Explorations of autonomous prosthetic grasping via proximity vision and deep learning

E. Mastinu, *Member, IEEE*, A. Coletti, J. van den Berg, and C. Cipriani, *Senior Member, IEEE*

***Abstract*— The traumatic loss of a hand is usually followed by significant psychological, functional and rehabilitation challenges. Even though much progress has been reached in the past decades, the prosthetic challenge of restoring the human hand functionality is still far from being achieved. Autonomous prosthetic hands showed promising results and wide potential benefit, a benefit that must be still explored and deployed. Here, we hypothesized that a combination of a radar sensor and a low-resolution time-of-flight camera can be sufficient for object recognition in both static and dynamic scenarios. To test this hypothesis, we analyzed via deep learning algorithms HANDdata, a human-object interaction dataset with particular focus on reach-to-grasp actions. Inference testing was also performed on unseen data purposely acquired. The analyses reported here, broken down to gradually increasing levels of complexity, showed a great potential of using such proximity sensors as alternative or complementary solution to standard camera-based systems. In particular, integrated and low-power radar can be a potential key technology for next generation intelligent and autonomous prostheses.**

***Index Terms*—autonomous, computer vision, deep learning, grasping, hand prosthesis, inertial, prosthetics, proximity, sensors.**

I. INTRODUCTION

THE traumatic loss of a hand is usually followed by significant psychological, functional and rehabilitation challenges. Engineers and researchers have, for a long time, made efforts to restore the functionality of a lost limb by developing prosthetic hands. The prosthetic hands currently available can be electrically powered and operated via myoelectric signals, or body-powered and operated via shoulder movements on the contralateral limb [1]. For those myoelectrically operated, the human-machine interface (HMI) relies on myoelectric (or electromyographic) sensors placed on the surface of the residual limb. Signals picked up from these sensors can be used to drive the prosthesis in a one-muscle-one-movement simplistic approach (e.g., flex biceps to close the hand). Despite several major and well-known challenges related to surface myoelectric signal acquisition (noise, motion artifacts, interface impedance changes due to humidity, temperature and pressure) [2], [3] whose inconsistencies make the grasp of an object hardly reliable and repeatable, this simple control approach proved to be somehow functional with basic prosthetic grippers, but failed to translate efficiently when modern multi-articulated (or multi-grasp) prosthetic hands reached the market (e.g., iLimb Ultra, Össur, Iceland and BeBionic, Ottobock, Germany). When operating multi-articulated prostheses, muscular contraction patterns are often used as the switching mechanism between different hand grasps (e.g., flex wrist three times to enable pinch grasp), killing the intuitiveness of the control. That is why in the past decades, researchers spent extensive efforts trying to relieve the amputee users from the burden of a non-intuitive HMI, moving the learning to the machine instead via artificial intelligence algorithms [4]–[11]. Thanks to those efforts, it is now widely accepted that machine learning algorithms applied to myoelectric signals can indeed facilitate the user when operating prosthetic hands with more than one degree-of-freedom, and this is confirmed also by the commercial interests behind this solution (e.g., Complete Control, COAPT, USA, and Myo Plus, Ottobock, Germany). However, the aforementioned challenges related to surface myoelectric signal acquisition still remain, latently contributing to disrupt the control experience.

Even though acceptance rate surveys surely need a more frequent update, they seem to unanimously capture strong trends of prosthesis rejection, sometimes as high as 40% [12], [13]. Obviously, the prosthetic challenge of restoring the human hand functionality and dexterity is still far from being achieved. Arguably, this is in part due to major challenges related to the acquisition of myoelectric signals from the surface of the skin, or to the total lack of tactile sensory feedback, challenges that coming-soon implanted solutions are successfully overcoming [14]–[16]. However, these implanted solutions are still under clinical investigation and will not reach the mass before a decade, and most importantly, they cannot provide a full answer to the complex problem of restoring the human hand functionality. Such articulated problem must be addressed in parallel from different directions: more intelligent hardware must be developed for the HMI as much as for the robotic prostheses. Unfortunately, the efforts spent so far on more intelligent and autonomous robotic hardware are far from being satisfactory, both from an engineering and a clinical perspective. The idea of semi-autonomous prosthetic hands was proposed decades ago and

Manuscript received November 7, 2023; revised January 16, 2024; and accepted February 21, 2024.

This work was supported by European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 101029946, and the Italian National PhD program in Artificial Intelligence, XXXVIII cycle, course on Health and Life Sciences, organized by Università Campus Bio-Medico di Roma.

This work involved human subjects. The ethical approval was provided by Ethical Committee of the Scuola Superiore Sant'Anna (ref. 12/2022).

E. Mastinu (corresponding author), A. Coletti and C. Cipriani are with the Artificial Hands Area, BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy. (e-mail: enzo.mastinu@santannapisa.it).

J. van den Berg was with Delft University of Technology, Delft, Netherlands.

previous work already attempted to address this need [17], [18]. Shape recognition for automatic grasp adaptation via cameras and image processing techniques was proven possible for industrial automation [19] and prosthetic purposes [20]–[22]. Došen *et Al.* explored object size and shape recognition via an RGB camera and a laser depth sensor located on a prosthetic hand [23]. Markovic, Mouchoux, *et Al.* explored similar purpose (i.e., object recognition for the identification of the most adequate prosthetic grasp) as well as automatic wrist orientation, via augmented reality glasses, depth and inertial sensors [24]–[26]. Automatic grasp selection, pre-shaping and wrist orientation was also recently pursued by Nobre Castro and Došen via an infrared depth camera placed on the dorsal side of a prosthetic hand [27]. Lastly, Starke *et Al.* recently presented a semi-autonomous control strategy for object recognition and grasp selection via RGB camera, inertial and distance sensors, directly interfaced with the user via a display and a single myoelectric channel, completely self-contained within the KIT robotic hand [28], [29].

Unfortunately, despite the promising results none of the proposed prosthetics solutions still has reached any clinical implementation. Arguably, this might be due to the difficult portability of computer vision solutions into a self-contained wearable system (i.e., too demanding hardware and data processing) and into the uncertainties of unconstrained environments. Therefore, it still remains a wide potential benefit in exploring other integrating and concurrent solutions pushing towards prosthetic hands which are semi (or potentially even completely) autonomous from the conventional myoelectric HMI. To this goal, a variety of proximity sensors can be further explored [30]–[32]. For instance, radars proved high potential for an accurate recognition of materials and body parts [33], and they are progressively becoming essential in fields such as automotive [34], civil engineering [35], contactless vital sign monitoring systems [36], [37] and novel human-machine interfaces [38], [39].

Here, in an attempt to contribute to the field, we hypothesized that a combination of a radar sensor and a low-resolution time-of-flight camera can provide sufficient data for autonomous control approaches, in particular for shapes and materials recognition in both static and dynamic scenarios. To test this hypothesis, we analyzed via deep learning algorithms a human-object interaction dataset (HANDdata [40]) comprising of first-person data recorded from a variety of sensors, including proximity (i.e., state-of-the-art radar and time-of-flight sensors), inertial, load cells and fingers stretch. This dataset includes almost 6000 human-object interactions focused on the reach-to-grasp action performed by 29 different able-bodied individuals, with 10 standardized objects of 5 different shapes and 2 kinds of materials. Moreover, further inference testing was performed on unseen data that was purposely collected following HANDdata original protocol. The analyses reported here, broken down to gradually increasing levels of complexity, showed a great potential of using such proximity sensors as alternative or complementary solution to standard camera-based systems. In particular, integrated and low-power radar can be a potential key technology for next generation intelligent and autonomous prostheses, and perhaps even for other applications in healthcare, social (e.g., mobile servant) and industrial (e.g., warehouse robots) robotics.

