ADVANCED INTELLIGENT SYSTEMS
Open Access
www.advintellsyst.com

Context-Aware Semiautonomous Control for Upper-Limb Prostheses

Gianmarco Cirelli, Enrica Stefanelli, Roberto Billardello, Christian Tamantini, Silvia Leonelli, Loredana Zollo, Luigi Pietro Cordella, and Francesca Cordella*

High rejection rates in upper-limb prosthetics stem from limited usability and excessive cognitive workload (CW), as traditional electromyographic (EMG) control strategies suffer from signal variability and require continuous user effort. Vision-based semiautonomous control strategies (SCS) are proposed to combine user input with contextual information. However, these approaches often provide limited user involvement in the loop and are not broadly validated in prosthetic applications. Their performance in usability and subjective perception remains unexplored. To address these limitations, an innovative SCS is proposed that integrates EMG signals with visual perception to enhance prosthetic performance, usability, and user experience, reducing CW. The proposed SCS allows users to intentionally select a grasp via EMG signals, while computer vision refines wrist orientation and adapts the grasp to the detected object, even in complex scenarios involving multiple objects or conflicting inputs between EMG and visual context. To assess performance, the proposed SCS is compared with the traditional EMG-based control strategy. 10 able-bodied subjects participate in comparative analysis. Results show that SCS achieves a 100% success rate (97.69% for the EMG-based strategy) while reducing task completion time by up to 52.66%. Usability increases (15-point improvement in the system usability scale score), and CW decreases, evidenced by physiological and subjective measures.

1. Introduction

Upper-limb prosthesis rejection represents a significant challenge, with reported rates reaching up to 44%.[1] Despite notable advancements in prosthetic technology, many users prioritize cosmetic over powered prostheses, since they report frustration resulting from inaccurate control, which decreases the usability.[2,3] Most of the commercial devices use electromyography (EMG) to decode motion intention in prosthesis control. However, EMG presents inherent drawbacks: its effectiveness relies on maintaining a balance between the number of gestures to be classified and the overall stability of the system[4] while being hindered by algorithmic constraints and the difficulty amputees face in producing consistent muscle contractions.[5] Over time, factors such as muscle fatigue, sweating, and electrode shift within the prosthetic socket positioning can further compromise control reliability, highlighting the urgent need for more robust prosthetic control solutions.[6,7] To overcome the limitations of EMG-based control, in recent years, computer vision has been integrated to enhance the reliability of prosthetic control and object manipulation. In particular, vision-based control strategies have been used to empower prostheses in environmental perception by mimicking human vision. This approach aims at extracting key features of the object to be grasped and optimizing the preshaping of the prosthetic hand to enhance the robustness of the systems and reduce the cognitive workload (CW) while maintaining intuitive control.[8]

Going into details, several approaches in the literature have proposed a semiautonomous control strategy (SCS) in which EMG signals mainly function as triggers within the control pipeline to increase prosthesis functionalities. For instance, early methods relied on RGB cameras and depth sensors to estimate object geometry and autonomously select the hand gestures and estimate wrist orientation[9] or on the use of cumbersome RGB-D cameras.[10] However, these approaches were unsuitable for prosthetic integration due to their high computational cost, reliance on controlled experimental conditions, and the lack of sensor integration into the prosthetic device (i.e., external ultrasonic depth sensor and cameras). To increase the number of grasp

G. Cirelli, E. Stefanelli, R. Billardello, C. Tamantini, S. Leonelli, L. Zollo, F. Cordella
Research Unit of Advanced Robotics and Human-Centred Technologies
Università Campus Bio-Medico di Roma
Rome 00128, Italy
E-mail: christian.tamantini@cnr.it

C. Tamantini
Institute of Cognitive Sciences and Technologies
National Research Council of Italy
Via Gian Domenico Romagnosi 18a, Rome 00196, Italy

L. P. Cordella
Università degli studi di Napoli Federico II
Naples 80125, Italy

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.202501043.

DOI: 10.1002/aisy.202501043

classes and improve classification reliability, convolutional neural networks were employed for RGB image classification while keeping computational costs low. Convolutional neural networks were employed for RGB image classification,[11,12] but relinquishing the estimate of the wrist orientation.

On the other hand, recent studies have explored the combination of EMG signals and visual sensors, not simply using EMG as a trigger, but integrating both modalities to enhance prosthetic control.[13,14] Despite their effectiveness, these systems have significant limitations in real-world scenarios, as they do not account for factors such as the complexity of the background or the detection of multiple objects in the frame. Moreover, they do not adjust wrist orientation based on the position and shape of the object, and they lack quantitative details on system latency, which is crucial for real-time prosthetic applications. In addition, they have not yet been validated on a real prosthetic system and with subjects, a crucial aspect for evaluating the usability of the device.

This work addresses a critical gap in the current literature on semiautonomous prosthetic control. Although numerous studies propose vision-driven or context-aware strategies and frequently claim a reduction in user cognitive load, such claims have not been quantitatively validated nor compared against conventional EMG-based direct control. To tackle this limitation, the present study introduces and evaluates a multimodal, context-aware SCS designed to enhance the reliability and intuitiveness of prosthetic manipulation by jointly accounting for user motor intention and object-related contextual information. The proposed method integrates the user directly into the control loop to enable intentional grasp selection, while the vision module autonomously regulates wrist orientation and compensates for potential EMG misclassifications, thereby ensuring consistent and context-appropriate grasp execution, even in the presence of multiple objects or unstructured scenes.

A preliminary version of this architecture was previously assessed in a constrained laboratory evaluation focusing on the algorithmic design, operational principles, and technical performance of the underlying components.[15] However, it had never been implemented in a real prosthetic device nor examined in comparison with standard EMG-based control. Building on this methodological foundation, the current study embeds the full SCS into a commercial hand–wrist prosthesis and performs the first holistic validation against state-of-the-art EMG-based strategies.[16] Beyond functional and usability metrics, we introduce an objective CW assessment, combining subjective questionnaires with physiological monitoring, to rigorously quantify whether the proposed semiautonomous approach effectively reduces user effort and improves interaction, as long hypothesized but never empirically demonstrated.

In this way, the present contribution goes beyond foundational proof-of-concept work, providing both a validated integration of multimodal sensing and a comprehensive experimental framework for assessing its benefits and limitations. This establishes a concrete platform for subsequent advances in sensing fusion methodologies and adaptive control strategies, while already demonstrating measurable improvements in prosthetic usability.

2. Experimental Section

An overview of the experimental workflow is presented in **Figure 1**, summarizing the four main phases of the process: development of the control strategy, validation through task execution, and data processing. Each phase is described in detail in the following sections.

2.1. Control Strategy

2.1.1. Semiautonomous Control Strategy

The proposed SCS integrates user intention with computer vision through four key components: EMG gesture classifier, object detection, grasp selection, and orientation estimation.

