nick S. Ward. Restoring brain function after stroke — bridging the gap between animals and humans, 3 2017.

[12] WHO. Spinal cord injury. https://www.who.int/news-room/fact-sheets/detail/spinal-cord-injury. Accessed: 04/04/2023.

[13] Christopher S. Ahuja, Jefferson R. Wilson, Satoshi Nori, Mark R.N. Kotter, Claudia Druschel, Armin Curt, and Michael G. Fehlings. Traumatic spinal cord injury, 4 2017.

[14] Steven C. Kirshblum, William Waring, Fin Biering-Sorensen, Stephen P. Burns, Mark Johansen, Mary Schmidt-Read, William Donovan, Daniel Graves, Amit Jha, Linda Jones, M. J. Mulcahey, and Andrei Krassioukov. Reference for the 2011 revision of the International Standards for Neurological Classification of Spinal Cord Injury, 2011.

[15] Jennifer A. Piatt, Shinichi Nagata, Melissa Zahl, Jing Li, and Jeffrey P. Rosenbluth. Problematic secondary health conditions among adults with spinal cord injury and its impact on social participation and daily life. *Journal of Spinal Cord Medicine*, 39(6):693–698, 11 2016.

[16] Mir Hojjat Khorasanizadeh, Mahmoud Yousefifard, Mahsa Eskian, Yi Lu, Maryam Chalangari, James S. Harrop, Seyed Behnam Jazayeri, Simin Seyedpour, Behzad Khodaei, Mostafa Hosseini, and Vafa Rahimi-Movaghar. Neurological recovery following traumatic spinal cord injury: A systematic review and meta-analysis, 5 2019.

[17] J. W. Fawcett, A. Curt, J. D. Steeves, W. P. Coleman, M. H. Tuszynski, D. Lammertse, P. F. Bartlett, A. R. Blight, V. Dietz, J. Ditunno, B. H. Dobkin, L. A. Havton, P. H. Ellaway, M. G. Fehlings, A. Privat, R. Grossman, J. D. Guest, N. Kleitman, M. Nakamura, M. Gaviria, and D. Short. Guidelines for the conduct of clinical trials for spinal cord injury as developed by the ICCP panel: Spontaneous recovery after spinal cord injury and statistical power needed for therapeutic clinical trials, 3 2007.

[18] I Laffont, E Briand, O Dizien, M Combeaud, B Bussel, M Revol, and A Roby-Brami. Kinematics of prehension and pointing movements in C6 quadriplegic patients. Technical report, 2000.

[19] Julie Bernhardt, Kathryn S. Hayward, Gert Kwakkel, Nick S. Ward, Steven L. Wolf, Karen Borschmann, John W. Krakauer, Lara A. Boyd, S. Thomas Carmichael, Dale Corbett, and Steven C. Cramer. Agreed Definitions and a Shared Vision for New Standards in Stroke Recovery Research: The Stroke Recovery and Rehabilitation Roundtable Taskforce. *Neurorehabilitation and Neural Repair*, 31(9):793–799, 9 2017.

[20] Ryan J. Visee, Jirapat Likitlersuang, and Jose Zariffa. An Effective and Efficient Method for Detecting Hands in Egocentric Videos for Rehabilitation Applications. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(3):748–755, 3 2020.

[21] Andrea Bandini and Jose Zariffa. Analysis of the hands in egocentric vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 4 2020.

[22] Thomas Brott, Harold P Adams, Charles P Olinger, John R Marler, William G Barsan, Jose Biller, Judith Spilker, Renee Holleran, Robert Eberle, Vicki Hertzberg, Marvin Rorick, Charles J Moomaw, and Michael Walker. Measurements of Acute Cerebral Infarction: A Clinical Examination Scale. Technical report, 1989.

[23] Scott E Kasner. Clinical interpretation and use of stroke scales. Technical report, 2006.

[24] Strokengine. Fugl-Meyer Assessment of Sensorimotor Recovery After Stroke (FMA).

[25] Strokengine. Action Research Arm Test (ARAT). https://strokengine.ca/en/assessments/action-research-arm-test-arat/. Accessed: 10/04/2023.

[26] Stephen Ashford, Mike Slade, Fabienne Malaprade, and Lynne Turner-Stokes. Evaluation of functional outcome measures for the hemiparetic upper limb: A systematic review. *Journal of Rehabilitation Medicine*, 40(10):787–795, 11 2008.

