

Article

Bio-Adaptive Robot Control: Integrating Biometric Feedback and Gesture-Based Interfaces for Intuitive Human–Robot Interaction (HRI)

Antonio Di Tecco ^{1,*} , Daniele Leonardis ¹ , Edoardo Ragusa ² , Antonio Frisoli ¹  and Claudio Loconsole ^{1,3} 

¹ Department of Excellence in Robotics and AI, Sant’Anna School of Advanced Studies, 56127 Pisa, Italy; daniele.leonardis@santannapisa.it (D.L.); antonio.frisoli@santannapisa.it (A.F.); claudio.loconsole@unimercatorum.it (C.L.)

² Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture, University of Genoa, 16145 Genoa, Italy; edoardo.ragusa@unige.it

³ Department of Engineering and Science, Universitas Mercatorum, 00186 Rome, Italy

* Correspondence: antonio.ditecco@santannapisa.it

Abstract

AI-driven assistance can help the user perform complex teleoperated tasks, introduce autonomous patterns, or adapt the workbench to objects of interest. On the other hand, the level of assistance should be responsive to the user’s response and adapt accordingly to promote a positive and effective experience. Envisaging this final goal, this article investigates whether physiological signals can be used to estimate the user’s performance and response in a teleoperation setup, with and without AI-driven assistance. In more detail, a teleoperated pick-and-place task was performed with or without AI-driven assistance during the grasping phase. A deep-learning algorithm for affordance detection provided assistance, helping participants align the robotic hand with the target object. Physiological and kinematic data were measured and processed by machine learning models to predict the effects of AI assistance on task performance during teleoperation. Results showed that AI-driven assistance, as expected, affected pick-and-place performance. Beyond this, the assistance affected the participant’s fatigue level, which the machine learning models could predict with an average accuracy of 84% based on the physiological response. In addition, the success or failure of the pick-and-place task could be predicted with an average accuracy of 88%. These findings highlight the potential of integrating deep learning with biometric feedback and gesture-based control to create more intuitive and adaptive HRI systems.

Keywords: Human–Robot Interaction (HRI); bio-adaptive control; biometric feedback; gesture-based interfaces; teleoperation; affordance detection; deep learning; physiological signals; fatigue prediction; machine learning



Academic Editor: Girijesh Prasad

Received: 27 November 2025

Revised: 30 January 2026

Accepted: 9 February 2026

Published: 17 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The rapid advances in robotics and the growing demand for seamless human–robot collaboration have necessitated the development of innovative control mechanisms to bridge the gap between human cognition and machine operation. This paper explores the use of biometric feedback as a tool to estimate and monitor the user’s performance and response within a Human–Robot Interaction (HRI) scenario. The envisaged goal is to use this information to adapt the system to the user’s needs, leading to more natural and effective interaction.

Although intuitive interaction is a well-established concept in human–robot interaction, its interpretation varies significantly across the literature. In this work, we do not aim to provide a general or theoretical definition of intuitive interaction. Instead, we adopt a pragmatic, task-oriented perspective. Specifically, intuitive interaction is here understood as an interaction paradigm that enables users to perform teleoperated manipulation tasks with stable performance and reduced physical effort and fatigue, without requiring explicit mode switching or additional cognitive load. Examples from the literature on telemanipulation studies include [1], which addresses intuitive grasping pose assistance, and [2], which focuses on intuitive target-object reaching assistance. Accordingly, the main goal of this study is to investigate whether AI-driven affordance-based assistance, combined with physiological and kinematic feedback, can support more intuitive teleoperation by improving task performance while mitigating user fatigue. The feasibility of using biometric signals as indicators of user state and as a basis for adaptive assistance is also explored.

Integration of biometric feedback into robotic systems interfaced with humans has been a subject of growing interest in the field of HRI and teleoperation [3]. These signals can then be used to control the robot's movements, enabling a more natural and adaptive control mechanism [3,4]. In teleoperation, monitoring and assisting the operator can be pivotal due to the usability limitations that similar technologies still face. In particular, operators face difficulties due to altered or cluttered multi-sensory feedback involving vision, proprioception, and touch, all mediated by the teleoperation system. Visual feedback through screens or VR headsets may suffer from latency, limited field of view, or depth perception mismatches, which can affect coordination and situational awareness [5,6]. The absence of direct proprioceptive and tactile sensations further hampers the operator's ability to drive and perceive motion, contact, and force, increasing cognitive load and reducing precision [7]. These challenges become more pronounced as robotic systems grow in mechanical complexity and degrees of freedom, demanding greater mental effort from the operator to maintain control and performance [8–10].

Automated assistance for the operator can be introduced to facilitate task execution [11]. It can be developed and implemented in different ways, such as by automating parts of the task sequence or by adapting the workspace and trajectories to the task and target objects [12,13]. However, assistance inherently limits the user's agency, suggesting that its optimal level should be continuously adapted to the user's needs in order to improve the quality and effectiveness of teleoperation. Therefore, we investigated whether the biometric signals introduced above can be used to estimate the user's response and performance, thereby laying the foundation for using such data to effectively adapt assistance.

Several recent studies explored the potential of biometric feedback for HRI. Various signals, including physiological signals, speech, facial expressions, gaze, and gestures, have been investigated. Physiological signals can indicate cognitive load or fatigue, which is valuable for adaptive robot behavior. Monitoring of physiological signals and adaptation of the robot in collaborative environments have been proposed in [14] using EMG signals and in [15] using ECG and HRV. Other physiological signals were also used, such as the photoplethysmogram (PPG) in [16] or the Galvanic Skin Response (GSR) to estimate the user's level of engagement or attention, as in [17].

Measurement of physiological signals has also been facilitated by the more specific field of social robotics within the HRI domain. Ref. [18] reviewed the state of the art in emotion recognition for HRI, highlighting the significance of emotional intelligence in successful interactions. More specifically, ref. [19] discusses how physiological signals can be combined to achieve more robust emotion recognition, while [20] shows how physiological data can enhance robots' understanding of user intent and emotions, especially for

more natural interactions. The idea that various physiological signals can be integrated to provide a holistic view of the human body is proposed in [21].

Recent surveys consolidate the role of biometric signals in estimating user workload, highlighting the importance of multimodal data measurement and emphasizing feature engineering and supervised models [22]. Monitoring human behavior to improve HRI also extends to kinematic features, as shown in [23] for human–robot trust estimation and in [4] for human–robot collaborative tasks. Regarding teleoperation assistance, a widely investigated approach is collision avoidance, as in [24–26]. Recent works [27] propose more advanced system architectures that integrate vision-based affordance models, task typology classification, and visual interfaces to allow the user to confirm or abort proposed autonomous assistance. Interestingly, the developed system also highlights the need to ask the user whether autonomous intervention is desired. An extended study, taking into account both different assistance strategies (i.e., assistive autonomous actions, directional precision motion control, and larger motion mapping scaling) and methods to estimate the user’s behavior, evidencing how the most effective solutions vary with respect to the task and human factors [28].

Additional information to enhance interaction can be gained from advanced processing of the remote environment [29,30]. Indeed, by teleceptive sensing [31], i.e., sensing without contact, enables sensing information from the environment, for example, by using cameras. Visual Affordance Segmentation (VAS) divides an object into functional parts, allowing for the planning of fine-grained semiautonomous actions [32–34]. The State-of-the-Art approaches use Deep Neural Networks (DNNs) [33,35] to tackle this problem. However, the typical hardware resources available in real-world pipelines prevent real-time processing, which calls for custom approaches that balance hardware constraints and model capabilities [36]. A recent work [37], which uses a simplified version of the VAS problem specifically designed for semi-autonomous scenarios, achieved state-of-the-art performance on standard benchmarks while maintaining low computing requirements that meet the low-latency demands of teleoperation scenarios.

This work presents a teleoperation setup with affordance-driven assistance and investigates whether physiological and kinematic signals can be used to predict teleoperation performance. The main research questions of the presented study are as follows:

- **RQ1:** Can AI-driven robotic assistance effectively improve user performance (time, correct rate) in a teleoperation task?
- **RQ2:** Can biometric data provide a prediction of user performance and fatigue in a teleoperation task?