II. Materials and Methods

This study aims to investigate the feasibility of using state-of-the-art proximity sensors, such as a radar and a time-of-flight camera, for object recognition in the context of autonomous prosthetic hands. To this aim, a dataset of human-object interactions was analyzed via deep learning. In order to investigate which sensor can be better suited for the aimed goal of recognizing the target object, all analyses were performed in 3 conditions: 1) radar data alone (RAD), 2) time-of-flight data alone (TOF), and 3) a combination of radar and time-of-flight data (RADTOF). Additionally, to further investigate the grasping scenarios, a fourth data condition was defined with the addition of the inertial sensor (RADTOFIMU). This last condition was deemed helpful to explore effects of fusion of data from both the environment and the participant's behavior.

A. Human-object interactions dataset

Most of the analyses reported in the following were performed on the HANDdata dataset [40], a data collection specifically tailored for autonomous grasping of a robotic hand and with particular attention to the reaching phase. This dataset included almost 6000 human-object interactions recorded via radar and time-of-flight sensors mounted on the forearm of able-bodied participants. These interactions focused on one of the most representative actions of daily life object manipulation, namely the reach to grasp. All details about the dataset are included in its dedicated report, but we briefly summarize the most important ones here for readers' convenience.

29 healthy adults participated to the data collection. Participants were asked to reach-grasp-lift-and-replace different objects while wearing an instrumented glove (Figure 1, left). The glove was instrumented so to track fingers stretching, kinematics of the forearm, and proximity measurements of the target objects. Forearm kinematics was tracked via 3-axes accelerometer and gyroscope of the BMI160 (Bosch, Germany) inertial sensor acquiring at 120 samples per second. The proximity sensors used were a pulsed coherent radar (A121, Acconeer, Sweden) and time-of-flight sensor (VL53L5CX, STmicroelectronics, France-Italy). The A121 radar was set to emit trains of 60 GHz pulses with known starting phase in pulsing/silent cycles alternating at 13 MHz. Knowing the velocity of the emitted pulses, the radar then tried to reconstruct reflections at certain discrete distances by analyzing the time taken to receive the echoes. In the configuration used in this study, the radar received pulses at times related to the distances of 35 range points, equally distributed from 6 cm to 41 cm (i.e., each range point is separated by 1 cm). Then, a complete reading of all range points, also defined as a sweep, was repeated 40 times with a rate of 800 Hz. Lastly, these 40 sweeps were collected in frames (i.e., frame size 40x35) and acquired with a rate of 15 frames per second. The radar field of view covered a 3D region of approximately 65x53 degrees at 50% of beam

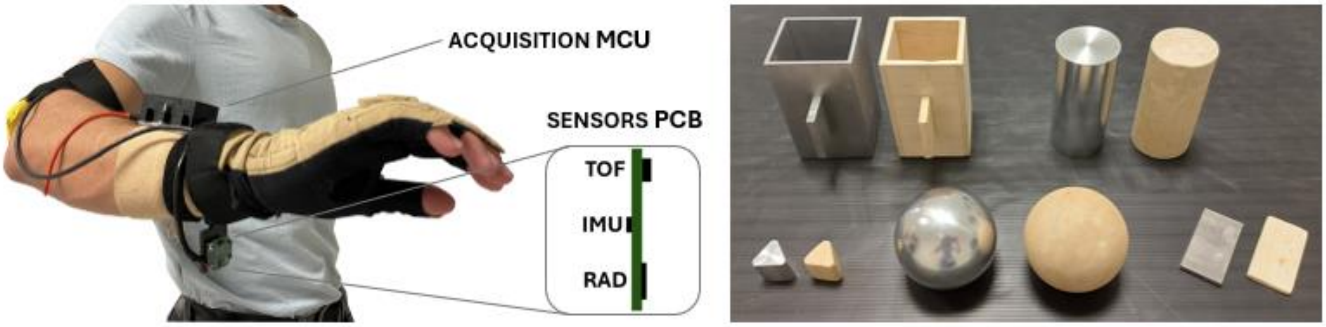

Figure 1. Instrumented glove (left) and target objects (right). The dataset analyzed in this study includes human-object interaction data acquired from 29 participants manipulating 10 different objects while wearing an instrumented glove. The instrumented glove (left) was based on a CyberGlove to which proximity and inertial sensors were added via a custom circuit board mounted on a customizable wrist band. Ten different target objects were used (right), each linked to a particular hand grasp. Abstract objects like a sphere, a cylinder, a triangular prism, a cuboid with a thin rectangular prism 'handle' and a thin rectangular prism were meant to trigger the spherical, power, tri-digit, lateral and pinch grip pattern, respectively.

power. The VL53L5CX time-of-flight sensor tried to estimate target's distance by illuminating the region of interest with 940 nm photons (i.e., invisible light). It was set for 64 pixels resolution at 15 frames per second (i.e., frame size 8x8). Its field of view covered a 3D region of approximately 45x45 degrees at 75% of beam power.

Ten different target objects were used in the data collection protocol and each object was meant to trigger a certain grasp pattern (Figure 1, right). Specifically, abstract objects like a sphere, a cylinder, a triangular prism, a cuboid with a thin rectangular prism 'handle' and a thin rectangular prism were meant to trigger the spherical, power, tri-digit, lateral and pinch grip pattern, respectively. Each object was available in two materials, wood or aluminium. Objects dimensions, materials and weights are standardized within the Southampton Hand Assessment Procedure [41], a clinically validated functional assessment, well-established in the field of prosthetics.

The dataset included human-object interactions from four different scenarios with gradually increasing complexity (for explicative figures and further details check HANDdata dedicated publication [40]), namely:

1) Bench-static, ideal scenario in which the proximity sensors were fixed in a clamp and steadily facing the target objects at a fixed distance and position.
2) User-static, static scenario in which the participants were steadily standing in front of the target objects with the proximity sensors partially facing the targets.
3) Pick-and-Lift, dynamic scenario in which the participants were reaching the target, to pick it, lift it about 10 cm, and then reposition it on the same start-area.
4) Pick-Lift-and-Move, dynamic scenario in which participants were reaching the target, to pick it, lift it about 10 cm while transporting it to a land-area different from the start-area, 40 cm away. Start- and land-area were alternated at each trial ultimately providing data for pick-and-lift from two different approach directions.

For user-static and grasping scenarios, the participants were asked to start and end each trial with the arm in rest position. The rest position was defined as the upper arm adjacent to the body trunk, with the elbow joint bent at 90 degrees, and with the hand palm perpendicular to the floor. Moreover, the subject's alignment in respect to the target object changed depending on the scenario. For the user-static and pick-and-lift scenarios, the subject's arm in rest position was aligned on the single area of interest where the target objects were located (i.e., left instrumented platform). Instead, for the pick-lift-and-move scenario, the subject's arm in rest position was aligned with the center of the two start and land areas (i.e., in between the left and right instrumented platforms) causing the target objects to be poorly or not in view at movement start.