[27] Christian Schuld, Steffen Franz, Karin Brüggemann, Laura Heutehaus, Norbert Weidner, Steven C. Kirshblum, Rüdiger Rupp, and on behalf of the EMSCI study group. International standards for neurological classification of spinal cord injury: impact of the revised worksheet (revision 02/13) on classification performance. *Journal of Spinal Cord Medicine*, 39(5):504–512, 9 2016.

[28] Inge Marie Velstra, Marc Bolliger, Jörg Krebs, Johan S. Rietman, and Armin Curt. Predictive value of upper limb muscles and grasp patterns on functional outcome in cervical spinal cord injury. *Neurorehabilitation and Neural Repair*, 30(4):295–306, 5 2016.

[29] Jörg Wissel, John Olver, and Katharina Stibrant Sunnerhagen. Navigating the poststroke continuum of care, 1 2013.

[30] Andrea Bandini, Mehdy Dousty, Sander L. Hitzig, B. Catharine Craven, Sukhvinder Kalsi-Ryan, and José Zariffa. Measuring Hand Use in the Home after Cervical Spinal Cord Injury Using Egocentric Video. *Journal of neurotrauma*, 39(23-24):1697–1707, 12 2022.

[31] Olivier Lambercy, Ludovic Dovat, Roger Gassert, Etienne Burdet, Chee Leong Teo, and Theodore Milner. A haptic knob for rehabilitation of hand function. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(3):356–366, 9 2007.

[32] J. H. Crosbie, S. Lennon, J. R. Basford, and S. M. McDonough. Virtual reality in stroke rehabilitation: Still more virtual than real. *Disability and Rehabilitation*, 29(14):1139–1146, 2007.

[33] Pablo MacEira-Elvira, Traian Popa, Anne Christine Schmid, and Friedhelm C. Hummel. Wearable technology in stroke rehabilitation: Towards improved diagnosis and treatment of upper-limb motor impairment, 11 2019.

[34] Juan Pablo Gomez-Arrunategui, Janice J. Eng, and Antony J. Hodgson. Monitoring Arm Movements Post-Stroke for Applications in Rehabilitation and Home Settings. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:2312–2321, 2022.

[35] Peter S. Lum, Liqi Shu, Elaine M. Bochniewicz, Tan Tran, Lin Ching Chang, Jessica Barth, and Alexander W. Dromerick. Improving Accelerometry-Based Measurement of Functional Use of the Upper Extremity After Stroke: Machine Learning Versus Counts Threshold Method. *Neurorehabilitation and Neural Repair*, 34(12):1078–1087, 12 2020.

[36] Dwaipayan Biswas, Daniele Corda, Giovanni Baldus, Andy Cranny, Koushik Maharatna, Josy Achner, Jasmin Klemke, Michael Jöbges, and Steffen Ortmann. Recognition of elementary arm movements using orientation of a tri-axial accelerometer located near the wrist. *Physiological Measurement*, 35(9):1751–1768, 9 2014.

[37] Gautam P. Sadarangani, Xianta Jiang, Lisa A. Simpson, Janice J. Eng, and Carlo Menon. Force myography for monitoring grasping in individuals with stroke with mild to moderate upper-extremity impairments: A preliminary investigation in a controlled environment. *Frontiers in Bioengineering and Biotechnology*, 5(AUG), 7 2017.

[38] Lisa A. Simpson, Amanda Mow, Carlo Menon, and Janice J. Eng. Preliminary examination of the ability of a new wearable device to capture functional hand activity after stroke. *Stroke*, 50(12):3643–3646, 12 2019.

[39] Chieh Ling Yang, Johnson Liu, Lisa A. Simpson, Carlo Menon, and Janice J. Eng. Real-World Functional Grasping Activity in Individuals With Stroke and Healthy Controls Using a Novel Wearable Wrist Sensor. *Neurorehabilitation and Neural Repair*, 35(10):929–937, 10 2021.

[40] Debeshi Dutta, Srinivasan Aruchamy, Soumen Mandal, and Soumen Sen. Poststroke Grasp Ability Assessment Using an Intelligent Data Glove Based on Action Research Arm Test: Development, Algorithms, and Experiments. *IEEE Transactions on Biomedical Engineering*, 69(2):945–954, 2 2022.