The study enrolled 30 participants in a teleoperation pick-and-place task and measured performance metrics such as success rate and execution time. The feasibility of estimating perceived fatigue from GSR and PPG features is also analyzed, supporting operator-aware teleoperation assistance.

2. Methods

This section presents the methodology for developing and evaluating a robotic teleoperation system for pick-and-place tasks, as shown in Figures 1 and 2. The system integrates a vision-based hand-tracking sensor for user input, an anthropomorphic robotic arm, a custom robotic hand, and an RGB-D camera, with an affordance-detection algorithm to enhance teleoperation capabilities. The experimental process involved a user study assessing the impact of AI-driven assistance on task performance and user experience. An overview of the setup and components is shown in Figures 3 and 4. Data acquisition and intelligent algorithm evaluation are also performed to analyze physiological responses and improve interaction efficiency.

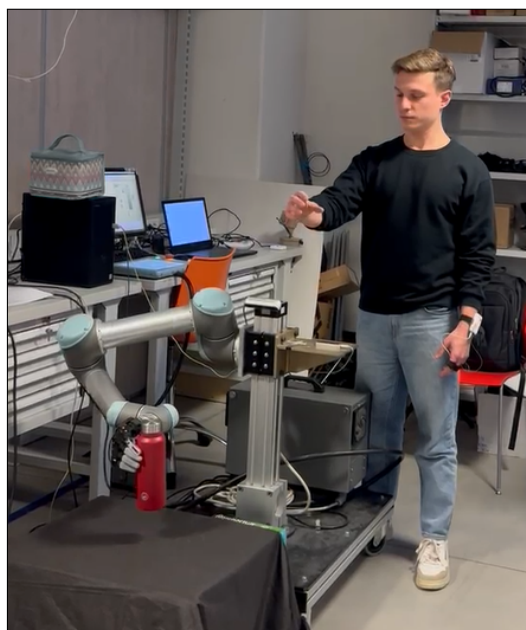


Figure 1. Experimental setup involving the robotic teleoperation pick-and-place task. The Leap Motion on the right for gesture tracking, whereas the participant wears the Shimmer3 GSR+ on the left for physiological acquisition.

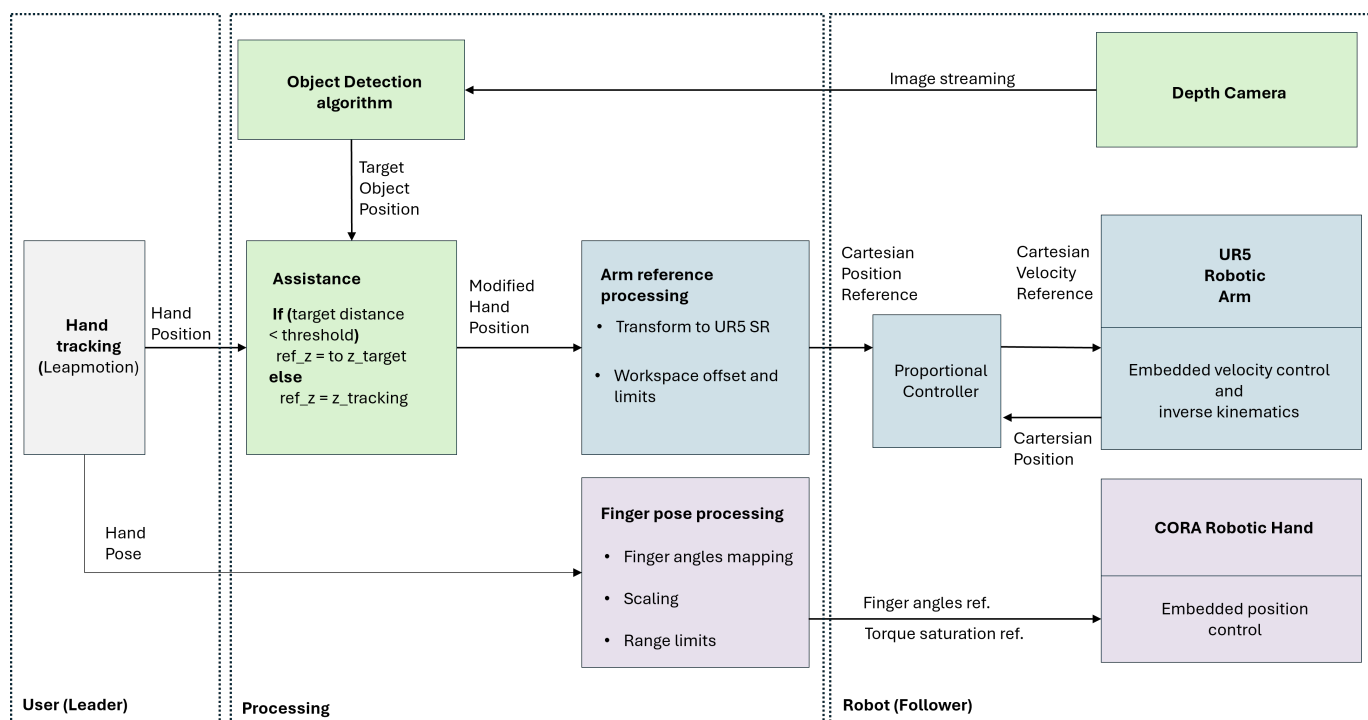


Figure 2. Scheme of the teleoperation setup.

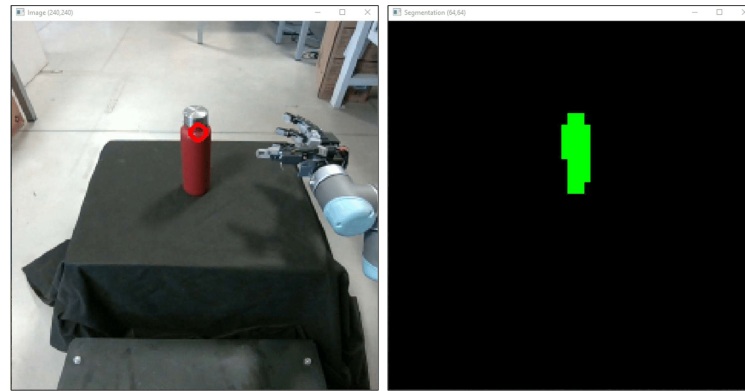


Figure 3. Camera view of the teleoperated scene and output of the vision-based affordance detection at the basis of the operator's assistance. In particular, the green background suggests regions to grasp, whereas the eventually blue background shows regions of the target object [38].

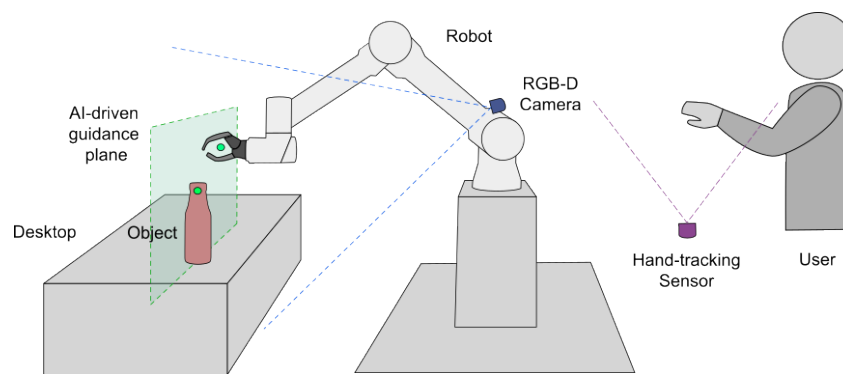


Figure 4. The diagram illustrates the experimental scenario, showing the relative positions of the camera device, the UR5 robotic arm, the object on the table, and the operator's viewpoint.

2.1. Robotic Teleoperation System for Pick-and-Place Tasks

A robotic teleoperation system is implemented to perform pick-and-place tasks performed using a robotic arm. Teleoperation is performed using vision-based position tracking without force feedback.