EMG Classifier: User motion intent was decoded using a nonlinear regression (NLR) classifier,[16] directly processing raw EMG signals. The recorded EMG data were segmented into 200 ms time windows with a 50 ms overlap. The classifier output comprised eight gesture classes (G_{EMG}), which can be grouped into three categories: 1) nongrasping motions: rest and open hand; 2) wrist motion commands: pronation and supination; and 3) grasping macrocategories: power (medium wrap), pinch (2-digits), pointing and lateral.

These gesture classes were selected to cover the essential functional movements required for daily prosthetic use. Here, the term "macrocategory" refers to a coarse functional grouping of hand-grasping gestures. These categories represent the essential control primitives for prosthetic use, while finer, object-specific grasp variations are introduced through the vision-based module.

Moreover, to enhance prediction robustness, a majority voting scheme was implemented,[17] aggregating the 200 classifier outputs generated within each time window to determine the final gesture prediction.

Object Detection: The object detection module relies on the YOLOv5s model trained on the COCO dataset,[18] a large-scale dataset containing 80 object classes. YOLO-based architectures are known for their fast inference and robust performance even in the presence of partially occluded objects, which makes them suitable for real-time prosthetic control scenarios.[19] For this application, 13 object-related gesture classes (G_{CV}) were considered, focusing on objects frequently involved in activities of daily living (ADLs).

Each detected object is semantically linked to a specific grasp type, following the grasp taxonomy by Feix et al.[20] By leveraging this object-to-grasp mapping, the system facilitates context-aware control, selecting grasps that align with the functional requirements dictated by the object properties. The strategy enables the EMG classifier to maintain high accuracy by limiting the number of directly recognized grasp classes to those previously reported. At the same time, it expands the effective set of possible grasps by combining EMG inputs with vision-based contextual information, thus enhancing the prosthesis overall grasping functionality. Specifically, it supports 13 gestures: 1) nongrasping motions: rest and open hand; 2) wrist motion commands: pronation and supination; and 3) object-specific grasps: medium wrap, adducted thumb, fixed hook, lateral, pointing, 2-digit pinch, 3-digit pinch, spherical, and parallel extension.

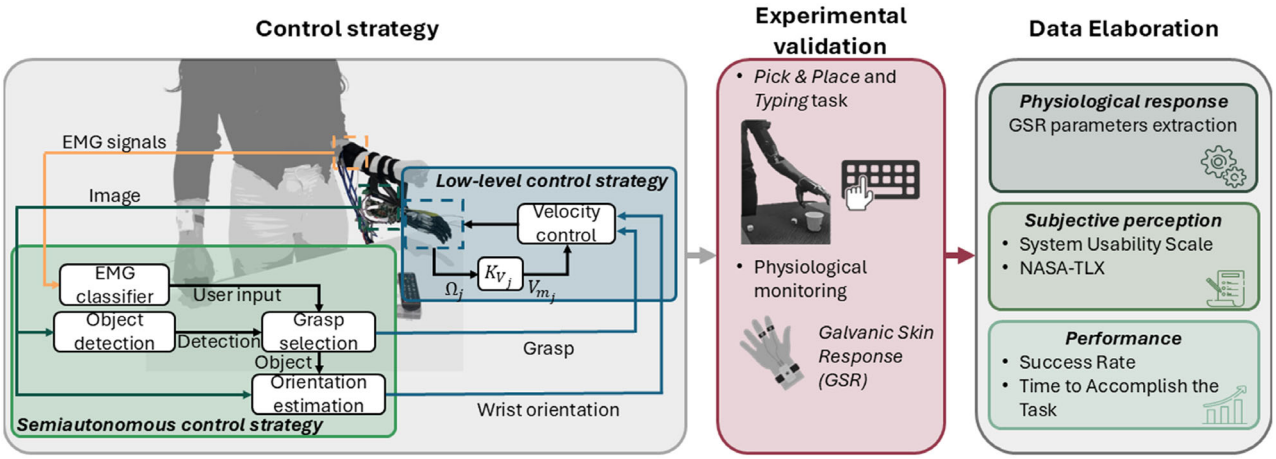

Figure 1. Schematic representation of the experimental workflow, divided into three main phases: control strategy implementation, task execution, and data processing. This workflow ensures a structured approach to evaluating the proposed SCS, by comparing it with an EMG-based strategy to highlight the methodology used for assessing performance and usability. In the low-level control strategy block, V_{m_j} is the back electromotive force, Ω_{m_j} is the angular velocity of the motors, and K_{V_j} is the voltage constant.

In this way, the multimodal approach combines high EMG decoding accuracy with vision-based contextual adaptation, enhancing the overall grasping functionality of the prosthesis.

Grasp Selection: The grasp selection module is responsible for resolving ambiguities between the user motor intention, decoded via EMG signals, and the contextual information provided by the vision system. Specifically, the EMG classifier predicts a macro-category of grasp, while the vision system suggests object-specific grasps based on the detected object. The module compares these two sources to determine the most appropriate grasp to execute. The aim is to identify the object most likely intended by the user and determine the most appropriate grasp to execute.

This decision-making process is governed by a heuristic procedure detailed in **Algorithm 1**. When no objects are detected in the scene, the system defaults to a rest state, assuming the user is not attempting to grasp anything. If the camera frames one or more objects, the system evaluates whether the gestures predicted by the vision module ($G_{\mathrm{CV}}(i)$) are consistent with the gesture inferred from the EMG classifier (G_{EMG}). In the absence of an object that is compatible with the gesture, the system selects the object with the highest confidence score and assigns the corresponding visual gesture as the final command. In instances where a single object exhibits a gesture that precisely aligns with the EMG prediction, the system will select that object with a high degree of confidence and proceed accordingly.

If multiple candidate objects are associated with gestures compatible with the EMG prediction, the system restricts the analysis to those located within a central region of the image, defined as the set of objects whose centroid positions $\mathbf{P}_i$ lie within a radius δ from the image center $\mathbf{c}$, that is, $||\mathbf{p}_i - \mathbf{c}|| < \delta$. This reflects the assumption that objects appearing near the center of the visual field are more likely to be the user's intended targets. Among these central objects, the system selects the one with the highest confidence score whose associated visual gesture $G_{\mathrm{CV}}(i)$ is a subset of the EMG prediction G_{EMG}. The corresponding gesture is then selected as the final grasp command G_{C} and transmitted to the prosthesis controller.

This strategy allows the system to prioritize agreement between user input and contextual cues, promoting intuitive and context-aware control. Moreover, it supports robust interaction in dynamic environments by resolving conflicts between EMG and vision in a deterministic and explainable way. The combination of gesture compatibility filtering and confidence-based ranking ensures that the prosthesis responds to the most relevant and intention-consistent object in the scene, ultimately enhancing usability and control efficiency.