All target objects were acquired in all scenarios. Moreover, "no-object" acquisitions were included for the static scenarios. These were achieved for the bench-static scenario by having no target under the proximity sensors, and for the user-static by having no hit within 1 meter range. Unfortunately, HANDdata did not include "reaching to no target" trials for the grasping scenarios.

B. Deep Learning models

Deep learning models were trained and tested so to assess the feasibility of shapes and materials recognition via proximity sensors. The models consisted of a convolutional neural network (CNN) for the RAD and TOF conditions, and a mixed-input neural network for the RADTOF condition. All models were implemented with TensorFlow framework, trained and tested via GPU (T4, Nvidia, USA). The models' architectures and hyperparameters were found via preliminary testing which involved also two steps random search via Keras tuner for number of layers and units, for dropout, activations, and learning rate, over 100 trials with 3 executions each.

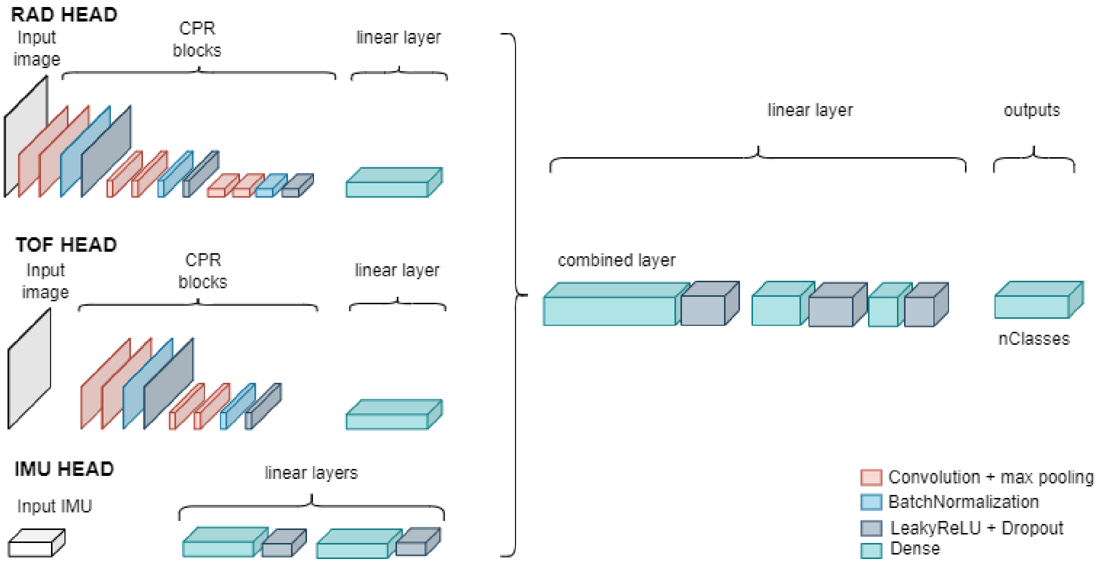

Figure 2. Illustration of the architecture of the mixed-input neural network used for the combination of radar (RAD), time-of-flight (TOF) and inertial sensors (i.e., RADTOFIMU). The single RAD and TOF CNNs are depicted in their respective branch of the mixed-input net.

CNN

The CNNs consisted of sequential CPR blocks (Figure 2) including four layers (convolution 3x3 kernel + max-pooling + batch-normalization + leaky-ReLU activation). A regularization dropout layer with rate 0.25 was attached after the activation function in each CPR block. The number of CPR blocks was defined as three for the RAD condition and two for the TOF condition. For the RAD condition, the CNN's input was defined as 64x35 2-channels images composed of real and imaginary parts of the 64-points Fast Fourier Transform of each radar's frame along the sweeps dimension (i.e., doppler map). For the TOF condition, the CNN's input was defined as 8x8 1-channel images composed of the sensor's raw frame. For the CNNs output, a final dense layer allowed to adapt each model to each particular classification problem.

Mixed-Input NN

For the RADTOF condition (i.e., the combination of radar and time-of-flight data), a mixed-input neural network was used (Figure 2) composed by two CNN input branches, respectively for RAD and TOF data, defined as above, followed by four linear layers with leaky-ReLU activation functions. Two dropout layers with rate 0.5 were attached to the second and to the fourth linear layer. For the output, a final dense layer allowed to adapt the model to each particular classification problem.

In order to further investigate the grasping scenarios, another data condition was defined with the addition of the inertial sensor. For this condition, namely RADTOFIMU, a third branch was added to the mixed-input neural network (Figure 2) composed by two linear layers with leaky-ReLU activation functions and dropout layers with rate 0.1. The input of this branch was defined as 6x1 (3-axes accelerometer and gyroscope) and fed with samples of the inertial sensor which were closest in time to the radar and time-of-flight frames.

C. Training and offline testing

For each of the deep learning model, a similar procedure was followed for the training and testing. Each dataset was shuffled and split between train, validation and test data sets taking respectively 80%, 10% and 10% of the data, ensuring balance among the different classes' samples. Networks were trained with the Adam stochastic optimizer up to 200 epochs, with batch size of 32 and learning rate 0.001. Then, at the end of the training epochs the network was further tested on the test_set data, from which the accuracy was computed. Test accuracy is reported as raw, thus no post-classification filtering was applied.

The data considered for the training and offline testing was:

- All available frames for bench- and user-static scenarios,
- frames relevant to the reaching motion for pick-and-lift grasping scenario, thus all frames included in the time range from -0.6 to -0.2 seconds before contact with the object (i.e., from -9 to -3 frames).

Importantly, no data from the pick-lift-and-move scenario was used for networks training nor for offline testing. We deemed more relevant to the explorative narrative of this study to leave that scenario for the inference testing part to showcase performances at worst case.

The instant of contact with the object was approximated from the inertial sensor data as the time instant of the first maxima in the z-axis (i.e., the axis parallel to the reaching direction), thus the instant of zero in the acceleration towards the target object. A simple sanity check was performed on pick-and-lift data discarding trials that were too fast (1.6% discarded data, 45 out of 2900 trials).

For the user-static and grasping scenarios, the training and testing was performed considering:

- aggregated data, thus from the 29 participants together,

- individual data, thus only relevant to each participant. Individual results are reported with the format MED:IQR (median: interquartile-range).

Even though aggregating the participants' data was our main strategy (see inference testing), we intended to include here also subject-specific explorations so to provide an interesting glance on alternative approaches tailored on each end-user.

The classification problems were defined according to the scenario of interest so to showcase the potential for recognizing different objects and object properties in static situations or even during the reaching motion (TABLE 1).

TABLE 1.
CLASSIFICATION PROBLEMS DEFINED
FOR TRAINING, OFFLINE AND INFERENCE TESTING

Scenario	Problem	# classes	No-object class?	Only metal?
Bench-static	Which object?	11	Yes	No
	Which shape?	6	Yes	No
	Which material?	3	Yes	No
User-static	Which object?	11	Yes	No
	Which shape?	6	Yes	Yes
	Which grasp?	3	Yes	Yes
Pick-and-Lift **Pick-Lift-and-Move**	Which object?	10	No	No
	Which shape?	5	No	Yes
	Which grasp?	2	No	Yes

Materials recognition was explicitly defined only for bench-static, however it was implicitly reported also for the other scenarios by the "which object?" classification problems. We considered more relevant to the explorative narrative to provide results about this worst-case. It is important to note that some explorations (i.e., "which grasp?" classification problems) ultimately aimed to demonstrate potential for the autonomous selection of the most adequate hand prosthesis grasp for manipulating a certain object. To this aim, we intended to keep these explorations tied to a realistic scenario of a modern multi-functional prosthetic hand with a limited selection of grasps, namely power and pinch grasps, a common clinical setup for prosthesis users.