2.1.1. Robotic Arm

The implemented robotic arm is a Universal Robot UR5, a six-axis collaborative robot with a 5 kg payload, a reach of 850 mm (fully extended), and a repeatability of 0.06 mm. The UR5 features a lightweight aluminum structure and built-in safety checks on motor loads, enhancing safety during interaction with the environment. In the developed setup, the low-level velocity control mode provided by the UR5 is used to ensure smooth operation when commanded by a continuous stream of references from the operator side. The velocity reference is computed as proportional to the distance between the reference and the current hand position. References are processed in Cartesian coordinates, leaving the inverse-kinematics computation and joint control to the built-in control unit. The wrist orientation was kept constant in base coordinates to maintain an end-effector pose suitable for grasping objects on a horizontal, planar surface.

2.1.2. Robotic Hand

A redesigned version of the robotic hand developed in the laboratory, the CORA-hand [39], is used as the end effector. The hand features three degrees of freedom (DOFs) corresponding to the thumb abduction, index flexion, and combined flexion of the middle, ring, and little fingers. 3D printed flexible parts (material TPU 95A) in the finger links

ensured compliance of the hand to undesired contacts and forces, i.e., lateral forces with respect to the actuation plane of the fingers and normal forces acting on the closing direction of the fingers. The actuators are three Feetech STS3215 servo motors operated with the built-in, closed-loop position control. This control loop can be programmed to saturate the output voltage at a certain threshold, thereby introducing saturation of the maximum output torque in stall conditions (i.e., when holding a grasped object).

2.1.3. RGB-D Camera and the Affordance Detection Algorithm

To assist the operator, a RealSense D455 RGB-D camera (HFOV 87°, VFOV 58°, depth resolution 1280 × 720) was used with Intel Realsense SDK 2.0 (v2.55.1), combined with an artificial-vision algorithm. The camera was mounted on the robot base at 0.6 m from the center of the workbench and pointed 45° downward, as illustrated in Figure 1. A vision-based affordance detection algorithm has been implemented to extract the information needed to assist the operator. For the implementation of the affordance segmentation algorithm, a custom network architecture based on MobileNetV3, as proposed in [37], has been used. The architecture is developed in Python 3.12 using Keras and TensorFlow libraries. The training dataset comprises two well-known datasets: the UMD [40] and the Italian Institute of Technology (IIT) [41] datasets. The IIT dataset has 8835 images. These images differ in framing, occlusion levels, lighting conditions, and resolutions. The University of Maryland (UMD) dataset comprises 28,843 RGB-D images, with multiple views of each object, enabling an assessment of the models' ability to handle framing variations. A custom head that targets multi-task learning was used. It forced the model to distinguish the object from the background and discriminate graspable and non-graspable parts of the object using two dedicated outputs. Figure 5 shows the template architecture. The subset of connections to the backbone is selected to control the number of FLOPs and ensure good feature extraction capabilities. In the proposed implementation, the base blocks always contain a depthwise separable convolutional layer followed by a standard nearest upsampling layer. To optimize the network weights, a hierarchical version of the learning problem was proposed. The first segmentation head supports a binary object segmentation (OS) classification task. The computational graph includes an additional SH, which tackles the actual VAS problem. This component processes an input representation optimized to isolate the object and to focus on its distinctive parts, thus biasing the feature extraction process. The loss function (\mathcal{L}) used to train the affordance detection network is formalized as:

$$\mathcal{L} = \mathcal{L}_{seg} + \lambda \mathcal{L}_{aff} \quad (1)$$

where \mathcal{L}_{seg} is the binary cross-entropy loss for object segmentation (OS), \mathcal{L}_{aff} is the segmentation loss specific to the VAS task, and λ balances the relative weight of the two terms.

Both the losses measure the prediction error at the pixel level and can be formalized as:

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^N \frac{1}{|\Omega|} \sum_{j \in \Omega} \sum_{c=1}^C y_{n,j,c} \log \hat{y}_{n,j,c}, \quad (2)$$

where N is the size of the training set, $|\Omega|$ is the number of pixels of the output mask, C is the number of classes, y is the real value of the output mask, and \hat{y} is the prediction performed by the model.

This formula reflects a multi-task learning approach in which the network is trained to perform both object-background segmentation and affordance-based region detection. The \mathcal{L}_{seg} guides the shared encoder to extract generic object features, while \mathcal{L}_{aff} fine-tunes these features to identify graspable regions. This layered optimization allows the model to

maintain generalization while improving task-specific performance under computational constraints, as also supported by [37].

Using the above processing, the centroid of the affordance area is computed and matched with the camera's depth information, yielding the three-dimensional coordinates of the point most suitable for grasping the acquired object. This information is used to compute the robotic assistance intervention, as described in the following section.

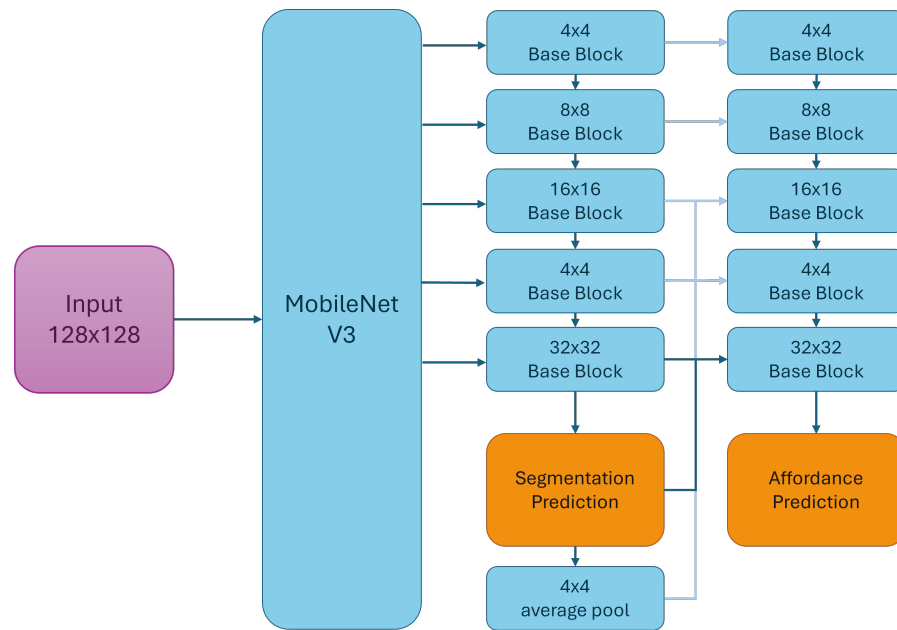


Figure 5. General scheme of the proposed network used in the Affordance Detection module [42].

2.1.4. Hand Tracking and Kinematic Data Processing

A hand-tracking, vision-based sensor (*Leap motion* controller by Ultraleap (SDK 2.3.1)) is used on the operator side.

The sensor is connected to a host PC to process the measured hand data and communicate with the follower side. Kinematic data processing, communication, and data recording are performed using a MATLAB Simulink model R2025B executed with a real-time pacing of 100 Hz. The sensor is placed on a horizontal support in front of the user, at hip level, facing upward.

The hand position measured by the sensor is normalized and scaled to pass from the operator workbench dimensions ($0.4 \times 0.4 \times 0.4$ m volume) centered in mid-air in front of the user (0.4 m vertical distance from the sensor), to the intended manipulation workbench at the follower side ($0.6 \times 0.6 \times 0.6$ m volume) set with lower limits according to the planar surface of a table. Then, references are transformed into the system of reference of the robotic arm base. To operate the robot in velocity control mode, the Simulink model closes a position control loop by comparing the reference position with the robot's current position. The position error is weighted, saturated, and sent to the robotic arm via UDP protocol as a Cartesian velocity reference for the end effector. The position of the palm, thumb, index, and middle fingers, as measured by the sensor, is used to compute the resulting thumb abduction, index flexion, and middle flexion angles. The angle of the index finger is determined using the palm's coordinates and reference system, and by projecting the index position onto the axis normal to the palm. A similar strategy is used for the middle finger, which controls all the middle, ring, and pink fingers on the CORA hand with a single DOF. For the thumb, the above projection is used to compute the abduction angle of the corresponding actuator on the CORA hand. These references are saturated according

to the robotic hand's limits and then sent to the robotic hand's control electronics through UDP communication.