Orientation Estimation: The orientation estimation module determines the optimal wrist orientation, specifically the pronation/supination (P/S) angle, by analyzing the geometry of the selected object using principal component analysis (PCA). This process is applied exclusively to the region of interest (ROI), defined by the bounding box of the object selected through the grasp selection procedure, thereby reducing computational

Algorithm 1. Grasp selection

Input: $\{G_{\mathrm{CV}}(i), \mathbf{p}_i, \mathrm{score}(i)\}$ for $i = 1, \ldots, N$; G_{EMG}; image center $\mathbf{c}$
Output: G_{C}

if no object is framed **then**
  $G_{\mathrm{C}} \leftarrow$ Rest
else
  if $G_{\mathrm{CV}}(i) \not\subseteq G_{\mathrm{EMG}}, \forall i \in \{1, \ldots, N\}$ **then**
    $k \leftarrow \arg\max_i \mathrm{score}(i)$
    $G_{\mathrm{C}} \leftarrow G_{\mathrm{CV}}(k)$
  else if $\exists! k \in \{1, \ldots, N\} : G_{\mathrm{CV}}(k) = G_{\mathrm{EMG}}$ **then**
    $G_{\mathrm{C}} \leftarrow G_{\mathrm{CV}}(k)$
  else
    given $\mathcal{C} = \{i \in \{1, \ldots, N\} \mid ||\mathbf{p}_i - \mathbf{c}|| < \delta\}$
    $L \leftarrow \{i \in \mathcal{C} \mid G_{\mathrm{CV}}(i) \subseteq G_{\mathrm{EMG}}\}$
    $k \leftarrow \arg\max_{i \in L} \mathrm{score}(i)$
    $G_{\mathrm{C}} \leftarrow G_{\mathrm{CV}}(k)$
  end if
end if

cost and minimizing the influence of background noise. The P/S angle is computed as the angle between the first principal component vector and the horizontal axis, and the resulting orientation estimate is then used to configure the wrist angle prior to grasp execution. In parallel, visual feedback is provided to the user via light-emitting diodes (LEDs), indicating the presence of detected objects and potential mismatches between the EMG-based classification and the computer vision outputs. This simple yet effective solution was chosen to enhance system transparency and user awareness during operation. LEDs offer a lightweight, low-cost, and nonintrusive feedback mechanism, making them ideal for real-time communication without interfering with the user's natural interaction with the prosthesis.

The described control pipeline is activated whenever the user performs a grasp gesture and remains active as long as objects are detected in the scene for at least five consecutive frames. Given that the analysis frequency is 2.07 frames per second,[15] the average execution time for this detection process is ≈2.4 s. Once this condition is fulfilled, the computer vision system (CVS) is turned off to minimize energy consumption and enhance the overall efficiency of the algorithm, particularly in its communication with the prosthesis. After grasping the object, the user can either rotate the hand or release it by performing the open-hand gesture, while all other hand gestures are not transmitted as input to the prosthesis. This allows the subject to adopt a rest configuration during task execution without maintaining muscle contraction, thereby reducing fatigue.

2.1.2. Low-Level Control of the Hand–Wrist Prosthesis

As shown in Figure 1, the low-level control received the class to be performed by the SCS. As outlined in 2.1.1, the SCS managed nine grasp classes, reported in **Figure 2**, together with rest and open-hand gestures, and wrist pronation-supination, based on the visualized object. Wrist orientation refers to the optimal position of the wrist before the grasping is performed. In particular, the wrist class is provided for a fixed time interval calculated from the wrist rotation speed, set at 30° per s, that is , the average value calculated on subjects without physical impairments performing ADLs.[21]

A velocity controller was implemented to control the wrist motor and the degrees of freedom of the hand. In particular, the hand motors independently controlled the movements of the fingers: the adduction/abduction and flexion/extension of the thumb and the flexion/extension of the index, middle, ring, and little fingers. The voltage error was calculated for each finger and wrist motor (j) as

$$e_{r_j} = V_{R_j} - V_{m_j} \tag{1}$$

where V_{m_j} is the back electromotive force and V_{R_j} is the reference voltage.[22] V_{m_j} is obtained by multiplying the angular velocity of the motors (Ω_{m_j}), and the voltage constant (K_{V_j}), which depends on the wrist and fingers physical building details.

2.2. Experimental Validation

To assess the performance and the usability of the proposed control strategies, an experimental validation was conducted. In accordance with the ethical guidelines and regulations governing research involving human subjects, this study was deemed exempt from approval by a relevant review board, as it presents minimal risk to participants and primarily involves observation without any invasive or potentially harmful procedures. A detailed description of the experimental protocol and setup, including the procedures followed and the equipment used to evaluate the system effectiveness, is provided in the following.

2.2.1. Experimental Protocol

A comparative analysis with the EMG-based control strategy was conducted. Specifically, the EMG-based approach is relying on the gesture classifier described in Section 2.1.1. The output of the classifier—one of the eight possible classes—is provided as input to the prosthetic device.

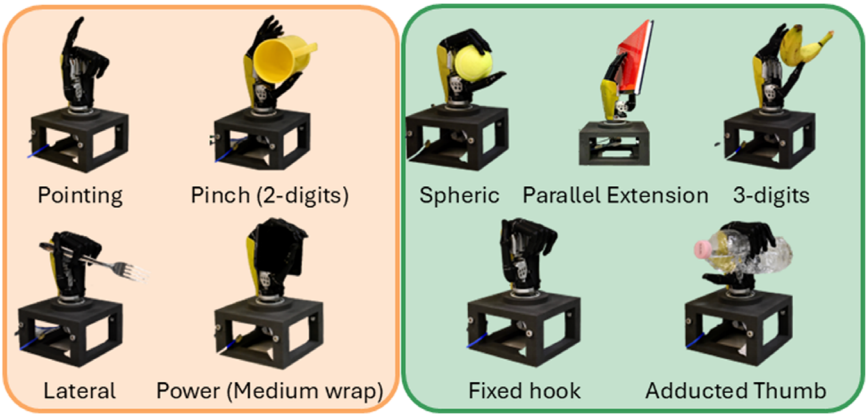

Figure 2. List of the hand-grasping gestures used in the EMG-based control strategy (left) and those introduced in the multimodal SCS (right). The objects associated with each hand gesture are: keyboard (pointing), spoon and fork (lateral), cup (2-digit), phone and remote (medium wrap), ball (sphenic), book (parallel extension), banana and wine glass (3-digits), suitcase and backpack (fixed hook), and bottle (adducted thumb).

10 able-bodied subjects (equally divided between males and females, average age = 27.6 ± 2.8 years) were recruited to perform a pick & place task with 12 selected objects and a typing task using a keyboard. Objects are reported in Figure 2.

The experimental sessions were conducted in a randomized order across subjects. However, to ensure balance, half of the subjects began with the semiautonomous strategy, while the other half started with the EMG-based one. Moreover, the 13 tasks were performed randomly within each experimental session, with three repetitions per task, resulting in a total of 39 repetitions.