[41] T. Scott Saponas, Desney S. Tan, Dan Morris, and Ravin Balakrishnan. Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 515–524, 2008.

[42] Weidi Tang, Xu Zhang, Xiao Tang, Shuai Cao, Xiaoping Gao, and Xiang Chen. Surface electromyographic examination of poststroke neuromuscular changes in proximal and distal muscles using clustering index analysis. *Frontiers in Neurology*, 8(JAN), 1 2018.

[43] Christopher L. Jones and Derek G. Kamper. Involuntary neuromuscular coupling between the thumb and finger of stroke survivors during dynamic movement. *Frontiers in Neurology*, 9(MAR), 3 2018.

[44] Negin Hesam-Shariati, Terry Trinh, Angelica G. Thompson-Butel, Christine T. Shiner, and Penelope A. McNulty. A Longitudinal Electromyography Study of Complex Movements in Poststroke Therapy. 1: Heterogeneous Changes Despite Consistent Improvements in Clinical Assessments. *Frontiers in Neurology*, 8, 7 2017.

[45] Sang Wook Lee, Kristin Wilson, Blair A Lock, and Derek G Kamper. Subject-specific Myoelectric Pattern Classification of Functional Hand Movements for Stroke Survivors NIH Public Access. *IEEE Trans Neural Syst Rehabil Eng*, 19(5):558–566, 2011.

[46] I. Mendez, B. W. Hansen, C. M. Grabow, E. J.L. Smedegaard, N. B. Skogberg, X. J. Uth, A. Bruhn, B. Geng, and E. N. Kamavuako. Evaluation of the Myo armband for the classification of hand motions. In *IEEE International Conference on Rehabilitation Robotics*, pages 1211–1214. IEEE Computer Society, 8 2017.

[47] Zhen Zhang, Kuo Yang, Jinwu Qian, and Lunwei Zhang. Real-time surface EMG pattern recognition for hand gestures based on an artificial neural network. *Sensors (Switzerland)*, 19(14), 7 2019.

[48] Muhammad Ziaur ur Rehman, Asim Waris, Syed Omer Gilani, Mads Jochumsen, Imran Khan Niazi, Mohsin Jamil, Dario Farina, and Ernest Nlandu Kamavuako. Multiday EMG-Based classification of hand motions with deep learning techniques. *Sensors (Switzerland)*, 18(8), 8 2018.

[49] Tian Xie, Yan Leng, Yihua Zhi, Chao Jiang, Na Tian, Zichong Luo, Hairong Yu, and Rong Song. Increased Muscle Activity Accompanying With Decreased Complexity as Spasticity Appears: High-Density EMG-Based Case Studies on Stroke Patients. *Frontiers in Bioengineering and Biotechnology*, 8, 11 2020.

[50] S Tam, G Bilodeau, J Brown, G Gagnon-Turcotte, A Campeau-Lecours, and B Gosselin. *A Wearable Wireless Armband Sensor for High-Density Surface Electromyography Recording; A Wearable Wireless Armband Sensor for High-Density Surface Electromyography Recording.* 2019.

[51] Ali Moin, Andy Zhou, Abbas Rahimi, Simone Benatti, Alisha Menon, Senam Tamakloe, Jonathan Ting, Natasha Yamamoto, Yasser Khan, Fred Burghardt, Luca Benini, Ana C. Arias, and Jan M. Rabaey. An EMG Gesture Recognition System with Flexible High-Density Sensors and Brain-Inspired High-Dimensional Classifier. 2 2018.

[52] Umberto Barone and Roberto Merletti. Design of a portable, intrinsically safe multichannel acquisition system for high-resolution, real-time processing HD-sEMG. *IEEE Transactions on Biomedical Engineering*, 60(8):2242–2252, 2013.

[53] Zhiyuan Lu, Kai yu Tong, Henry Shin, Sheng Li, and Ping Zhou. Advanced myoelectric control for robotic hand-assisted training: Outcome from a stroke patient. *Frontiers in Neurology*, 8(MAR), 3 2017.

[54] Qiuyang Qian, Xiaoling Hu, Qian Lai, Stephanie C. Ng, Yongping Zheng, and Waisang Poon. Early stroke rehabilitation of the upper limb assisted with an electromyography-driven neuromuscular electrical stimulation-robotic arm. *Frontiers in Neurology*, 8(SEP), 9 2017.