2.1.5. Human and Robot Modes (AI vs. MAN)

The hand-tracking does not provide a binary open/close command. It drives (i) the robot end-effector from the hand position, and (ii) the robotic hand (thumb abduction, index flexion, and combined middle–ring–little flexion).

The operator guidance relies on the vision-based hand tracker only: no external tracking of the human arm segments (shoulder–elbow–wrist) is performed, as the control mapping uses the tracked hand position and finger pose to generate the end-effector reference. Human arm motion is unconstrained and serves only to place the hand within the sensor's field of view. The robot state (end-effector and joint signals) is acquired from the UR5 internal encoders. Therefore, two modes are defined and used. They are the AI-driven assistance modes based on machine-learning support, as described in Section 2.1.3, whereas the other mode does not use AI assistance (MAN). AI and MAN share the same input channels and sampling, differing by the absence of the depth-alignment cue in MAN.

When AI mode is enabled, the coordinates of the target object from the processed camera frames are forwarded to the Simulink model via the vision processing software named *Intelligent Processing System* [42]. Here, coordinates are transformed from the camera system of reference to the room coordinates. Thus, if the hand-tracked position is below a given distance from the target object (distance threshold set to 200 mm as absolute distance in the Cartesian space), the robot's reference position is then constrained on the z-axis (depth) to match the depth of the target object. References remained unconstrained on the vertical and lateral axes.

2.2. Experimental Procedure

The user study evaluates the effects of an AI-driven assistance mode on a human operator during a common teleoperation task.

Thirty participants (2 female, 28 male; age 30.13 ± 5.20) enrolled in this research study provided informed consent. Informed consent is used to invite participants to participate in the experiment, in accordance with the study approval from the *Joint Ethics Committee of the Scuola Normale Superiore and Sant'Anna School of Advanced Studies* (Aut. Prot. No. 62/2024 on 16 January 2025).

The experimental procedure is divided into five phases named *Welcome*, *Learning*, *Warm-up*, *Experiment*, and *Conclusion*, as follows:

1. During the *Welcome* phase, the participant is welcomed, and the experimental activity is explained (3 min).
2. The *Learning* phase follows (5 min): the supervisor teaches the participant how to move the robot's arm using the *Leap Motion* controller. During this phase, the participant observes how to pick and place the object of interest, having the robotic workbench in direct sight (1 m distance). The object of interest is a bottle, as illustrated in Figure 1.
3. The *Warm-up* phase follows. The participant dresses the *Shimmer3 GSR+* sensor on the left wrist, and learns how to move the robot's arm through the right hand to pick-and-place the object of interest (3–5 min).
4. After the *Experiment* phase begins. This phase is composed of two sessions. Each session is based on the AI or MAN modes, as written in Section 2.1.5. Each session consists of 5 task repetitions, with a total of 10 tasks. The order of sessions is inverted across participants (AI→MAN and MAN→AI) to minimize cognitive switching costs and allow familiarization within a session before transitioning. For each task, the *success* and *fault* pick-and-place outcomes and the participant's reply to a question are

noted according to predefined criteria; specifically, the *success* is recorded when the grasped object of interest is placed correctly on the workbench, while any deviation from correct placement is labeled as *fault*. Instead, the following question is asked to participants: “What is your perceived level of fatigue on a scale of 1 (lowest) to 5 (highest)?”. The previous notes were recorded 10 times for each participant.

5. When participants conclude both sessions, there is a *Conclusion* phase. The participant is thanked, and the next participant is invited.

The session duration is about 4 min. There is a session break of about 3 min between sessions. In conclusion, an experiment takes about 15 min per participant.

2.3. Data Signal

Different signals are collected using dedicated devices, including the *Shimmer3 GSR+* sensor, the *Leap Motion* controller, and the *Universal Robot UR5* robot, as shown in Figure 1.

The *Shimmer3 GSR+* is used to acquire electrodermal activity (GSR) and photoplethysmography (PPG) signals. This device is worn on the left wrist, not involved in robot control, to: (i) minimize motion artifacts caused by muscle contractions and grip adjustments in the controlling arm; (ii) preserve the naturalness of gestures on the control side; (iii) maintain stable electrode contact and more consistent PPG unaffected by rapid motion. Therefore, the *ConsensysPRO* software Rev1.6a is used to collect these signals from the *Shimmer GSR+* device.

Robust pose estimation requires continuous visibility of the right hand and wrist performing the gestures. Therefore, the *Leap Motion* sensor is positioned on the right side to maximize line of sight during the pick-and-place task and reduce occlusion of the wrist and fingers. Components of the robotic experimental setup, described in Section 2.1, are used to record the position of the user’s hand pose (hand tracking sensor), and the position of the robotic hand, pose, and fingers from the robot side. These signals are gathered throughout MATLAB’s Simulink model.

Data are stored in the *RobotSense25-DashGrasp Dataset* (see Data Availability Statement) for analysis and the development of machine learning models.

The gathered signals used in this study are presented in Table 1.

Table 1. Signals collected and used in the study.

Modality	Device	Signal Groups	Notation
Physiological	Shimmer3 GSR+ (left wrist)	GSR; PPG	Scalars
Human Hand Kinematics	Leap Motion (right hand)	HH: hand position + finger pose	x,y,z-axis/ angle
Robot Hand Position (global)	UR5 encoders	GRH: end-effector position in global/ base frame	x,y,z-axis
Robot Hand Position (robot frame)	UR5 encoders	RCH: end-effector position in robot frame	x,y,z-axis
Robot Hand Euclidean Coordinates	UR5 encoders	RECH: Euclidean distances of end-effector in robot frame	x,y,z-axis

2.4. Feature Extraction

Global and statistical features are extracted from each signal presented in previous subsection, as in [43], including head value (first sample of the signal), tail value (last

sample of the signal), maximum, minimum, ratio between the maximum and minimum values, difference between maximum and minimum value (range), root mean square, ratio between root mean square and median absolute deviation, mean absolute deviation, median absolute deviation, mean, median, mean of absolute values, mean square, power, geometric mean, harmonic mean, 10% trimmed mean, variance, standard deviation, 2nd-, 3rd, and 4th-order moments, 33rd-percentile, 1st-, and 3rd-quantile, interquartile, skewness, kurtosis, crest factor, clearance factor, impulse factor, peak value, and shape factor.

2.5. Datasets

The extracted features from the signals are used to define and create two datasets. The datasets, denoted as D_1 and D_2 , are used to study the two classification problems P_1 and P_2 , respectively.

The first problem (P_1) is to predict the participant's reply to the following question: "What is your perceived level of fatigue on a scale of 1 (lowest) to 5 (highest)?" The question was asked at the end of each repetition to record the participant's fatigue.

The second problem (P_2) is to predict whether the participant picked up and placed an object correctly (1 = fault) or incorrectly (2 = success). It is useful to predict and prevent an object from falling when grasped, for improving the robot's feedback.

These datasets are labeled with the labels of P_1 and P_2 , respectively. Each dataset contained 300 samples (30 participants \times 2 sessions \times 5 tasks), corresponding to the total number of experimental tasks. D_1 's samples are labeled in the range 1–5, whereas D_2 's samples are labeled between classes 1 and 2. Then, they are split into internal and external sets (90%/10%), with the same proportion of samples per class. Thus, 9 tasks per participant form the internal set and 1 task forms the external set; across all participants, this yields 270 internal and 30 external samples. This disjoint split prevents information leakage while preserving the personalized approach that the system targets for each operator. To reduce class imbalance within the dataset's internal set, the Synthetic Minority Over-sampling Technique (SMOTE) is applied [44]. This oversampling method generates new synthetic instances of the minority class by interpolating between existing samples in the feature space, thereby improving class balance before training the classification algorithm. This oversampling algorithm generates synthetic samples in the feature space, improving class balance before training the classification algorithm. The datasets are as presented in Table 2.