The objects were positioned on a 70 cm high table and were displaced 30 cm to the right from their initial location during the pick & place task. In this task, subjects were instructed to reach the object, grasp it, move it to a designated target position on the table, release it, and then return their hand to the initial rest position. Positions remained fixed and consistent across subjects and sessions, while object orientations on the table were intentionally randomized across trials to evaluate the capability of the vision module to generalize grasp classification and orientation estimation under different viewpoints. In the typing task, subjects were instead required to type a five-letter word ("hello") on a standard keyboard, using the prosthesis to press the keys sequentially.

At the beginning of each experimental session, participants were instructed to perform each gesture five times, holding the position for 5 s while freely moving their arm within the workspace. Allowing arm movements during gesture execution aimed to capture the natural variability in muscle activation caused by changes in limb posture, thereby improving the robustness of the classifier across different positions within the workspace. During this phase, EMG signals were continuously recorded. Subsequently, the NLR classifier was tested offline using a simple hold-out approach (10% test, 90% training) and trained for the online test using all the subject-collected data.

CW is often assessed through subjective questionnaires such as NASA Task Load Index (NASA-TLX)[23] and SUS.[24] However, physiological measurements offer the possibility of quantifying CW through objective, continuous indicators. Among the available physiological metrics, galvanic skin response (GSR) is particularly suitable for capturing rapid, dynamic variations in workload, making it well aligned with the temporal structure of the tasks considered in this study.[25–27] Other cardiorespiratory indicators, such as heart rate, heart rate variability, and respiration rate, are also commonly employed in workload assessment, but their most reliable features typically require longer temporal windows than those available in our experimental design.[28] Moreover, these measures are generally more sensitive to physical rather than purely cognitive load.[29]

Given the suitability of GSR for capturing rapid fluctuations in CW, the signal was further processed to ensure comparability across participants. To normalize GSR data with respect to a resting condition, a 2 min baseline was recorded before the start of the experimental session. Each GSR time series was then normalized by subtracting the mean baseline value and dividing the result by the same baseline mean, yielding a dimensionless signal that captures relative variations with respect to each participant's physiological resting level.[30]

2.2.2. Experimental Setup

Figure 3 shows the experimental setup of both experimental sessions.

The user intent was decoded using a bracelet with six equally spaced 13E200 Myobock electrodes (Ottobock). Signal acquisition was performed via an NI USB-6002 DAQ, with a sampling rate of 1 kHz. To minimize crosstalk from the biceps brachii and optimize signal acquisition, the armband was placed on the forearm, 4 cm below the elbow, in accordance with the standard placement used in prosthetic devices.[31]

The CVS consists of a single board computer (SBC) as a computing unit, that is, the Raspberry Pi 4B (8 GB RAM), paired with a miniature 16 MP RGB camera with autofocus (Arducam 16 MP IMX519). The camera was placed in the ventral part of the prosthetic wrist, as it ensures the lowest level of occlusion due to the presence of the hand with respect to the other possible camera placements, that is, dorsal, internal, and external. In particular, a level of occlusion lower than 15% of the total number of pixels was calculated considering images acquired from 10 able-bodied subjects wearing camera in the four different configurations.[15] The resolution of the camera was set to 320 × 240 pixels, as previous works have demonstrated that it guarantees an optimal trade-off between the inference time and object detection.[15] Remote access to the SBC desktop was provided via virtual network computing, while a portable 5 V DC/3 A USB-C power bank supplied power, ensuring portability.

The prosthetic device, which includes the Touch Bionics RoboLimb hand (Össur) and the wrist rotator 10S17 (Ottobock), was powered by a bench power supply capable of delivering 7.4 V and 3 A. A custom nylon support, manufactured using additive technology (selective laser sintering), was attached to the subject's prosthetic-hand side with Velcro.

The GSR was measured using the Shimmer3 GSR+ unit, which assesses skin conductivity through two reusable electrodes placed on the index and middle fingers of the right hand with a sampling frequency of 52 Hz.

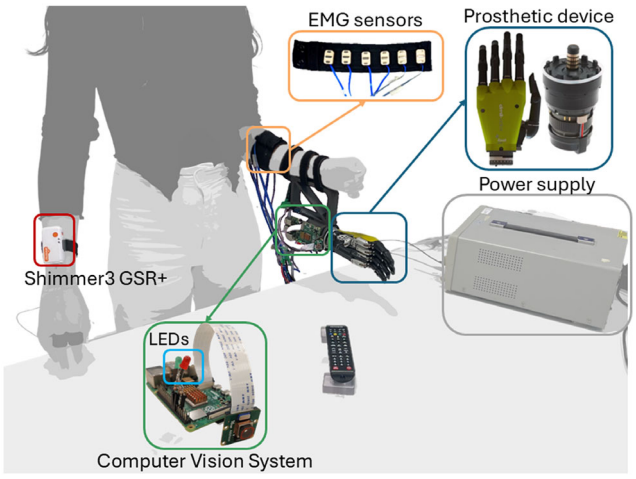

Figure 3. Illustration of the experimental setup, including the prosthetic system, EMG acquisition setup, GSR acquisition system, CVS, and task execution environment.

2.3. Data Elaboration

The extracted data and key performance indicators (KPIs) used to evaluate both system performance and user CW are presented below. The analysis includes physiological responses derived from GSR signals, subjective perceptions, and performance metrics related to task execution.

2.3.1. Physiological Response

The raw GSR signal was processed using a low-pass filter with a cutoff frequency of 5 Hz to remove noise. In the first step, the baseline was computed, corresponding to the subject's resting period. The mean baseline value was then used to normalize the data, minimizing intersubject and intersession variability and ensuring comparability of the results. The signal was segmented for each trial and decomposed into its main components, that is, the skin conductance level (SCL) and skin conductance response (SCR), using a fourth-order Butterworth low-pass filter with a cutoff frequency of 0.1 Hz and a fourth-order Butterworth high-pass filter with the same cutoff frequency, respectively. For each trial, the following parameters were computed, as they are considered indicative of cognitive load:[25,26] 1) number of peaks (NumPeaks[-]), which corresponds to the total number of responses detected from the SCR component during each task. According to the literature, a peak is typically characterized by an increase in the SCR signal exceeding 0.01–0.05 μS. In this study, the threshold for peak detection was set at 0.03 μS;[32] 2) mean SCL (SCLMean[-]), which represents the mean value of the tonic component throughout the entire task and reflects the slow variation in skin conductance; and 3) mean power of the spectrum of the GSR signal (MeanPower[-]), evaluated within the 0–1 Hz band, where the most significant spectral variations were observed.[33] It is computed as

$$\text{mean}(P(\omega)) = \text{mean}\left(\frac{1}{N} Y(\omega) Y^*(\omega)\right), \omega \in [0,1]\text{Hz} \quad (2)$$

where P represents the power spectrum, ω denotes the frequency, N is the signal length, and Y and Y^* correspond to the frequency spectrum and its complex conjugate, respectively.