A last "no-object" class was always added to classification problems related to the static scenarios of bench- and user-static, aiming to explore the capability to recognize the no-target situation in static conditions. Such ability could arguably stabilize the autonomous control and avoid misclassifications before any reaching motion would take place. Moreover, considering the preliminary investigation and demonstrative purposes of this study, some analyses regarding the user-static and grasping scenarios were performed considering only the objects made of metal (TABLE 1). This was deemed necessary due to the limited permittivity of the wood (i.e., high transparency to the radar) so to limit inherent obvious advantages between different data conditions and focus the exploration on object recognition during motion.

D. Inference: testing with new collected data

All models trained with the aggregated data from all 29 participants were further explored via inference testing, thus by measuring classification accuracy on unseen data not part of the original dataset. For this, new data was collected according to the particular scenario of interest trying to replicate as close as possible the original data acquisition protocol [40]. In particular, for inference testing on the bench-static scenario, new data was acquired for each object for two different orientations (1st and 2nd from the bench-static protocol), for about 6 seconds each. Then, for inference testing on user-static, grasping pick-and-lift and pick-lift-and-move scenarios, a new data collection was performed with a participant that was already part of the original dataset. However, CyberGlove and platforms data were not acquired as in the original protocol.

The participant signed informed consent for data and media acquisition and public release. The ethical approval was provided by Ethical Committee of the Scuola Superiore Sant'Anna (ref. 12/2022).

The actual inference testing was performed on Matlab (Mathworks, USA) by evaluating the TensorFlow nets exported in *.h5 format on the freshly acquired data. Specifically, the data considered for the inference testing was:

- Last 60 acquired frames (about 4 seconds) for each object for bench- and user-static scenarios,
- All frames included in the time range from -0.6 to -0.2 seconds before contact with the object (i.e., from -9 to -3 frames) for pick-and-lift and pick-lift-and-move grasping scenarios.

The classification problems were as defined in TABLE 1. Only for the grasping pick-and-lift and pick-lift-and-move scenarios a mode post-classification filter was applied, therefore considering as final prediction the most frequent predicted class from the analyzed frames related to each different trial.

E. Statistical analysis

A correlation analysis was performed between participants' height and resulting accuracies in static-user scenario (*corrcoef* function available on Matlab). No further statistical analysis was performed on the data.

III. RESULTS

A. Offline analysis

Bench-static: can we recognize the object on the bench?

An illustrative sample of the proximity sensors images provided to the CNNs for the bench-static scenario is shown in Figure 3.

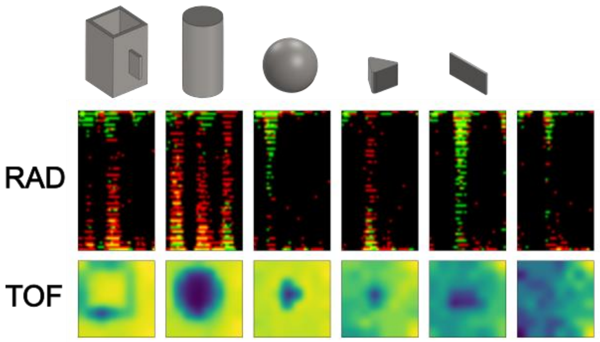

Figure 3. Images from the proximity sensors data. Visualization of the data acquired from the radar and time-of-flight sensors while steadily facing the different target objects. For illustration purposes, radar images were plotted as RGB images with blue channel set to zero, and time-of-flight images were plotted as pseudo-colour images with "viridis" colour mapping. Lastly, all images were smoothed via Gaussian interpolation.

Results from the bench-static scenario (Figure 4) showed perfect accuracies for all objects (100%), shapes (100%) and materials (100%) recognition with the radar data. Accuracies were lower with the TOF data for all objects (81.92%), shapes (69.43%) and materials (92.86%) recognition. Interestingly, the combination of radar and TOF data allowed top accuracies as the ones achieved with the radar data, resulting in 100% accuracy for all objects, shapes and materials recognition.

User-static: can we recognize the object from the user?

High recognition accuracies were found also in the user-static scenario when considering all participants together and RAD and RADTOF conditions (Figure 4). Indeed, when using radar data alone the offline accuracies were 95.42%, 98.62% and 97.34% for all objects, shapes and grasps recognition, respectively. For RADTOF, accuracies were 95.12%, 99.17% and 99.36% for objects, shapes and grasps. Lastly for TOF, accuracies were 59.05%, 75.41% and 90.22% for all objects, shapes and grasps recognition, respectively.

For the individual analysis, thus when considering data only relevant to each participant, RAD condition clearly outperformed the other two conditions. Indeed, average accuracies with radar data alone were 86.67:43.08%, 100:0.16% and 97.26:31.66% for all objects, shapes and grasps recognition, respectively. For TOF condition, accuracies were 66.67:43.97%, 75.00:35.42% and 91.30:15.88% for objects, shapes and grasps. Lastly, when combining the data from radar and time-of-flight sensors (i.e., RADTOF), offline accuracies were 65.52:39.69%, 85.42:50.00% and 92.11:45.01% for all objects, shapes and grasps recognition, respectively.

In general, the TOF sensor proved to be more sensitive to the less optimal centering of the object in its field of view, demonstrated also by the significant negative correlation between participants' height and resulting accuracies (p=0.03).

Pick-and-Lift: can we recognize the object while moving towards it?

Accuracies certainly dropped when analyzing the dynamic scenario of the user reaching to grasp the object. Even though results presented quite some variability, they show a trend of TOF outperforming the other data conditions in both aggregated and individual analyses (Figure 4). Specifically about the aggregated analysis, when using time-of-flight data alone the accuracies were 74.81%, 92.10% and 96.97% for all objects, shapes and grasps recognition, respectively. When using radar data alone the accuracies were considerably lower, to 36.21%, 62.58% and 66.67% for all objects, shapes and grasps. Recognition accuracies benefitted from considering more sensors together. Indeed, for RADTOF accuracies were 77.14%, 62.15% and 92.90% for all objects, shapes and