Table 2. Number of samples of the internal set without SMOTE and the external set for datasets D_1 and D_2 for classification problems P_1 and P_2 .

Class	Internal Set	External Set
(a)		
1	135	15
2	81	9
3	45	5
5	9	1
Total	270	30
(b)		
1	45	5
2	225	25
Total	270	30

2.6. Feature Selection

The internal training sets of D_1 and D_2 are used to find the optimal feature sets F_1 and F_2 , respectively. These feature sets are used to define the input to machine learning algorithms for solving problems P_1 and P_2 .

The Forward Sequential Feature Selection (FSFS) [45] is used to find the optimal feature set by minimizing the cross-entropy of an Artificial Neural Network (ANN). The ANN is chosen because it can learn complex linear and nonlinear input-output relationships [46].

A total of 100 FSFS executions are performed using stratified k -fold cross-validation on the training set, with $k = 10$, yielding 100 candidate feature sets. Each fold contains randomly selected samples; of course, folds do not share any samples.

A statistical evaluation leads to the most predictive feature set among the 100 candidates. The Student's t -test with a significance level (α) of 0.05 is used to determine the optimal feature set for each problem. In particular, Student's t -test is used to compare the means of the cross-validation errors generated by the ANN models using distinct candidate feature sets. The feature set that yields the highest number of null-hypothesis rejections is considered the most powerful for the classification task.

2.7. Algorithms Training and Evaluation

Many families of classifiers, such as Decision Trees, Discriminant Analysis, Naive Bayes, Logistic Regression, Support Vector Machines (SVMs), ANNs, k -Nearest Neighbors (k -NN), Kernels, and Ensembles, are considered for problems P_1 and P_2 . These families are used for evaluating architecture designs, also by using *Bayesian search* to fine-tune classifiers' hyperparameters. In particular, fine-tuned classifiers are named as *optimized* in this study.

In addition, different versions of these algorithms are designed. As reported in the scientific literature, these versions differ in architectural hyperparameters.

Performance metrics are measured to compare the algorithms to one another. These metrics are the mean value and standard deviation (Std) of accuracy, weighted precision, recall, and F1 score.

The algorithms are trained with the internal set and evaluated with the external set 30 times. The best classification models are selected based on the lowest test loss.

This analysis is performed on a personal computer equipped with an Intel i9-14900K CPU, 64 GB of DDR5 DRAM, and an NVIDIA GeForce RTX 4070 Ti GPU.

3. Results and Discussion

This section presents the experimental results for feature selection, signal importance analysis, and classifier performance evaluation, measured by classification accuracy, precision, recall, and F1-score, to identify the most effective models for each classification problem.

3.1. Selected Features

The feature sets F_1 and F_2 were selected via FSFS to define the input of models for problems P_1 and P_2 .

The features were extracted from physiological signals, such as GSR and PPG, as well as from robot and human hands. The hands' features have the same symbology used in Table 1. In addition, the features of the hands are distinguished by the first letter on the x -, y -, or z -axis. The feature's mean and standard deviation (Std) values are also specified. The feature sets are presented in Table 3.

Looking at the table, the feature sets F_1 and F_2 consist of 14 and 16 features, respectively. Moreover, while F_1 relies on physiological responses and a few kinematic features for the classification problem P_1 , F_2 requires more features to capture the complexity of the

classification problem P_2 . Thus, the importance of signal-specific metrics in both feature sets suggests that combining physiological and robotic kinematic signals provides a robust approach to classification, as further analyzed in the next section.

Table 3. Optimal feature sets F_1 (a) and F_2 (b).

(a)		
Signal	[Mean, Std]	Feature Set F_1
GSR	[5.8267, 5.9945]	Root Mean Square
PPG	[6.7874×10^6 , 3×10^6]	Mean Square
	[4.5898×10^4 , 8.3165×10^4]	4th-Moment
XGRHand	[0.1845, 0.0017]	Variance
YGRHand	[0.1845, 0.0691]	Range
	[0.0538, 0.0325]	Interquartile
XHHand	[0.2073, 0.9370]	Skewness
YHHand	[1.1796×10^4 , 1.4851×10^4]	3rd-Moment
ZHHand	[3.4949, 6.5435]	33rd-Percentile
XRCHand	[−0.6484, 0.0341]	Minimum
YRCHand	[0.2707, 0.0341]	Mean Abs. Deviation
ZRCHand	[0.0033, 0.0030]	3rd-Moment
XREHand	[2.2601, 0.4539]	Mean Square
YREHand	[−0.3930, 0.1851]	Minimum
(b)		
Signal	[Mean, Std]	Feature Set F_2
PPG	[2.743×10^3 , 3.5052×10^2]	Maximum
XGRHand	[0.1096, 0.0585]	Maximum
	[−0.0007, 0.0242]	3rd-Quantile
YGRHand	[0.0162, 0.0154]	Median Abs. Dev.
ZGRHand	[0.0001, 0.0005]	3rd-Quantile
XHHand	[1.906×10^3 , 1.5275×10^6]	4th-Moment
	[−4.2788, 5.1725]	Minimum
YHHand	[18.2186, 8.3759]	Standard Deviation
	[5.1199, 6.6678]	Kurtosis
	[60.3138, 20.8596]	Peak Value
ZHHand	[27.2655, 25.1983]	Interquartile
YRCHand	[0.07390, 0.0175]	Mean Square
ZRCHand	[0.0906, 0.0448]	Mean Abs. Deviation
	[0.1234, 0.0496]	Standard Deviation
	[0.0037, 0.0030]	3rd-Quantile
ZRECHand	[−0.8566, 0.2135]	Head Value

3.2. Signal Importance Analysis

Signals' predictive power was assessed using *feature permutation importance* (FPI) with optimal feature sets, as in [43]. In particular, each feature was assigned an *importance*

score 30 times, and the average and standard deviation were computed. Thus, the *average cumulative importance* (ACI) score was computed for each signal. Figure 6 shows bar plots with the ACI score related to each signal for each classification problem P_1 and P_2 . Looking at Figure 6a, the PPG signal has the highest ACI score, indicating that cardiovascular activity is crucial to distinguish classes. The GSR signal follows with a moderate importance score, reinforcing the relevance of physiological responses in the classification problem P_1 . Instead, hand signals, such as XGRHand, XHHand, XRCHand, and so forth, show lower ACI values, suggesting that their contribution is less significant than physiological signals. However, their use implies that they still provide complementary information that improves classification performance. In contrast, in Figure 6b, the YHHand signal stands out with a higher ACI score than the others, showing its important features in this classification problem P_2 . Regarding RQ2, the results suggest that kinematic information from the human hand is a key determinant of class distinctions in P_2 . Whereas the other signals, including physiological ones such as PPG, have lower ACI scores, indicating a shift in feature importance compared to P_1 . In addition, the broader feature set in P_2 compared to P_1 results in a more evenly distributed importance among robotic hand signals.

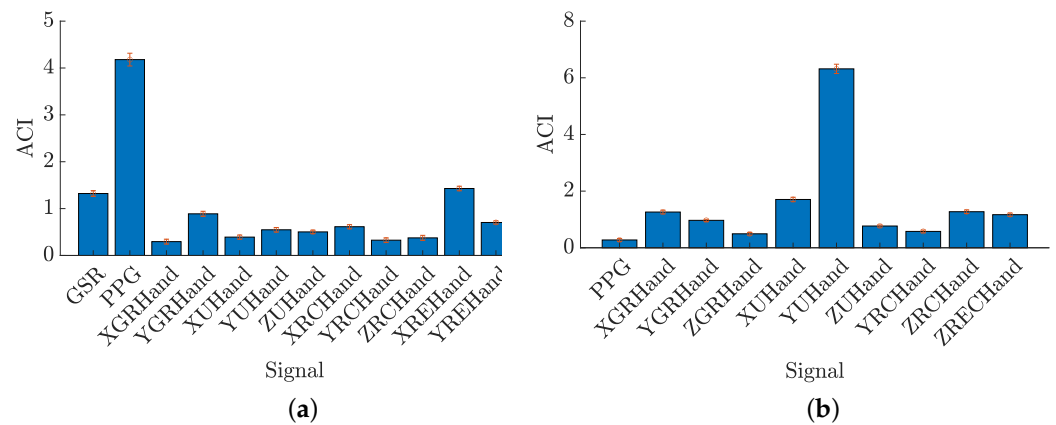


Figure 6. Average Cumulative Importance (ACI) score with a confidence interval of 95% for feature sets’ signals F_1 (a) and F_2 (b).