2.3.2. Subjective Perception

Subjective perception was assessed by administering two questionnaires, the NASA-TLX[23] and the system usability scale (SUS),[24] at the end of each experimental session, since they are gold-standard tools for assessing subjective workload and system usability, respectively. Furthermore, each subject was given a customized questionnaire to identify, for both control strategies, the macrocategory of hand gesture they found easiest to manage and the one they found most challenging.

2.3.3. Performance

For the performance assessment, the success rate (SR) and the time to accomplish the task (TAT) were considered as KPIs.

2.3.4. Statistical Analysis

The Wilcoxon paired-sample test was performed for both the involved KPIs and questionnaires to analyze the differences between the two different control strategies, that is, the SCS and the EMG-based one, with a significance level set at $p - value = 0.05$.

3. Results and Discussion

Figure 4 presents the results for the selected KPIs across both control strategies. The data is grouped into grasp macrocategories to enhance interpretation and readability. Although both control strategies resulted in high SR, the EMG-based control achieved an average SR = 97.69%, while the SCS approach outperformed it by reaching an SR = 100%. This result highlights how the integration of visual information in the SCS strategy can effectively compensate for EMG misclassifications, showing that, when users rely on the vision-based system in cases where EMG decoding fails, overall reliability increases.

Figure 4a shows the TAT values for the four macrocategories of grasp. It is evident that the TAT is statistically significantly lower in the SCS condition compared to the EMG-based control. This difference is particularly noticeable in the more complex typing task (pointing gesture), where the median TAT is 50.49 s (33.80, 80.16) for the EMG-based control and 26.59 s (21.86, 37.24) for the SCS approach. This result highlights how the EMG-based control, in the presence of limb movements (such as during repositioning or manipulation), suffers from an increased risk of misclassification due to the inherent variability in EMG signals under dynamic postural changes. Conversely, the SCS approach retains the previously classified grasp until an open-hand gesture is detected, thereby providing more robust and stable control.

Regarding the parameters extracted from the GSR, the number of peaks observed in the SCR shows a statistically significant difference between the two sessions, as reported in Figure 4b. This reduction in the number of peaks suggests that the proposed SCS leads to a more efficient distribution of cognitive resources, as indicated by the decrease in physiological responses.[26] Instead, the median SCLMean value, which reflects the overall activation of the sympathetic nervous system, does not show significant variations between the two sessions (see Figure 4c), suggesting that the tonic activation is not substantially influenced by the control strategy. Lastly, the median spectral power shows a slight decrease in the SCS session compared to the EMG control one, as illustrated in Figure 4d. Although a reduction in spectral power is observed across all gesture categories, a statistically significant difference emerges only in the pinch macrocategory. This result is particularly noteworthy, as low-frequency spectral power is often associated with the nervous system's ability to adaptively respond to external stimuli.[33] Hence, the observed reduction with the proposed SCS may reflect a decrease in signal complexity, which is in line with a reduced cognitive load. Moreover, the localized significance in the pinch gesture might reflect the higher sensorimotor demands or fine motor precision required in this condition,

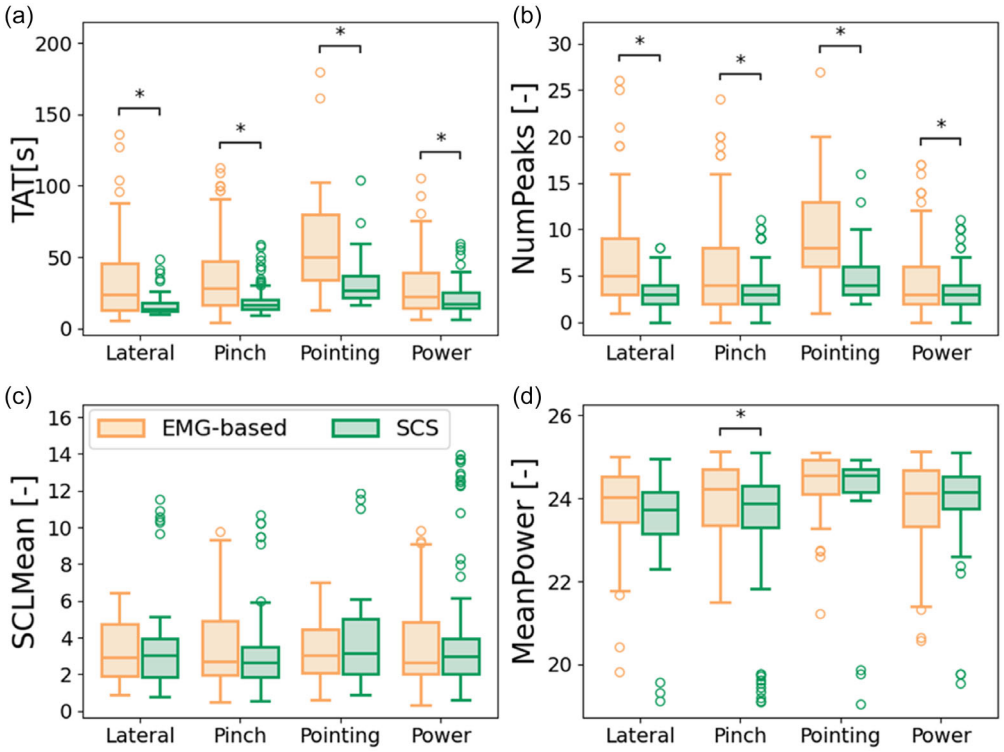

Figure 4. Box plots comparing KPIs for the EMG-based and semiautonomous strategies. The KPIs are: a) TAT, b) number of peaks (NumPeaks), c) mean SCL (SCLMean), and d) mean power of the spectrum of the GSR signal (MeanPower). Asterisk denotes statistically significant differences ($p - value = 0.05$).

making it more sensitive to changes in cognitive or neuromuscular load.

Figure 5 shows the results obtained from the administration of the questionnaires. Concerning the NASA-TLX questionnaire, the SCS outperformed the EMG-based approach, with a significant decrease for mental demand (MD), physical demand (PD), effort (E), and frustration (F). No statistically significant differences were found for either temporal demand (TD) or performance (P). Specifically, the result obtained for the P aligns with the SR, which remains comparable between the two tested strategies.

The median SUS score is significantly higher in the SCS condition compared to the EMG-based control condition. According to established SUS score rankings, the multimodal condition falls within the "Good/Excellent" usability range, while the EMG-based condition corresponds to the "OK/Poor" range, indicating a notably better user experience with the multimodal approach.[24]

The custom questionnaire results, based on participants' subjective perceptions, revealed differences in grasp difficulty between control methods. With EMG control, the lateral grasp was perceived as the most challenging (7 out of 10 subjects), whereas SCS did not highlight a specific difficulty (7 out of 10 subjects). For the easiest grasp, users predominantly favored the power grasp for the EMG-based control (7 out of 10), while the same users showed a more balanced distribution when using the semiautonomous approach, still leaning toward power (4 out of 10). These findings suggest that SCS offers more uniform perceived performance, reinforcing its advantages in real-world applications, while challenges with EMG-based control may arise from the effectiveness of the NLR classifier, with a subjective tendency for the lateral grasp.