[55] Marcello Mulas, Michele Folgheraiter, and Giuseppina Gini. An EMG-controlled exoskeleton for hand rehabilitation. In *Proceedings of the 2005 IEEE 9th International Conference on Rehabilitation Robotics*, volume 2005, pages 371–374, 2005.

[56] Meng Fen Tsai, Andrea Bandini, Rosalie H. Wang, and José Zariffa. Capturing representative hand use at home using egocentric video in individuals with upper limb impairment. *Journal of Visualized Experiments*, 2020(166), 12 2020.

[57] Alejandro Betancourt, Pietro Morerio, Carlo S. Regazzoni, and Matthias Rauterberg. The evolution of first person vision methods: A survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(5):744–760, 5 2015.

[58] Nandana Welage, Michelle Bissett, Kristy Coxon, Kenneth N.K. Fong, and Karen P.Y. Liu. Development and feasibility of first- and third-person motor imagery for people with stroke living in the community. *Pilot and Feasibility Studies*, 9(1), 12 2023.

[59] Jirapat Likitlersuang, Elizabeth R. Sumitro, Tianshi Cao, Ryan J. Visée, Sukhvinder Kalsi-Ryan, and José Zariffa. Egocentric video: A new tool for capturing hand use of individuals with spinal cord injury at home. *Journal of NeuroEngineering and Rehabilitation*, 16(1), 7 2019.

[60] Meng Fen Tsai, Rosalie H. Wang, and Jose Zariffa. Identifying Hand Use and Hand Roles after Stroke Using Egocentric Video. *IEEE Journal of Translational Engineering in Health and Medicine*, 9, 2021.

[61] Meng-Fen Tsai, Rosalie H. Wang, and José Zariffa. Validity of Novel Outcome Measures for Hand Function Performance After Stroke Using Egocentric Video. *Neurorehabilitation and Neural Repair*, page 154596832311596, 3 2023.

[62] OT- Bioelettronica. Software. https://www.otbioelettronica.it/prodotti/software. Accessed: 28/03/2023.

[63] Nicolette Driscoll, Brian Erickson, Brendan B Murphy, and Andrew G Richardson. MXene-infused bioelectronic interfaces for multiscale electrophysiology and stimulation. 2021.

[64] J. Antonio Ruvalcaba, M. I. Gutiérrez, A. Vera, and L. Leija. Wearable Active Electrode for sEMG Monitoring Using Two-Channel Brass Dry Electrodes with Reduced Electronics. *Journal of Healthcare Engineering*, 2020, 2020.

[65] NIHR Southampton Biomedical Research Centre. Procedure for Measuring ADULT ULNA LENGTH.

[66] Giorgio Biagetti, Paolo Crippa, Laura Falaschetti, Simone Orcioni, and Claudio Turchetti. Human activity monitoring system based on wearable sEMG and accelerometer wireless sensor nodes. *BioMedical Engineering Online*, 17, 11 2018.

[67] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An Introduction to Statistical Learning with Applications in R Second Edition. Technical report, 2021.

[68] Mihir Patel and Matthew Varacallo. Anatomy, Shoulder and Upper Limb, Forearm Bones. Technical report, 2019.

[69] Clinical Gate. Compartment Syndromes and Volkmann Contracture.

[70] Sara Abbaspour, Maria Lindén, Hamid Gholamhosseini, Autumn Naber, and Max Ortiz-Catalan. Evaluation of surface EMG-based recognition algorithms for decoding hand movements. *Medical and Biological Engineering and Computing*, 58(1):83–100, 1 2020.

[71] J Too, A R Abdullah, T N S Tengku Zawawi, N Mohd Saad, and H Musa. Classification of EMG Signal Based on Time Domain and Frequency Domain Features. Technical Report 1, 2017.

[72] Christopher Spiewak. A Comprehensive Study on EMG Feature Extraction and Classifiers. *Open Access Journal of Biomedical Engineering and Biosciences*, 1(1), 2 2018.

[73] Dennis Tkach, He Huang, and Todd A Kuiken. JNER JOURNAL OF NEUROENGINEERING AND REHABILITATION Study of stability of time-domain features for electromyographic pattern recognition. Technical report, 2010.

[74] Mark Everingham, Luc Van Gool, Christopher K.I. Williams, John Winn, and Andrew Zisserman. The Pascal Visual Object Classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303–339, 6 2010.