3.3. Performance Evaluation

The Table 4 shows the performance of algorithms evaluated on the external dataset. In particular, as shown in Table 4(a), the k -NN models achieved the highest accuracy, both reaching a level of accuracy higher than 84% with lower standard deviations, indicating stable performance across different evaluations. The Optimized k -NN model used these architecture hyperparameters: four neighbors, a city-block distance, and a distance weight of 1/inverse distance. Instead, the Boosted Trees showed an accuracy level of 80%, indicating that it may be more sensitive to variations in the dataset. Instead, looking at Table 4(b), the best model was the Medium ANN, achieving a level of accuracy of 86.13% and a weighted F1-score of 0.87, along with a strong balance between precision and recall. The Narrow ANN followed with an accuracy of 86% and an F1-score of 0.8688, showing similar stability but slightly lower performance. The Bilayered ANN had a lower accuracy of 84.80% and the highest standard deviation (0.0148), suggesting that deeper architectures may introduce more variability in classification performance.

Table 4. Performance of P_1 (a) and P_2 (b) best classifiers.

Model	Mean	Std	Precision	Recall	F1 (↓)
(a)					
Optimized k -NN	0.8800	~0	0.8625	0.8800	0.8618
Weighted k -NN	0.8400	0	0.8225	0.8400	0.8129
Boosted Trees	0.8000	~0	0.7701	0.8000	0.7788
(b)					
Medium ANN	0.8613	0.0120	0.8867	0.8613	0.8700
Narrow ANN	0.8600	0.0109	0.8858	0.8600	0.8688
Bilayered ANN	0.8480	0.0148	0.8803	0.8480	0.8586

3.4. Pick-and-Place Analysis

The task completion time and the success rate in grasping the object were analyzed to evaluate the effectiveness of AI-driven robotic arm guidance compared to manual guidance.

The results were that the median completion time was slightly lower in the AI-driven condition (42 s) than in the manual condition (44 s). However, the high standard deviation observed in both conditions, 44 s for AI-driven and 28.88 s for manual, suggests a significant variability in execution times.

Regarding the mean success rate in pick-and-place, the measured value is around 83% (see Table 2(b)). In particular, both methods obtained similar results, such as 1.82 points for AI-driven and 1.84 points for manual, where higher is better, with standard deviations of 0.38 and 0.36, respectively.

In response to RQ1, these data show that AI-driven guidance did not significantly improve grasping accuracy compared to manual guidance in the proposed experimental setup. While the trend suggests a minor improvement in performance under the AI-driven condition, the relatively high variability in user performance across both conditions precludes a statistically significant difference.

At this point, further investigations can consider longer training periods. Although the system was relatively easy to use, and a short adaptation phase appeared to be sufficient to perform the task, a considerably longer training period might help stabilize participants' performance with the system.

Results also indicate that integrating AI does not compromise accuracy, maintains reliability comparable to manual driving, and ensures stable and safe execution. However, improving the effectiveness of the AI-driven assistance can lead to significant, measurable improvements in performance, as discussed in the following Section.

3.5. System Optimization Strategies for Improved Efficiency

Although the AI-driven assistance provided a minor reduction in the median task completion time, such a difference was not statistically significant. Further optimization of the assistance, discussed hereafter, might lead to significant performance gains. On the other hand, the result may also align with other studies (e.g., [4]), suggesting that humans' rapid adaptation to a given task can intrinsically limit the effectiveness of intuitive assistance.

Several directions are identified for future system optimization, such as:

- *Anticipatory Motion Planning*: Integrating predictive trajectory generation based on affordance centroids could allow the robot to begin pre-aligning its grasp before the user's input stabilizes, reducing idle or correction phases during object approach.
- *Low-Latency Processing*: Offloading the vision and segmentation pipeline to a dedicated edge computing device (e.g., NVIDIA Jetson Orin or Coral TPU) may decrease end-to-end system latency, improving responsiveness.

- *Adaptive Control Strategies:* By incorporating user modeling based on physiological and behavioral data, the assistance level could be adapted dynamically. For example, users showing signs of increased effort may receive more assertive alignment guidance, whereas confident users could benefit from greater freedom of control.

Beyond these directions, additional steps could further optimize the system and substantially reduce task completion time. For instance, the integration of predictive user intent modeling, where machine learning models anticipate the operator's next action based on multimodal signals, including physiological and kinematic. This would enable the robot to prepare its motion trajectory in advance, reducing idle phases and corrective movements. Another step involves refining the affordance detection algorithm through continuous online learning, allowing the system to adapt to new objects or changing environmental conditions. Moreover, combining AI-driven assistance with haptic or multimodal feedback channels could accelerate decision-making during teleoperation by enhancing user situational awareness. Together, these improvements are expected to yield more consistent reductions in execution time while preserving safety and reliability.

These directions and steps will be evaluated in future experimental sessions, where task completion time (gesture recognition, data transmission, robotic movement, etc.) will also be compared across different optimization levels.

4. Limitations

The experimental results provide insights into bio-adaptive robotic teleoperation; however, some limitations must be acknowledged.

The current user study was conducted with a sample of 30 participants (2 females and 28 males). The sample size allowed control and validation of the feasibility of the proposed bio-adaptive teleoperation system under consistent conditions. However, while the included sample is relatively large for conventional teleoperation setups, it is not gender-balanced, and the sample size may still be limited for generalizing results.

In addition, the evaluation focused on a single task (pick-and-place) performed in a controlled environment with no external disturbances or environmental variability. While this reduces confounds and improves reproducibility, it does not assess robustness under dynamic or cluttered settings.

Moreover, although the affordance model was trained on large-scale datasets, the end-to-end evaluation used only an object. This choice supports clear and reproducible measurements and should be interpreted as a proof of concept under controlled conditions rather than as a cross-sectional evaluation.

However, to improve real application validity, future research might investigate the system's performance under different conditions, such as:

- dynamic task scenarios;
- a multi-object assessment (3–5 object categories with different geometries/materials, poses, and orientations);
- cognitive dual-task setups to simulate multitasking;
- complete NASA-TLX or similar questionnaires;
- collaborative settings with shared control or assistance negotiation.

In the current system implementation, the orientation of the robotic wrist was held fixed in the base reference frame throughout the task execution. This design choice aimed to simplify the control mapping between the user's hand and the robot's end effector, ensuring consistent grasping behavior on the planar workbench. However, this constraint reduced the robot's dexterity, limiting its ability to adapt to objects in non-standard orientations. To address this limitation, future developments will explore more wrist control strategies.

These approaches aim to restore full 6-DOF manipulation while preserving the system's intuitive use and reliability.

Moreover, the constraint of maintaining a fixed robot wrist orientation simplified control mapping and reduced cognitive load for novice users; however, it also limited the system's dexterity. In particular, this restriction reduced the robot's ability to adapt to objects in non-standard poses, resulting in suboptimal alignment during grasping in some cases. Participants' feedback confirmed that while the fixed wrist facilitated learning and ensured consistent behavior on planar surfaces, it occasionally hindered performance in edge-case scenarios. To overcome this limitation, alternative control schemes are currently being investigated. These include variable wrist orientation control for full 6-DOF teleoperation, as well as hybrid schemes in which AI-based affordance detection provides alignment cues while leaving fine orientation adjustments to the user. Such approaches are expected to improve both dexterity and user experience without compromising user intuitiveness.