Taken together, these findings support the main hypothesis of this work, namely that augmenting EMG-based control with context-aware vision leads to more efficient and less demanding interaction than purely EMG-based direct control. In the proposed SCS, EMG provides the user intentional input, while vision contributes object-related information and wrist orientation cues that refine and stabilize grasp execution. The grasp selection module was specifically designed to integrate these two sources of information in a structured way, ensuring that coherent EMG–vision combinations are reinforced and that conflicting cases are handled through deterministic rules that preserve consistency and support the user's intended action. A detailed technical validation of the system's ability to handle the different conditions that naturally arise in real-world use was already presented in a previous work.[15] Specifically, the grasp selection module can correctly manage: i) coherent situations, in which the grasp inferred from the user's EMG signals matches the grasp associated with the object detected in the visual scene, including cluttered environments where multiple objects are present; and ii) noncoherent situations, in which EMG and vision propose conflicting actions, such as when the user produces the intention to grasp that does not match the one suggested by the vision system, framing the object. In both cases, the proposed grasp selection approach applies explicit

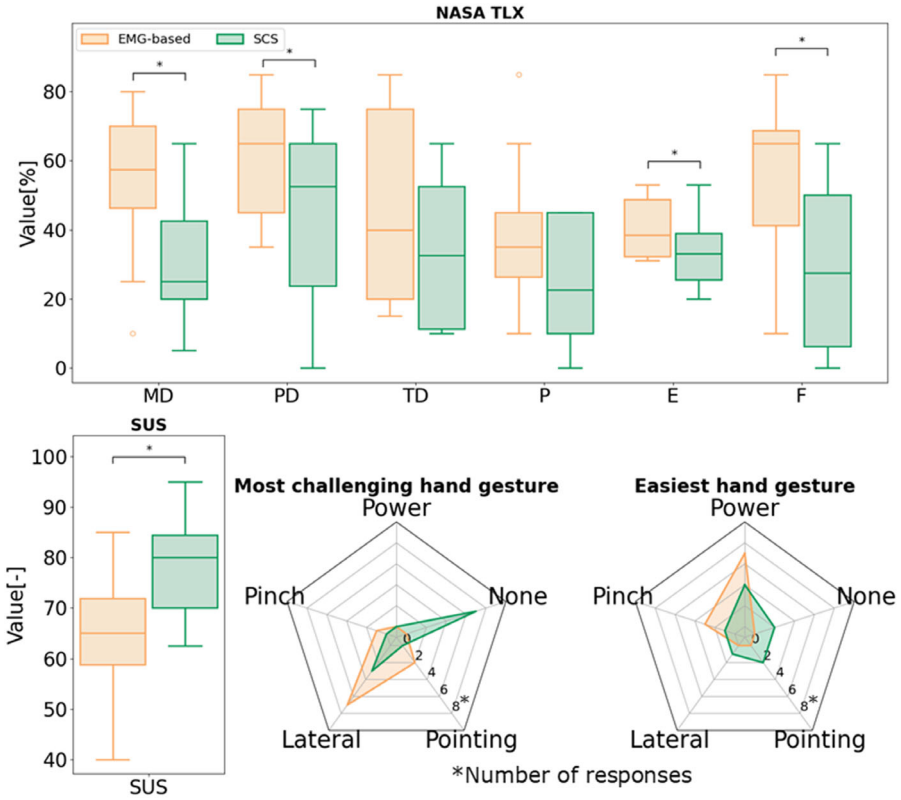

Figure 5. Results from subjective evaluations of CW (NASA-TLX), usability (SUS), and user perception of grasp difficulty across the two control strategies. The NASA-TLX items are: MD, PD, TD, P, E, and F. Asterisk denotes statistically significant differences ($p - value = 0.05$).

decision rules to identify and produce the most plausible and context-appropriate prosthesis command. The present study builds upon that foundation by demonstrating that this integration strategy not only operates reliably at the algorithmic level but also results in measurable benefits for users during real prosthesis interaction.

The improved SR, reduced task completion times, and lower physiological and subjective indicators of CW observed in this study show that combining user intention with contextual visual information achieves a more intuitive and reliable control experience. These results confirm, for the first time through quantitative evidence, a reduction in cognitive demand, an advantage frequently claimed in the literature for semiautonomous strategies but never empirically demonstrated in direct comparison with traditional EMG-based control.

On the other hand, it is worth noting that the task execution times observed in this study are considerably longer than those typically reported for able-bodied individuals performing comparable pick & place and typing tasks. In the literature, healthy subjects generally accomplish similar actions in ≈3 s.[21,34] **Figure 6** shows the complete timeline of a representative trial, illustrating the sequence of events from object framing to task completion. In this example, the object detection block accounts for ≈97% of the vision processing time, confirming that detection is the most computationally demanding stage of the pipeline. This suggests that optimizing or accelerating the detection module could substantially reduce perceptual latency.

Nevertheless, the overall task duration is largely dictated by the user's physical interaction with the object and by the mechanical execution of the grasp, rather than by computational delays. While the proposed SCS mitigates the well-known limitations of EMG-based direct control—thereby facilitating prosthesis use and improving the fluidity of the interaction—execution times remain substantially higher than physiological values. This gap reflects both the cognitive and motor demands associated with operating an artificial device and the current limits of prosthetic hardware. Future work should therefore address not only algorithmic and multimodal control improvements but also advances in actuation and embedded computation to move toward prostheses capable of delivering more natural speed and responsiveness.

3.1. Limitations

Despite the promising results, some aspects of this study warrant consideration. First, experiments were conducted exclusively with able-bodied participants. Although the proposed semiautonomous, context-aware control strategy showed reductions in CW and improved functional performance, further investigations involving individuals with upper-limb loss will be important to confirm the generalizability of these findings, given the distinct motor and cognitive demands they may experience.