[75] Dandan Shan, Jiaqi Geng, Michelle Shu, and David F Fouhey. Understanding Human Hands in Contact at Internet Scale. Technical report, 2020.

[76] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. 6 2015.

[77] Ross Girshick. Fast R-CNN. Technical report, 2012.

[78] MATLAB - MathWorks Italia. Rank features for classification using minimum redundancy maximum relevance (MRMR) algorithm. https://it.mathworks.com/help/stats/fscmrmr.html. Accessed 10/04/2023.

[79] Chittaranjan Andrade. Z Scores, Standard Scores, and Composite Test Scores Explained. *Indian Journal of Psychological Medicine*, 43(6):555–557, 11 2021.

[80] Iqbal H. Sarker. Machine Learning: Algorithms, Real-World Applications and Research Directions, 5 2021.

[81] Shahadat Uddin, Arif Khan, Md Ekramul Hossain, and Mohammad Ali Moni. Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, 19(1), 12 2019.

[82] IBM Documentation. How SVM Works. https://www.ibm.com/docs/en/spss-modeler/saas?topic=models-how-svm-works. Accessed: 06/04/2023.

[83] Vasileios Apostolidis-Afentoulis. SVM Classification with Linear and RBF kernels. 2015.

[84] scikit-learn. RBF SVM parameters. https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html. Accessed: 06/04/2023.

[85] scikit-learn. Random Forest Classifier. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomFore Accessed: 06/04/2023.

[86] scikit-learn. machine learning in Python. https://scikit-learn.org/stable/. Accessed: 28/03/2023.

[87] Eva Ostertagová, Oskar Ostertag, and Jozef Kováč. Methodology and application of the Kruskal-Wallis test. *Applied Mechanics and Materials*, 611:115–120, 2014.

7. Abstract in lingua italiana

La ripresa della funzionalità dell'estremità superiore (UE) è essenziale per migliorare la qualità della vita ed è una priorità assoluta per le persone con lesioni al midollo spinale (SCI) e ictus. Negli ultimi anni sono stati fatti progressi nella valutazione della vita quotidiana e nella riabilitazione utilizzando tecnologie indossabili come la visione in prima persona (FPV), che tuttavia ha alcune limitazioni. Lo scopo di questa tesi è quello di sviluppare un sistema multimodale che combina FPV ed elettromiografia superficiale (sEMG) per registrare attività che coinvolgono l'UE e convalidare le sue prestazioni su adulti sani durante attività standardizzate in un ambiente di laboratorio. Le prestazioni di sEMG nel rilevare interazioni funzionali (ad esempio, la manipolazione di un oggetto per uno scopo funzionale) sono state valutate addestrando e testando quattro algoritmi di machine learing. È stato utilizzato un algoritmo di deep learing per rilevare le interazioni tra mano e oggetto dai frame FPV. Le rilevazioni risultanti sono state analizzate per ottenere le prestazioni di classificazione FPV. Sono stati quindi testati tre diverse combinazioni di FPV e EMG. La prima ha coinvolto la concatenazione dello stato di interazione dall'analisi FPV con le caratteristiche più rilevanti estratte dal segnale EMG. Il secondo e il terzo approccio hanno utilizzato gli operatori logici AND e OR, rispettivamente, per combinare lo stato di interazione predetto dal segnale EMG e dall'analisi FPV. Le prestazioni degli approcci single-modal e multi-modal sono state valutate utilizzando metriche di accuratezza, F1-score, precisione, recall e specificità. Le interazioni tra mano e oggetto sono state rilevate automaticamente con una mediana di accuratezza del 0,653 (0,044) per sEMG, 0,650 (0,067) per FPV, 0,716 (0,067) per la combinazione FPV e sEMG. I nostri risultati hanno dimostrato che la strategia multimodale ha superato i due approcci single-modal, come dimostrato dalla maggior parte delle metriche di valutazione utilizzate in questo studio. Questo studio suggerisce che la combinazione di FPV ed EMG è un modo efficace per catturare interazioni tra mano e oggetto funzionali e non funzionali in individui sani. Ricerche future prevederanno la convalida dei nostri risultati su soggetti con SCI o ictus, sia in un contesto clinico che nelle loro case.

Parole chiave: visione in prima persona, elettromiografia di superficie, approccio multimodale, interazione mano-oggetto, funzione della mano