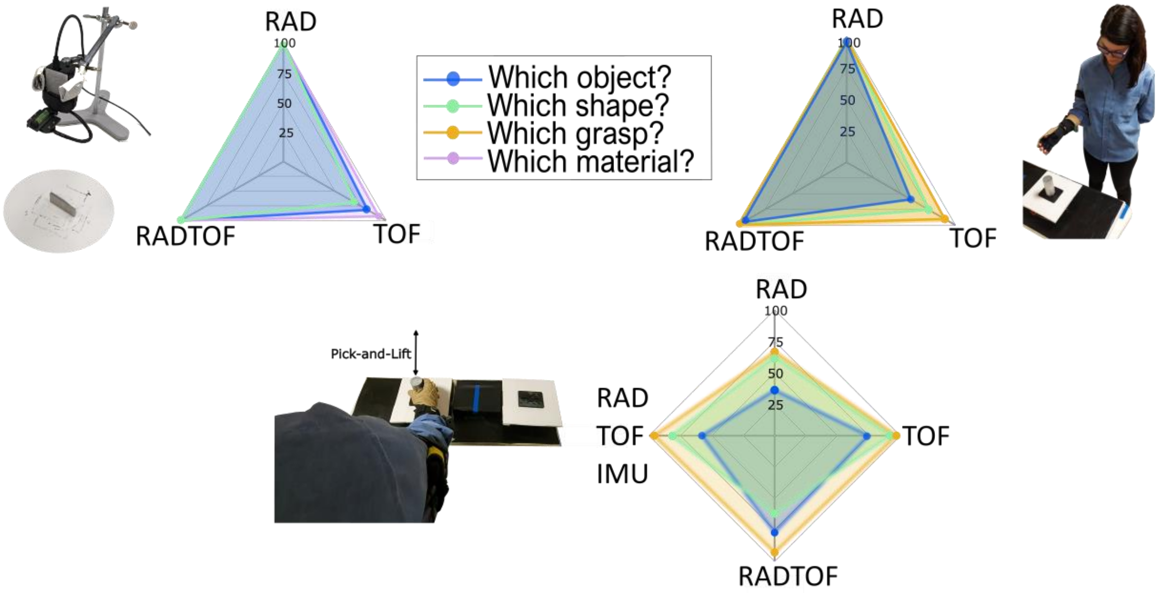

Figure 4. Offline test accuracies for the different scenarios, data conditions and classification problems. The data conditions were defined as RAD when considering radar data alone, as TOF when considering time-of-flight data alone, as RADTOF when considering a combination of radar and time-of-flight, and as RADTOFIMU when considering also the inertial sensor. The scenarios were bench-static (top-left), user-static (top-right) and pick-and-lift (bottom). The different classification problems are explained in the spiderplots legend.

TABLE 2.
INFERENCE ACCURACIES (%)

	Bench-static			User-static			Pick-and-Lift		
	Which Object?	Which shape?	Which material?	Which object?	Which shape?	Which grasp?	Which object?	Which shape?	Which grasp?
RAD	66.67	68.18	99.85	27.58	59.72	82.22	37.00	58.00	92.00
TOF	49.55	51.21	71.52	24.39	13.06	16.39	17.00	60.00	72.00
RADTOF	86.21	93.79	99.85	27.27	55.83	66.67	50.00	24.00	94.00
RADTOFIMU	/	/	/	/	/	/	34.00	76.00	82.00

grasps. Lastly, the addition of inertial sensor data (i.e., RADTOFIMU) allowed accuracies of 57.87%, 81.91% and 95.84%.

Similar findings were reached also for the individual analysis, with TOF overall leading the recognition performances, with RAD quite unreliable, and with the accuracies from all data conditions converging to high values when considering a simple 2-classes problem of power grasp versus precision pinch grasp. Specifically, for TOF average accuracies were 86.76:9.04%, 93.55:14.71% and 91.89:10.91% for all objects, shapes and grasps recognition, respectively. For RAD data condition, accuracies were 41.18:33.17%, 63.64:52.59% and 71.79:20.99% for objects, shapes and grasps. When it comes to the mixed-input neural networks, RADTOF accuracies were 82.81:15.80%, 45.71:47.22% and 91.43:27.68%, while RADTOFIMU accuracies were 85.48:18.01%, 57.58:34.44% and 85.37:14.74% for all objects, shapes and grasps recognition, respectively.

B. Inference analysis on new data

All models trained with the aggregated data from all 29 participants were further explored via inference testing, thus by measuring classification accuracy on unseen data not part of the original dataset. Results for all data condition, scenario and classification problem are reported in TABLE *2*. A summary of the most promising results for each scenario and classification problem is depicted in Figure 5. via confusion matrices. As expected, the bench-static scenario proved to be the easiest to solve with RADTOF being the condition which closest followed the accuracies seen during offline testing. Here, beside showing misclassifications between the wooden sphere and cylinder and between the wooden prisms, resulting accuracies were relatively high even when dealing with all objects available placed with different orientations (Video 1). Instead, user-static proved to be the most challenging scenario in which a poorly controlled alignment with the target object can considerably deteriorate recognition. Indeed, classifications were quite random when trying to recognize all objects, and misclassifications were mostly related to the small prisms and the large cuboid and sphere when trying to recognize the shapes. Nevertheless, the RAD condition allowed discrete performances when the classification problem was simplified in choosing among only two hand grasps, power or precision grasp. The inference accuracies for the grasping pick-and-lift scenario were quite unexpected. Here, results showed promising performance for the recognition of shapes and grasps even when sensors were moving towards the target, reaching classification accuracies as high as 94% with RADTOF condition on the 2-classes problem. The inclusion of inertial sensor data proved to be beneficial when matching the classification problem to the number of different reaching trajectories of each shape, with misclassifications mostly related to the small prisms.

Pick-Lift-and-Move: what happens if we include more directions of approach?

The pick-lift-and-move trials included an increasing level of complexity, namely a different and more disadvantageous starting position as well as two more directions of approach towards the target object, rightwards with the target poorly in view at start and leftwards with the target not in view at start. The same models trained with the aggregated data from all 29 participants were further inference tested on this unseen and diverse data. Results are reported in TABLE 3. Overall and as expected, accuracies dropped from the pick-and-lift inference tests, particularly for all neural networks which depended on time-of-flight sensor data. Highest accuracy was 68% and reached by RAD CNN in the 2-classes problem. Moreover, solving the recognition problem for ten different objects with two shapes and two materials seemed just unfeasible for all data conditions.

TABLE 3.
INFERENCE ACCURACIES (%) FOR PICK-LIFT-AND-MOVE

	Pick-Lift-and-Move		
	Which object?	Which shape?	Which grasp?
RAD	21.00	34.00	68.00
TOF	11.00	12.00	48.00
RADTOF	26.00	20.00	62.00
RADTOFIMU	22.00	48.00	52.00