Second, the experimental tasks focused on isolated pick & place and typing actions. Although these tasks are widely used

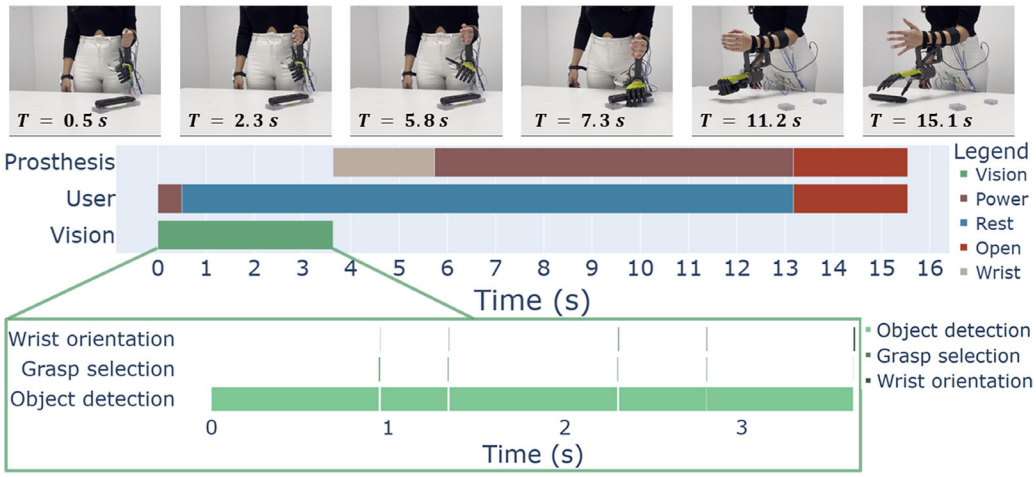

Figure 6. Timeline of user intention, vision system, and prosthesis activity during the execution of a representative grasping task using the proposed SCS. It provides a detailed temporal breakdown of the actions performed by the user, and the prosthesis, and when the vision module is activated. The zoomed view highlights the internal processing stages of the computer vision pipeline (i.e., object detection, grasp selection, and wrist orientation estimation), showing their contribution within the first seconds of the task. The upper frame sequence illustrates key moments of the interaction, from object framing to grasp completion.

benchmarks in prosthetic research, they only partially capture the diversity and richness of daily-living activities. Additionally, participants operated using predefined grasp types; for some objects, alternative grasps may be equally viable depending on the approach direction. This constraint allowed for controlled evaluation but may limit ecological realism.

From a technological perspective, the system relies on a trained object detector (YOLOv5s), whose performance is bounded by the training dataset. Consequently, objects that are visually complex, severely occluded, or not represented in the training set may be more challenging to recognize, potentially affecting the generation of contextual commands in highly unstructured scenarios.

4. Conclusion

This study proposes and validates a SCS that integrates EMG signals with visual perception to enhance the usability, the efficiency, and the reliability of upper-limb prosthetic control. Unlike previous approaches that rely solely on EMG signals or fully autonomous vision-based systems, the proposed method enables users to intentionally select grasp types while leveraging computer vision to refine wrist orientation and provide contextual adjustments. Importantly, this multimodal control strategy has been successfully implemented within a commercial prosthetic system, demonstrating its practical applicability and readiness for real-world use.

To evaluate its effectiveness, the SCS was compared to an EMG-based control strategy in terms of performance, usability, and CW. Experimental results with able-bodied participants demonstrated that the proposed SCS outperforms the EMG-based approach, achieving an SR = 100% and reducing the median TAT by 52.66%.

Usability was higher with the SCS, as reflected in a 15-point increase in the SUS score. CW also appeared lower in this condition, as indicated by physiological and subjective measures. The multimodal approach resulted in a fewer number of responses detected from the SCR component of the GSR during tasks and a lower median GSR power spectrum, suggesting reduced autonomic activation. This trend was further supported by NASA-TLX scores, which showed a statistically significant reduction in MD, PD, E, and F.

By addressing the limitations of traditional EMG-based control and integrating vision for enhanced adaptability, this work contributes to the development of more intuitive and effective prosthetic control strategies.

Future efforts will be devoted to evaluating the usability and robustness of the proposed approach in ecologically valid scenarios, progressively moving toward autonomous at-home use by end-users, that is, the amputees. Assessing the system outside controlled laboratory conditions, over longer and more naturalistic sessions, will also enable the integration of additional unobtrusive physiological measurements of the CW. Moreover, from a technological perspective, future work will aim to extend the vision algorithm to handle objects that afford multiple grasp types, adapting the suggested grasp according to the approach direction and viewpoint of the prosthetic user.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was supported in part by the European Union - Next Generation EU - NRRP M4.C2 - Investment 1.5 Establishing and strengthening of Innovation Ecosystems for sustainability (Project n. ECS00000024 Rome Technopole), in part by the Italian Ministry of Research, under the complementary actions to the NRRP "Fit4MedRob - Fit for Medical

Robotics" Grant PNC0000007, (CUP: B53C22006990001), in part by INAIL Prosthetic Center with BioInterNect (CUP: E57G23000280005), and in part by Fondazione Compagnia di San Paolo within the project "Bio-inspired stimulation techniques for restoring multimodal somatic sensations in patients with sensorimotor limitations - BIO-MIME", funded under the Bando vEIColo - Accompagnamento per la valorizzazione della ricerca.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

computer vision, electromyographic, pattern recognition, semiautonomous control, upper-limb prosthetics

Received: August 31, 2025
Revised: December 1, 2025
Published online:

[1] S. Salminger, H. Stino, L. H. Pichler, C. Gstoettner, A. Sturma, J. A. Mayer, M. Szivak, O. C. Aszmann, "Current Rates of Prosthetic Usage in Upper-Limb Amputees–Have Innovations had an Impact on Device Acceptance?," *Disability and Rehabil.* **2022**, *44*, 3708.

[2] L. C. Smail, C. Neal, C. Wilkins, T. L. Packham, "Comfort and Function Remain Key Factors in Upper Limb Prosthetic Abandonment: Findings of a Scoping Review," *Disability and Rehabil. Assistive Technol.* **2021**, *16*, 821.

[3] C. L. McDonald, S. Westcott-McCoy, M. R. Weaver, J. Haagsma, D. Kartin, "Global Prevalence of Traumatic Non-Fatal Limb Amputation," *Prosthet. Ortho. Int.* **2021**, *45*, 105.

[4] D. Yadav, K. Veer, "Recent Trends and Challenges of Surface Electromyography in Prosthetic Applications," *Biomed. Eng. Lett.* **2023**, *13*, 353.

[5] D. Farina, I. Vujaklija, R. Brånemark, A. M. J. Bull, H. Dietl, B. Graimann, L. J. Hargrove, K.-P. Hoffmann, H. Huang, T. Ingvarsson, et al., "Toward Higher-Performance Bionic Limbs for Wider Clinical Use," *Nat. Biomed. Eng.* **2023**, *7*, 473.

[6] D. Yeung, I. M. Guerra, E. S. Ian Barner-Rasmussen, D. Farina, I. Vujaklija, "Co-Adaptive Control of Bionic Limbs via Unsupervised Adaptation of Muscle Synergies," *IEEE Trans. Biomed. Eng.* **2022**, *69*, 2581.

[7] P. Gulati, Q. Hu, S. Farokh Atashzar, "Toward Deep Generalization of Peripheral Emg-Based Human-Robot Interfacing: a Hybrid Explainable Solution for Neurorobotic Systems," *IEEE Rob. Auto. Lett.* **2021**, *6*, 2650.