5. Conclusions

This study introduced a robotic teleoperation system for pick-and-place tasks, leveraging a vision-based hand-tracking sensor for operator control and an affordance detection algorithm for AI-driven assistance. The system used a Universal Robot UR5 robotic arm, a custom-designed CORA-hand robotic end effector, and a Realsense D455 RGB-D camera to enable intuitive and efficient manipulation. The AI-driven assistance was implemented using a MobileNetV3-based neural network to enhance grasp prediction through multi-task learning. The experimental results from the user study demonstrated that AI-driven assistance impacted user activity. In particular, the system's effectiveness was measured through both physiological and kinematic signal analysis, which revealed that combining robotic and physiological data enabled robust classification models. In addition to the effects of the AI-driven assistance proposed in the scenario, the analysis showed that the intelligent algorithms successfully predicted user fatigue levels and object manipulation outcomes with over 84% accuracy, potentially contributing to the development of safer and more reliable teleoperation scenarios. In conclusion, the integration of human physiological signals, robotic control data, and AI-driven assistance represents a promising approach to enhancing teleoperation performance. Future work could explore real-time adaptive AI models to further refine assistance mechanisms and investigate the system's applicability to more complex manipulation tasks and real-world industrial settings. Expanding the study with a more diverse participant pool and additional experimental conditions could provide further insights into optimizing human–robot collaboration.

Author Contributions: Conceptualization, A.D.T., D.L. and C.L.; methodology, A.D.T., D.L., E.R. and C.L.; software, A.D.T., D.L. and E.R.; validation, A.D.T., D.L. and E.R.; formal analysis, A.D.T. and D.L.; investigation, A.D.T. and D.L.; resources, A.D.T., D.L. and E.R.; data curation, A.D.T. and D.L.; writing—original draft preparation, A.D.T., D.L. and E.R.; writing—review and editing, A.D.T., D.L. and E.R.; visualization, A.D.T., D.L. and E.R.; supervision, A.D.T., D.L., E.R., C.L. and A.F.; project administration, A.D.T., D.L., E.R., C.L. and A.F.; funding acquisition, C.L. and A.F. All authors have read and agreed to the published version of the manuscript.

Funding: This project has been funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.1—Call for tender No. 104 published on 2 February 2022, by the Italian Ministry of University and Research (MUR), funded by the European Union—NextGenerationEU—Project Title “AVATAR: Enhanced AI-enabled Avatar Robot for Remote Telepresence”—CUP J53D23000860006, D53D23001490008—Grant Assignment Decree No. 960 adopted on 30 June 2023 by the Italian MUR.



Institutional Review Board Statement: This study involved humans in its research. The Joint Bioethical Committee of Scuola Normale Superiore and Sant’Anna School of Advanced Studies approved all ethical and experimental procedures and protocols (Aut. Prot. No. 62/2024).

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: The data presented and used in this study are contained in the *RobotSense25-DashGrasp Dataset* available on Zenodo: <https://zenodo.org/records/15272120>, accessed on 12 February 2026.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bowman, M.; Zhang, J.; Zhang, X. Intent-based Task-Oriented Shared Control for Intuitive Telem Manipulation. *J. Intell. Robot. Syst.* **2024**, *110*, 167. [[CrossRef](#)]
2. Laghi, M.; Raiano, L.; Amadio, F.; Rollo, F.; Zunino, A.; Ajoudani, A. A target-guided telem Manipulation architecture for assisted grasping. *IEEE Robot. Autom. Lett.* **2022**, *7*, 8759–8766. [[CrossRef](#)]
3. Adawy, M.; Abualese, H.; El-Omari, N.K.T.; Alawadhi, A. Human–Robot Interaction (HRI) using Machine Learning (ML): A Survey and Taxonomy. *Int. J. Adv. Soft Comput. Its Appl.* **2024**, *16*. [[CrossRef](#)]
4. Ajoudani, A.; Zanchettin, A.M.; Ivaldi, S.; Albu-Schäffer, A.; Kosuge, K.; Khatib, O. Progress and Prospects of the Human–Robot Collaboration. *Auton. Robot.* **2018**, *42*, 957–975. [[CrossRef](#)]
5. Coelho, A.; Singh, H.; Kondak, K.; Ott, C. Whole-Body Bilateral Teleoperation of a Redundant Aerial Manipulator. In *Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA)*; IEEE: New York, NY, USA, 2020; pp. 9150–9156. [[CrossRef](#)]
6. Draper, J.V. Teleoperators for Advanced Manufacturing: Applications and Human Factors Challenges. *Int. J. Hum. Factors Manuf.* **1995**, *5*, 53–85. [[CrossRef](#)]
7. Shao, S.; Zhou, Q.; Liu, Z. Mental Workload Characteristics of Manipulator Teleoperators with Different Spatial Cognitive Abilities. *Int. J. Adv. Robot. Syst.* **2019**, *16*, 1729881419888042. [[CrossRef](#)]
8. Dafarra, S.; Pattacini, U.; Romualdi, G.; Rapetti, L.; Grieco, R.; Darvish, K.; Milani, G.; Valli, E.; Sorrentino, I.; Viceconte, P.M.; et al. icub3 Avatar System: Enabling Remote Fully Immersive Embodiment of Humanoid Robots. *Sci. Robot.* **2024**, *9*, eadh3834. [[CrossRef](#)]
9. Zhang, G.; Luo, X.; Zhang, L.; Li, W.; Wang, W.; Li, Q. A Framework of Indicators for Assessing Team Performance of Human–Robot Collaboration in Construction Projects. *Buildings* **2025**, *15*, 2734. [[CrossRef](#)]
10. Park, S.; Kim, J.; Lee, H.; Jo, M.; Gong, D.; Ju, D.; Won, D.; Kim, S.; Oh, J.; Jang, H.; et al. A whole-body integrated AVATAR system: Implementation of telepresence with intuitive control and immersive feedback. In *IEEE Robotics & Automation Magazine*; IEEE: New York, NY, USA, 2023.
11. Crandall, J.; Goodrich, M. Characterizing efficiency of human robot interaction: A case study of shared-control teleoperation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*; IEEE: New York, NY, USA, 2002; Volume 2, pp. 1290–1295. [[CrossRef](#)]
12. Luo, J.; Lin, Z.; Li, Y.; Yang, C. A Teleoperation Framework for Mobile Robots Based on Shared Control. *IEEE Robot. Autom. Lett.* **2020**, *5*, 377–384. [[CrossRef](#)]
13. Gottardi, A.; Tortora, S.; Tosello, E.; Menegatti, E. Shared Control in Robot Teleoperation With Improved Potential Fields. *IEEE Trans. Hum.-Mach. Syst.* **2022**, *52*, 410–422. [[CrossRef](#)]
14. Peternel, L.; Tsagarakis, N.; Caldwell, D.; Ajoudani, A. Robot adaptation to human physical fatigue in human–robot co-manipulation. *Auton. Robot.* **2018**, *42*, 1011–1021. [[CrossRef](#)]
15. Hopko, S.K.; Khurana, R.; Mehta, R.K.; Pagilla, P.R. Effect of cognitive fatigue, operator sex, and robot assistance on task performance metrics, workload, and situation awareness in Human–Robot collaboration. *IEEE Robot. Autom. Lett.* **2021**, *6*, 3049–3056. [[CrossRef](#)]
16. Savur, C.; Kumar, S.; Sahin, F. A Framework for Monitoring Human Physiological Response during Human Robot Collaborative Task. In *Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*; IEEE: New York, NY, USA, 2019; pp. 385–390. [[CrossRef](#)]
17. Li, K.; Wu, J.; Zhao, X.; Tan, M. Real-Time Human–Robot Interaction for a Service Robot based on 3D Human Activity Recognition and Human-Mimicking Decision Mechanism. In *Proceedings of the 2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*; IEEE: New York, NY, USA, 2018; pp. 498–503.
18. Spezialetti, M.; Placidi, G.; Rossi, S. Emotion Recognition for Human–Robot Interaction: Recent Advances and Future Perspectives. *Front. Robot. AI* **2020**, *7*, 532279. [[CrossRef](#)]