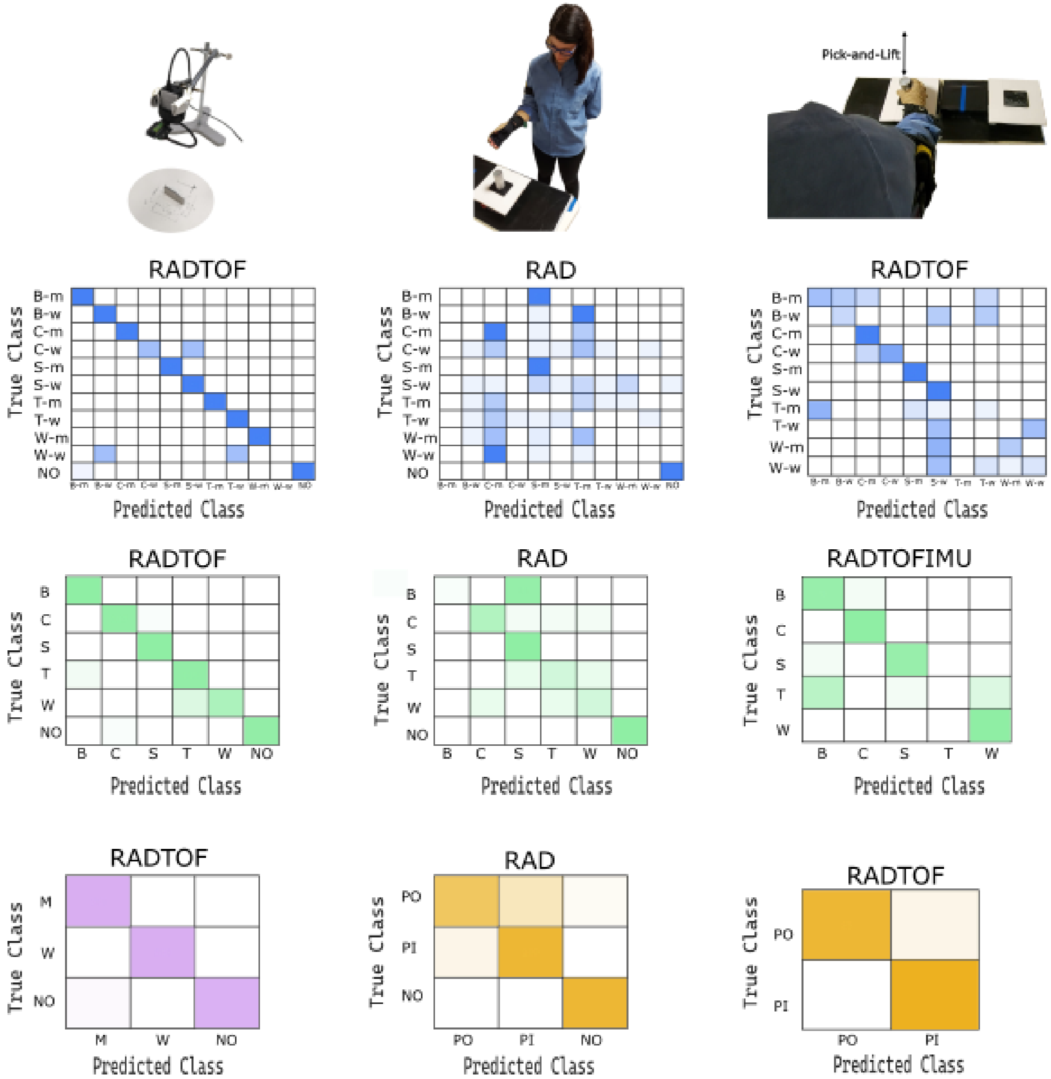

Figure 5. Confusion matrices of the most promising results from the inference test. The matrices visualize classifications for the three scenarios (columns) and for the different classification problems (rows). The different problems are indicated with different colours (legend as in Fig.4), while classes are indicated with initials. For the shapes, sphere = S, cylinder = C, triangular prism = T, cuboid with a thin rectangular prism 'handle' = B, thin rectangular prism = W, no-object = NO. For the different materials, metal = M and wood = W. Lastly, for the grasps, power = PO, precision pinch = PI.

IV. Discussion

The feasibility analyses reported in this study seem to confirm the original hypothesis of state-of-the-art proximity sensors being a possible alternative for shapes and materials recognition of objects. Indeed, data from a modern integrated radar and a time-of-flight 8x8 pixels camera proved to be sufficient for the reliable and repeatable recognition of ten different objects in an ideal static scenario (Video 1), thus with the targets steady and located within the sensors' field of view. Moreover, it was investigated how gradually increasing the complexity of this ideal scenario would change the recognition accuracy. Specifically, promising results were found also in a less ideal but still static scenario of subjects standing in front of the target object with limited control on the sensors' field of view centring and alignment in respect to the targets. Here, data from a single radar sensor was sufficient to achieve discrete performances in recognising the most indicated hand grasp to use before starting the reaching motion towards the object. Against expectations, such radar sensor (and its combination with other sensors) proved also promising results when facing dynamic scenarios, thus showing the potential to recognize the target object even during reaching motion.

Interestingly, the bench-static setup unveiled the potential of using a radar sensor for the recognition of materials. Even if assessed briefly and with two very different materials, these results are in harmony with previous literature [33], showing a great potential for radar sensor to characterize complex objects based on their permittivity. Such considerations are to be taken into account for the design of modern intelligent prosthetic hands, likely based on a mixed-sensors system where, hypothetically, materials can be recognized in a later stage of the grasping action so to allow minor grip adjustments and prevent object slippage.

As seen in the inference tests, the trained networks responded differently to unseen data in different scenarios and classification problems. Due to the limited testing is it therefore hard to make any claim on the data generalizability. While aggregating the data from multiple participants seems to be the way to go, it might be also beneficial, for an actual clinical translation point of view, to better tailor the data collection and network training to the individual characteristics of the participant of interest (i.e., the end-user), probably also including different object reaching directions. The negative correlation found between participants' height and resulting accuracies would suggest that further sanity checks on the data might be needed in order to reduce unhelpful data variability.

This study intended to compare two state-of-the-art proximity sensors, so to understand their different potentials for the given problem. Moreover, we intended also to verify the original hypothesis that a combination of these sensors can be of value for the given problem. Overall, the data from the radar sensor was better suited for the different problems in all scenarios, sometimes even outperforming the combination with inertial and TOF sensors. However, combining radar and time-of-flight data proved to be successful for an ideal bench-static setup. Oppositely, the time-of-flight sensor heavily suffered from misalignments with the target objects, in line with its reduced field-of-view compared to the radar (45x45° vs 65x53°), but its data became more and more relevant as the sensor progressed towards the target. Additionally, the radar performance might be improved with different data acquisitions (e.g., different range points and sweeps settings for the radar) and different processing (e.g., alternative to common doppler maps, advanced background removal techniques). Nevertheless, a combination of the two proximity sensors might still be the way to go, perhaps optimizing the deep learning models so to reduce the impact of a sensor in the final prediction accordingly to the scenario of interest.

While it is clear that introducing a less controlled alignment within the sensors' field of view as well as diverse data from dynamic situations (like reaching and grasping the objects from different directions) can have considerable effects on the final accuracy, the results are still remarkable considering the simplicity of these sensors compared to those normally operated in similar research, such as RGB-D cameras. Such considerations can have implications in the development of future autonomous robotic grippers, especially prosthetic hands. At first, proximity sensors can be a valid alternative to more power and computation demanding camera-based systems for the recognition of objects or their basic characteristics relevant to the grasp. Moreover, such recognition seems feasible either before and after starting the approach movement towards the target, provided that the objects are somehow within the sensors' field of view. The latter finding opens up the possibility of performing and/or updating the intended grasp on the robotic hand while the amputee user moves towards the target, a challenge and achievement that is still unprecedented in the field. Ultimately, we envision a system in which a complex target recognition can be achieved via three phases: 1) during the static phase to allow a correct preshape of the prosthesis, 2) during the reaching phase to allow corrections of the grasp or to alternatively alert the user of a potential misclassification, and 3) during the target-is-grasped phase to allow recognition of further details of the object and consequent grip adjustments. It remains of primary interest to understand how each of these phases would impact the naturalness of prosthesis use and thus the final user's acceptance.