[8] L. Gionfrida, D. Kim, D. Scaramuzza, D. Farina, "Wearable Robots for the Real World Need Vision," *Sci. Rob.* **2024**, *9*, eadj8812.

[9] S. Došen, C. Cipriani, "Cognitive Vision System for Control of Dexterous Prosthetic Hands: Experimental Evaluation," *J. Neuroeng. Rehabil.* **2010**, *7*, 1.

[10] M. N. Castro, S. Dosen, "Continuous Semi-Autonomous Prosthesis Control using a Depth Sensor on the Hand," *Front. Neurorob.* **2022**, 814973.

[11] M. C. F. Castro, W. C. Pinheiro, G. Rigolin, "A Hybrid 3d Printed Hand Prosthesis Prototype Based on Semg and a Fully Embedded Computer Vision System," *Front. Neurorob.* **2022**, *15*, 751282.

[12] P. Weiner, J. Starke, S. Rader, F. Hundhausen, T. Asfour, "Designing Prosthetic Hands with Embodied Intelligence: the Kit Prosthetic Hands," *Front. Neurorob.* **2022**, *16*, 815716.

[13] S. Deshmukh, V. Khatik, A. Saxena, "Robust Fusion Model for Handling emg and Computer Vision Data in Prosthetic Hand Control," *IEEE Sensors Lett.* **2023**, *7*, 1.

[14] M. Zandigohar, M. Han, M. Sharif, "Multimodal Fusion of emg and Vision for Human Grasp Intent Inference in Prosthetic Hand Control," *Front. Rob. AI* **2024**, *11*, 1312554.

[15] G. Cirelli, C. Tamantini, L. P. Cordella, F. Cordella, "A Semiautonomous Control Strategy Based on Computer Vision for a Hand–Wrist Prosthesis," *Robotics* **2023**, *12*, 152.

[16] F. Leone, F. Mereu, C. Gentile, F. Cordella, E. Gruppioni, L. Zollo, "Hierarchical Strategy for Semg Classification of the Hand/Wrist Gestures and Forces of Transradial Amputees," *Front. Neurorob.* **2023**, *17*, 03.

[17] M. F. Wahid, R. Tafrershi, R. Langari, "A Multi-Window Majority Voting Strategy to Improve Hand Gesture Recognition Accuracies Using Electromyography Signal," *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *12*, 1.

[18] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. Lawrence Zitnick, *Computer Vision – ECCV 2014*. Lecture Notes in Computer Science (Eds: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars), Springer, Cham **2014**, pp. 740–755.

[19] Y. Li, S. Li, H. Du, L. Chen, D. Zhang, Y. Li, "Yolo-Acn: Focusing on Small Target and Occluded Object Detection," *IEEE access* **2020**, *8*, 227288.

[20] T. Feix, J. Romero, H.-B. Schmiedmayer, A. M. Dollar, D. Kragic, "The Grasp Taxonomy of Human Grasp Types," *IEEE Trans. Human-Machine Syst.* **2015**, *46*, 66.

[21] E. Stefanelli, M. Lapresa, F. Cordella, D. D'Accolti, C. Cipriani, L. Zollo, "A Hand-Wrist Control Strategy Based on Human Upper Limb Kinematics," in *2024 10th IEEE RAS/EMBS Int. Conf. for Biomedical Robotics and Biomechatronics (BioRob)*, IEEE, Heidelberg, Germany, September **2024**, pp. 1029–1034.

[22] S. Bruno, S. Lorenzo, V. Luigi, O. Giuseppe, *Robotics: Modelling, Planning and Control*, Vol. *1*, Springer **2010**.

[23] A. Cao, K. K. Chintamani, A. K. Pandya, R. Darin Ellis, "Nasa Tlx: Software for Assessing Subjective Mental Workload," *Behav. Res. Methods* **2009**, *41*, 113.

[24] A. Bangor, P. Kortum, J. Miller, "Determining what Individual Sus Scores Mean: Adding an Adjective Rating Scale," *J. Usability Stud.* **2009**, *4*, 114.

[25] N. Nourbakhsh, F. Chen, Y. Wang, R. A. Calvo, "Detecting Users' Cognitive Load by Galvanic Skin Response with Affective Interference," *ACM Trans. Interact. Intell. Syst.* **2017**, *7*, 12.

[26] G. Luzzani, I. Buraioli, D. Demarchi, G. Guglieri, "A Review of Physiological Measures for Mental Workload Assessment in Aviation: A State-Of-The-Art Review Of Mental Workload Physiological Assessment Methods in Human-Machine Interaction Analysis," *Aeronaut. J.* **2024**, *128*, 928.

[27] C. Tamantini, M. L. Cristofanelli, F. Fracasso, A. Umbrico, G. Cortellessa, A. Orlandini, F. Cordella, "Physiological Sensor Technologies in Workload Estimation: A Review," *IEEE Sensors J.* **2025**, *25*, 34298.

[28] K. Li, H. Rüdiger, T. Ziemssen, "Spectral Analysis of Heart Rate Variability: Time Window Matters," *Front. Neurol.* **2019**, *10*, 545.
[29] C. Tamantini, F. Cordella, N. L. Tagliamonte, I. Pecoraro, I. Pisotta, A. Bigioni, F. Tamburella, M. Lorusso, M. Molinari, L. Zollo, "A Data-Driven Fuzzy Logic Method for Psychophysiological Assessment: an Application to Exoskeleton-Assisted Walking," *IEEE Trans. Med. Rob. Bionic.* **2024**, *6*, 695.
[30] D. Novak, M. Mihelj, M. Munih, "A Survey of Methods for Data Fusion and System Adaptation Using Autonomic Nervous System Responses in Physiological Computing," *Interact. Comput.* **2012**, *24*, 154.
[31] F. Riillo, L. Quitadamo, F. Cavrini, E. Gruppioni, C. Pinto, N. Pastò, L. Sbernini, L. Albero, G. Saggio, "Optimization of emg-based Hand Gesture Recognition: Supervised vs. Unsupervised Data Preprocessing on Healthy Subjects and Transradial Amputees," *Biomed. Signal Process. Control* **2014**, *14*, 11.
[32] R. Cittadini, C. Tamantini, F. S. di Luzio, C. Lauretti, L. Zollo, F. Cordella, "Affective State Estimation Based on Russell's Model and Physiological Measurements," *Sci. Rep.* **2023**, *13*, 9786.
[33] J. Zhou, J. Y. Jung, F. Chen, "Dynamic Workload Adjustments in Human-Machine Systems Based on gsr Features," in *Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 Int. Conf.*, Bamberg, Germany, Springer **2015**, pp. 550–558.
[34] C. Tamantini, F. Cordella, C. Lauretti, L. Zollo, "The wgd—a Dataset of Assembly Line Working Gestures for Ergonomic Analysis and Work-Related Injuries Prevention," *Sensors* **2021**, *21*, 7600.