19. Abiri, R.; Zhao, X.; Heise, G.; Jiang, Y.; Abiri, F. Brain Computer Interface for Gesture Control of a Social Robot: An Offline Study. In *Proceedings of the 2017 Iranian Conference on Electrical Engineering (ICEE)*; IEEE: New York, NY, USA, 2017; pp. 113–117.
20. Kim, D.Y.; Lym, H.J.; Lee, H.; Lee, Y.J.; Kim, J.; Kim, M.G.; Baek, Y. Child-Centric Robot Dialogue Systems: Fine-Tuning Large Language Models for Better Utterance Understanding and Interaction. *Sensors* **2024**, *24*, 7939. [[CrossRef](#)]
21. Tapus, A.; Bandera, A.; Vazquez-Martin, R.; Calderita, L.V. Perceiving the Person and Their Interactions with the Others for Social Robotics—A Review. *Pattern Recognit. Lett.* **2019**, *118*, 3–13. [[CrossRef](#)]
22. Tamantini, C.; Cristofanelli, M.L.; Fracasso, F.; Umbrico, A.; Cortellessa, G.; Orlandini, A.; Cordella, F. Physiological Sensor Technologies in Workload Estimation: A Review. *IEEE Sens. J.* **2025**, *25*, 34298–34310. [[CrossRef](#)]
23. Campagna, G.; Lagomarsino, M.; Lorenzini, M.; Chrysostomou, D.; Rehm, M.; Ajoudani, A. Estimating Trust in Human–Robot Collaboration Through Behavioral Indicators and Explainability. *IEEE Robot. Autom. Lett.* **2025**, *10*, 10218–10225. [[CrossRef](#)]
24. Peng, S.; Cheng, X.; Yu, M.; Feng, X.; Geng, X.; Zhao, S.; Wang, P. Collision risk assessment and operation assistant strategy for teleoperation system. *Appl. Sci.* **2023**, *13*, 4109. [[CrossRef](#)]
25. Pan, M.; Li, J.; Yang, X.; Wang, S.; Pan, L.; Su, T.; Wang, Y.; Yang, Q.; Liang, K. Collision risk assessment and automatic obstacle avoidance strategy for teleoperation robots. *Comput. Ind. Eng.* **2022**, *169*, 108275. [[CrossRef](#)]
26. Zhang, X.; Zhong, Z.; Guan, W.; Pan, M.; Liang, K. Collision-risk assessment model for teleoperation robots considering acceleration. *IEEE Access* **2024**, *12*, 101756–101766. [[CrossRef](#)]
27. Penco, L.; Momose, K.; McCrory, S.; Anderson, D.; Kitchel, N.; Calvert, D.; Griffin, R.J. Mixed reality teleoperation assistance for direct control of humanoids. *IEEE Robot. Autom. Lett.* **2024**, *9*, 1937–1944. [[CrossRef](#)]
28. Lin, T.C.; Krishnan, A.U.; Li, Z. Perception and action augmentation for teleoperation assistance in freeform telemanipulation. *ACM Trans. Hum.-Robot Interact.* **2024**, *13*, 1–40. [[CrossRef](#)]
29. Castro, M.N.; Dosen, S. Continuous Semi-Autonomous Prosthesis Control Using a Depth Sensor on the Hand. *Front. Neurobotics* **2022**, *16*, 814973. [[CrossRef](#)]
30. Starke, J.; Weiner, P.; Crell, M.; Asfour, T. Semi-Autonomous Control of Prosthetic Hands based on Multimodal Sensing, Human Grasp Demonstration and User Intention. *Robot. Auton. Syst.* **2022**, *154*, 104123. [[CrossRef](#)]
31. Krausz, N.E.; Hargrove, L.J. A Survey of Teleceptive Sensing for Wearable Assistive Robotic Devices. *Sensors* **2019**, *19*, 5238. [[CrossRef](#)]
32. Xu, R.; Chu, F.J.; Tang, C.; Liu, W.; Vela, P.A. An Affordance Keypoint Detection Network for Robot Manipulation. *IEEE Robot. Autom. Lett.* **2021**, *6*, 2870–2877. [[CrossRef](#)]
33. Jiang, Z.; Zhu, Y.; Svetlik, M.; Fang, K.; Zhu, Y. Synergies between Affordance and Geometry: 6-DOF Grasp Detection via Implicit Representations. In *Proceedings of the Robotics: Science and Systems (RSS) Conference, Angeles, CA, USA, 12–16 July 2021*.
34. Wang, W.; Zhu, H.; Ang Jr, M.H. SGSIN: Simultaneous Grasp and Suction Inference Network via Attention-Based Affordance Learning. *IEEE Trans. Ind. Electron.* **2024**, *72*, 4990–5000. [[CrossRef](#)]
35. Nguyen, A.; Kanoulas, D.; Caldwell, D.G.; Tsagarakis, N.G. Object-based Affordances Detection with Convolutional Neural Networks and Dense Conditional Random Fields. In *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: New York, NY, USA, 2017; pp. 5908–5915.
36. Ragusa, E.; Dosen, S.; Zunino, R.; Gastaldo, P. Affordance Segmentation using Tiny Networks for Sensing Systems in Wearable Robotic Devices. *IEEE Sens. J.* **2023**, *23*, 23916–23926. [[CrossRef](#)]
37. Ragusa, E.; Canuti, G.P.; Lugani, S.; Zunino, R.; Gastaldo, P. Filling the Pareto Optimal Front for Affordance Segmentation on Embedded Devices using RGB-D Cameras. *IEEE Sens. J.* **2025**, *25*, 27467–27477. [[CrossRef](#)]
38. Di Tecco, A.; Leonardis, D.; Frisoli, A.; Loconsole, C. Grasping Task in Teleoperation: Impact of Virtual Dashboard on Task Quality and Effectiveness. *Robotics* **2025**, *14*, 92. [[CrossRef](#)]
39. Leonardis, D.; Frisoli, A. CORA Hand: A 3D Printed Robotic Hand Designed for Robustness and Compliance. *Meccanica* **2020**, *55*, 1623–1638. [[CrossRef](#)]
40. Myers, A.; Teo, C.L.; Fermüller, C.; Aloimonos, Y. Affordance detection of tool parts from geometric features. In *Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA)*; IEEE: New York, NY, USA, 2015; pp. 1374–1381.
41. Nguyen, A.; Kanoulas, D.; Caldwell, D.G.; Tsagarakis, N.G. Detecting object affordances with convolutional neural networks. In *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: New York, NY, USA, 2016; pp. 2765–2770.
42. Di Tecco, A.; Camardella, C.; Leonardis, D.; Loconsole, C.; Frisoli, A. Virtual Dashboard Design for Grasping Operations in Teleoperation Systems. In *Proceedings of the 2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE)*; IEEE: New York, NY, USA, 2024; pp. 994–999.
43. Di Tecco, A.; Frisoli, A.; Loconsole, C. Machine Learning for Predicting User Satisfaction in Human–Robot Interaction (HRI) Teleoperation Tasks. *IEEE Access* **2025**, *13*, 151567–151580. [[CrossRef](#)]
44. Udu, A.G.; Salman, M.T.; Ghalati, M.K.; Lecchini-Visintini, A.; Siddle, D.R.; Dong, H. Emerging SMOTE and GAN Variants for Data Augmentation in Imbalance Machine Learning Tasks: A Review. *IEEE Access* **2025**, *13*, 113838–113853. [[CrossRef](#)]

45. Jović, A.; Brkić, K.; Bogunović, N. A Review of Feature Selection Methods with Applications. In *Proceedings of the 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*; IEEE: New York, NY, USA, 2015; pp. 1200–1205. [[CrossRef](#)]
46. Montesinos López, O.A.; Montesinos López, A.; Crossa, J. Fundamentals of Artificial Neural Networks and Deep Learning. In *Multivariate Statistical Machine Learning Methods for Genomic Prediction*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 379–425.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.