Limitations

To the best of our knowledge, this study offers for the first time the exploration of radar and time-of-flight proximity sensors for object properties recognition and grasp facilitation in autonomous prosthetic hands. However, we report here only pure offline analysis. Even though the inference tests included here are of great help to assess the nets translational potential, such results are limited and to be intended as a feasibility check; as seen already in other technologies, the system performance can wildly change during real-time tests. For this reason, it is our intention to continue optimize the deep learning models, port them from the computer to a wearable format, and then test such approach in real-time with an actual robotic prosthetic hand.

This study analyzed a limited set of objects and materials, thus it is still unclear how results are transferrable to a realistic scenario of use with a prosthetic hand. Nevertheless, we argue that the target objects used here can be already fairly representative of certain daily-life tasks because these objects were purposely selected for the Southampton Hand Assessment Procedure, a well-known, clinically validated functional assessment for robotic hands intended for prosthetics purposes.

The ”no-object” class poses further challenges to the user-static scenario (i.e., when trying to recognize an object before actually starting the reaching motion). Here, the no-object class for user-static was simply represented by ideal data acquired by the proximity sensors while pointing at nowhere with no hit within a 1 meter range. However, such ideal data might be hard to acquire in a real implementation with the amputee hovering the prosthesis over a desk, or towards a group of objects. The risk of frustrating misclassifications is higher in a realistic scenario.

Even though shape recognition proved to be feasible while approaching the target object, no hand prosthesis currently on the market can change its fingers posture to a grasp in less than half a second. Thus, with current robotic technology amputee users would need to reach the target with slower speed than able-bodied individuals. Nevertheless, the idea and goal remain of interest because it would allow the hand prosthesis to better mimic its biological counterpart.

Architecture and hyperparameters optimizations on the deep learning models used here were limited. Preliminary explorative iterations were conducted to find a basic network architecture and a hyperparameters set that could suit all data conditions, after which the networks were not further optimized during the analyses. However, there is a large potential for improvement by performing more extensive systematic optimizations tailored to the situation of interest (i.e., which sensor, static or dynamic, which classification problem). This route will be surely further explored in the next steps of the project.

V. Conclusions

In this study we explored for the first time the use of state-of-the-art proximity sensors for object grasp facilitation in autonomous prosthetic hand. To this aim, an extensive human-object interactions dataset was analysed via deep learning models trying to classify the different target objects. Results showed a promising potential for grasp classification (i.e., selecting the adequate hand grasp to be used) before and also during the reaching-to-grasp motion. Results seem to suggest modern, low-power radar as a potential key technology for next generation intelligent and autonomous robotic hands. In particular, these results seek to contribute to the development of alternative control approaches for prosthetic hands, less dependent on the conventional but knowingly unreliable electromyographic human-machine interface. Autonomous prosthetic hands can be a game changer, envisioning a complex system that can interact in an intelligent fashion with the user and any object.

Acknowledgements

The authors would like to express their gratitude to Acconeer AB for supporting and freely sharing samples of the most recent radar A121.

Competing interests

CC is the founder of Prensilia SRL, a university spin-off that develops multi-articulated and sensorized robotic hands. All authors declare this research was conducted in the absence of any commercial or financial relationships that could constitute a potential conflict of interest.

References

[1] K. J. Zuo and J. L. Olson, “The evolution of functional hand replacement: From iron prostheses to hand transplantation,” *Can. J. Plast. Surg.*, vol. 22, no. 1, pp. 44–51, 2014.

[2] R. Merletti and P. A. Parker, *Surface Electromyography : Physiology, Engineering, and Applications*. Hoboken, NJ, USA: Wiley, 2016. doi: 10.1002/9781119082934.

[3] J. G. Webster, *Medical Instrumentation: Application and Design*.

[4] D. Graupe, A. A. Beex, W. J. Monlux, and I. Magnussen, “A multifunctional prosthesis control system based on time series identification of EMG signals using microprocessors.,” *Bull. Prosthet. Res.*, vol. 10, no. 27, pp. 4–16, 1977, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/603818

[5] B. Hudgins, P. Parker, and R. N. Scott, “A new strategy for multifunction myoelectric control,” *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, 1993, doi: 10.1109/10.204774.

[6] M. Ortiz-Catalan, B. Hakansson, and R. Branemark, “Real-Time and Simultaneous Control of Artificial Limbs Based on Pattern Recognition Algorithms,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 4, pp. 756–764, Jul. 2014, doi: 10.1109/TNSRE.2014.2305097.

[7] L. J. Hargrove, L. A. Miller, K. Turner, and T. A. Kuiken, “Myoelectric Pattern Recognition Outperforms Direct Control for Transhumeral Amputees with Targeted Muscle Reinnervation: A Randomized Clinical Trial,” *Sci. Rep.*, vol. 7, no. 1, 2017, doi: 10.1038/s41598-017-14386-w.

[8] J. M. Hahne, M. A. Schweisfurth, M. Koppe, and D. Farina, “Simultaneous control of multiple functions of bionic hand prostheses: Performance and robustness in end users,” *Sci. Robot.*, vol. 3, no. 19, p. eaat3630, Jun. 2018, doi: 10.1126/scirobotics.aat3630.

[9] E. Mastinu, J. Ahlberg, E. Lendaro, L. Hermansson, B. Hakansson, and M. Ortiz-Catalan, “An Alternative Myoelectric Pattern Recognition Approach for the Control of Hand Prostheses: A Case Study of Use in Daily Life by a Dysmelia Subject,” *IEEE J. Transl. Eng. Heal. Med.*, vol. 6, no. June 2017, pp. 1–12, 2018, doi: 10.1109/JTEHM.2018.2811458.

[10] A. M. Simon *et al.*, “User Performance With a Transradial Multi-Articulating Hand Prosthesis During Pattern Recognition and Direct Control Home Use,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 271–281, 2023, doi: 10.1109/TNSRE.2022.3221558.

[11] D. D’Accolti, F. Clemente, A. Mannini, E. Mastinu, M. Ortiz-Catalan, and C. Cipriani, “Online Classification of Transient EMG Patterns for the Control of the Wrist and Hand in a Transradial Prosthesis,” *IEEE Robot. Autom. Lett.*, vol. 8, no. 2, pp. 1045–1052, Feb. 2023, doi: 10.1109/LRA.2023.3235680.

[12] A. D. Roche, H. Rehbaum, D. Farina, and O. C. Aszmann, “Prosthetic Myoelectric Control Strategies: A Clinical Perspective,” *Curr. Surg. Reports*, vol. 2, no. 3, p. 44, Mar. 2014, doi: 10.1007/s40137-013-0044-8.

[13] S. Salminger *et al.*, “Current rates of prosthetic usage in upper-limb amputees – have innovations had an impact on device acceptance?,” *Disabil. Rehabil.*, vol. 44, no. 14, pp. 3708–3713, Jul. 2022, doi: 10.1080/09638288.2020.1866684.

[14] M. Ortiz-Catalan, E. Mastinu, P. Sassu, O. Aszmann, and R. Brånemark, “Self-Contained Neuromusculoskeletal Arm Prostheses,” *N. Engl. J. Med.*, vol. 382, no. 18, pp. 1732–1738, Apr. 2020, doi: 10.1056/NEJMoa1917537.

[15] J. Zbinden *et al.*, “Improved control of a prosthetic limb by surgically creating electro-neuromuscular constructs with implanted electrodes,” *Sci. Transl. Med.*, vol. 15, no. 704, Jul. 2023, doi: 10.1126/scitranslmed.abq3665.

[16] P. F. Pasquina *et al.*, “First-in-man demonstration of a fully implanted myoelectric sensors system to control an advanced electromechanical prosthetic hand,” *J. Neurosci. Methods*, vol. 244, pp. 85–93, Apr. 2015, doi: 10.1016/j.jneumeth.2014.07.016.

[17] M. Rakić, “An automatic hand prosthesis,” *Med. Electron. Biol. Eng.*, 1964, doi: 10.1007/BF02474360.

[18] Peter J. Kyberd, “The Southampton Hand: An

intelligent myoelectric prosthesis," *J. Rehabil. Res. Dev.*, no. November, pp. 326–334, 1994.

[19] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," 2008. doi: 10.1177/0278364907087172.

[20] S. Došen and D. B. Popović, "Transradial Prosthesis: Artificial Vision for Control of Prehension," *Artif. Organs*, vol. 35, no. 1, pp. 37–48, Jan. 2011, doi: 10.1111/j.1525-1594.2010.01040.x.

[21] J. Degol, A. Akhtar, B. Manja, and T. Bretl, "Automatic grasp selection using a camera in a hand prosthesis," 2016. doi: 10.1109/EMBC.2016.7590732.

[22] P. Weiner, J. Starke, F. Hundhausen, J. Beil, and T. Asfour, "The KIT Prosthetic Hand: Design and Control," 2018. doi: 10.1109/IROS.2018.8593851.

[23] S. Došen, C. Cipriani, M. Kostić, M. Controzzi, M. C. Carrozza, and D. B. Popovič, "Cognitive vision system for control of dexterous prosthetic hands: Experimental evaluation," *J. Neuroeng. Rehabil.*, 2010, doi: 10.1186/1743-0003-7-42.

[24] M. Markovic, S. Dosen, C. Cipriani, D. Popovic, and D. Farina, "Stereovision and augmented reality for closed-loop control of grasping in hand prostheses," *J. Neural Eng.*, vol. 11, no. 4, p. 046001, Aug. 2014, doi: 10.1088/1741-2560/11/4/046001.

[25] M. Markovic, S. Dosen, D. Popovic, B. Graimann, and D. Farina, "Sensor fusion and computer vision for context-aware control of a multi degree-of-freedom prosthesis," *J. Neural Eng.*, vol. 12, no. 6, p. 066022, Dec. 2015, doi: 10.1088/1741-2560/12/6/066022.

[26] J. Mouchoux, S. Carisi, S. Dosen, D. Farina, A. F. Schilling, and M. Markovic, "Artificial Perception and Semiautonomous Control in Myoelectric Hand Prostheses Increases Performance and Decreases Effort," *IEEE Trans. Robot.*, vol. 37, no. 4, pp. 1298–1312, Aug. 2021, doi: 10.1109/TRO.2020.3047013.

[27] M. N. Castro and S. Dosen, "Continuous Semi-autonomous Prosthesis Control Using a Depth Sensor on the Hand," *Front. Neurorobot.*, vol. 16, no. March, pp. 1–17, Mar. 2022, doi: 10.3389/fnbot.2022.814973.

[28] J. Starke, P. Weiner, M. Crell, and T. Asfour, "Semi-autonomous control of prosthetic hands based on multimodal sensing, human grasp demonstration and user intention," *Rob. Auton. Syst.*, vol. 154, p. 104123, Aug. 2022, doi: 10.1016/j.robot.2022.104123.

[29] P. Weiner, J. Starke, S. Rader, F. Hundhausen, and T. Asfour, "Designing Prosthetic Hands With Embodied Intelligence: The KIT Prosthetic Hands," *Front. Neurorobot.*, vol. 16, no. March, 2022, doi: 10.3389/fnbot.2022.815716.

[30] A. Saudabayev and H. A. Varol, "Sensors for Robotic Hands: A Survey of State of the Art," *IEEE Access*, vol. 3, pp. 1765–1782, 2015, doi: 10.1109/ACCESS.2015.2482543.

[31] N. E. Krausz and L. J. Hargrove, "A Survey of Teleceptive Sensing for Wearable Assistive Robotic Devices," *Sensors*, vol. 19, no. 23, p. 5238, Nov. 2019, doi: 10.3390/s19235238.

[32] W. Guo, W. Xu, Y. Zhao, X. Shi, X. Sheng, and X. Zhu, "Toward Human-in-the-Loop Shared Control for Upper-Limb Prostheses: A Systematic Analysis of State-of-the-Art Technologies," *IEEE Trans. Med. Robot. Bionics*, vol. 5, no. 3, pp. 563–579, Aug. 2023, doi: 10.1109/TMRB.2023.3292419.

[33] H.-S. Yeo, G. Flamich, P. Schrempf, D. Harris-Birtill, and A. Quigley, "RadarCat," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, Oct. 2016, pp. 833–841. doi: 10.1145/2984511.2984515.

[34] C. Waldschmidt, J. Hasch, and W. Menzel, "Automotive Radar — From First Efforts to Future Systems," *IEEE J. Microwaves*, vol. 1, no. 1, pp. 135–148, Jan. 2021, doi: 10.1109/JMW.2020.3033616.

[35] F. Tosti and C. Ferrante, "Using Ground Penetrating Radar Methods to Investigate Reinforced Concrete Structures," *Surv. Geophys.*, vol. 41, no. 3, pp. 485–530, May 2020, doi: 10.1007/s10712-019-09565-5.

[36] M. Mercuri, I. R. Lorato, Y.-H. Liu, F. Wieringa, C. Van Hoof, and T. Torfs, "Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor," *Nat. Electron.*, vol. 2, no. 6, pp. 252–262, Jun. 2019, doi: 10.1038/s41928-019-0258-6.

[37] M. Giordano, G. Islamoglu, V. Potocnik, C. Vogt, and M. Magno, "Survey, Analysis and Comparison of Radar Technologies for Embedded Vital Sign Monitoring," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Jul. 2022, pp. 854–860. doi: 10.1109/EMBC48229.2022.9871847.

[38] J. Lien *et al.*, "Soli," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, Jul. 2016, doi: 10.1145/2897824.2925953.

[39] M. Scherer, M. Magno, J. Erb, P. Mayer, M. Eggimann, and L. Benini, "TinyRadarNN: Combining Spatial and Temporal Convolutional Neural Networks for Embedded Gesture Recognition With Short Range Radars," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10336–10346, Jul. 2021, doi: 10.1109/JIOT.2021.3067382.

[40] E. Mastinu, A. Coletti, S. H. A. Mohammad, J. van den Berg, and C. Cipriani, "HANDdata – first-person dataset including proximity and kinematics measurements from reach-to-grasp actions," *Sci. Data*, vol. 10, no. 1, p. 405, Jun. 2023, doi: 10.1038/s41597-023-02313-w.

[41] C. M. Light, P. H. Chappell, and P. J. Kyberd, "Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: Normative data, reliability, and validity," *Arch. Phys. Med. Rehabil.*, vol. 83, no. 6, pp. 776–783, 2002, doi: 10.1053/apmr.2002.32737.