Smart Ankleband for Plug-and-Play Hand-Prosthetic Control

Dean Zadok[1], Oren Salzman[1], Alon Wolf[2], Alex M. Bronstein[1]
[1]*Dept. of Computer Science,* [2]*Dept. of Mechanical Engineering*
Technion - IIT,
Haifa, Israel
{deanzadok,osalzman,bron}@cs.technion.ac.il, alonw@me.technion.ac.il

***Abstract*—Building robotic prostheses requires a sensor-based interface designed to provide the robotic hand with the control required to perform hand gestures. Traditional Electromyography (EMG) based prosthetics and emerging alternatives often face limitations such as muscle-activation limitations, high cost, and complex calibrations. In this paper, we present a low-cost robotic system composed of a smart ankleband for intuitive, calibration-free control of a robotic hand, and a robotic prosthetic hand that executes actions corresponding to leg gestures. The ankleband integrates an Inertial Measurement Unit (IMU) sensor with a lightweight neural network to infer user-intended leg gestures from motion data. Our system represents a significant step towards higher adoption rates of robotic prostheses among arm amputees, as it enables one to operate a prosthetic hand using a low-cost, low-power, and calibration-free solution. To evaluate our work, we collected data from 10 subjects and tested our prototype ankleband with a robotic hand on an individual with an upper-limb amputation. Our results demonstrate that this system empowers users to perform daily tasks more efficiently, requiring few compensatory movements.**

***Index Terms*—Prosthetic Arms, Wearable Robotics, Deep Learning Methods.**

I. Introduction

People with upper-limb differences, especially amputees, struggle with Activities of Daily Living (ADLs) [1] due to the limitations of current prosthetics, which often lack functionality, intuitive control, and can cause discomfort or pain. Despite the dominance of muscle-based control technologies like electromyography (EMG) [2], [3] and ultrasound (US) [4], [5], their limitations, including muscle disuse, pain, and the need for complex training, result in low adoption rates of current prosthetics, decreased quality of life, and reluctance to experiment with newer developments.

In recent years, researchers have begun to explore alternative control methods for prosthetic arms that do not involve muscle activation. These include autonomous behaviors such as predicting the action type using motion [6], [7], detecting objects through vision sensing [8], [9], providing instructions using voice commands [10], and even control robotic devices through the interpretation of brain activity via electroencephalography (EEG) [11]. However, such methods face various operational limitations. For example, environmental factors such as lighting and motion may degrade the precision of vision-based methods, and to process high-resolution data, one will need expensive wearable hardware to allow precise prosthetic control. Addressing these limitations, including the discomfort associated with muscle-based control, is crucial for increasing the adoption of robotic prosthetic arms.

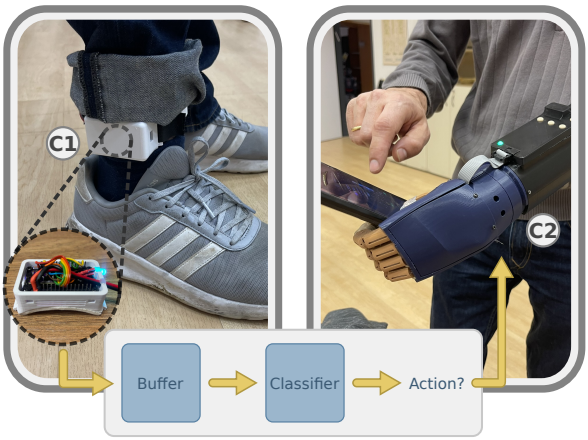


Fig. 1: Our robotic system is comprised of two components: **C1** smart ankleband (left) and **C2** robotic hand (right). An IMU placed in the smart ankleband collects and stores signals in a buffer which serves as an input to a machine-learning model able to classify user leg gestures. The corresponding action is transmitted to the robotic hand.

In response to these challenges, this paper presents an end-to-end low budget, low power, wearable system that allow upper-extremity amputees to seamlessly control a robotic hand without calibration.[1] Our system (Fig. 1) is based on a smart ankleband, composed of a lightweight microcontroller, a battery, and an IMU. The ankleband uses a pre-trained machine-learning model to infer leg gestures that the user performs, which correspond to predefined actions, given the motion data generated from the IMU. These are then transmitted to the robotic hand using a wireless interface which, in turn, executes the relevant action. In our empirical evaluation, we demonstrate the ability of our system to accurately infer gestures for unseen individuals, by collecting data from 10 subjects. Finally, we demonstrate via a user study in the lab how the system can be used by a user with an upper-limb amputation and enhance the performance of ADL.

[1]Our code and dataset are at https://github.com/deanzadok/ankleband.

II. SYSTEM DESIGN

A. Overview

Our design choices aim to increase the adoption of assistive devices for amputees are guided by three primary objectives:

- **O1** Affordability—the system cost should be minimized as this is a significant barrier to adoption.
- **O2** Reliability—the system should exhibit high performance even under low-energy and cost constraints.
- **O3** User independent—the system should be "plug and play", eliminating the need for calibration for new users.

To this end, we propose a robotic system (Fig. 1) consisting of component **C1**—a wearable device placed on the ankle called the *smart ankleband* and component **C2**—a robotic hand. The smart ankleband is designed to detect intended leg gestures performed by the user and transmit them to the robotic hand using a wireless interface which, in turn, executes the corresponding action. The smart ankleband (Fig. 1, left) includes an IMU motion sensor for motion detection, a microcontroller for leg-gesture classification, Bluetooth Low-Energy (BLE) command transmission, and a power source. The robotic hand (Fig. 1, right) is designed to be generic and can support a wide range of platforms. We based our choice on an open-source project designed to make robotic hands accessible and programmable[2].

Based on user feedback, we define four robotic hand actions: grasp, point, rotate wrist left, and rotate wrist right (Fig. 2). Accordingly, we define four different leg gestures corresponding to each action. To return to an open-hand state, the user should perform the last-performed gesture. Importantly, the specific leg gestures were selected based on the spatial displacement of the IMU sensor, with the four distinct gestures designed to simplify the classification process (Fig. 2). While the gestures are similar to those used in previous papers adopting leg-gesture classification [12], we require each gesture to be performed twice to minimize unintended activations.

B. Smart-Ankleband Physical Design

The smart ankleband integrates a microcontroller, IMU, and power source into a comfortable and compact wearable design (Fig. 1, left) that requires no additional clothing or user-specific adaptations. To optimize comfort and weight distribution, the microcontroller and IMU are housed in one component, with the power source in a separate one.

To achieve low cost (**O1**) while ensuring reliability (**O2**), we selected the ESP32 microcontroller for its balance of processing power, varied interfaces, and energy efficiency, and the Adafruit BNO08X IMU for its cost-effectiveness and ability to generate high-frequency 3D acceleration and angular velocity data (up to 1 kHz). Finally, the power source is a 3.7V battery with 950 mAh. The overall cost of the smart ankleband is roughly 30 USD, requiring only standard 3D printing and basic soldering.

[2]Full details on how to rebuild and program the hand are in https://github.com/Haifa3D/hand-mechanical-design.

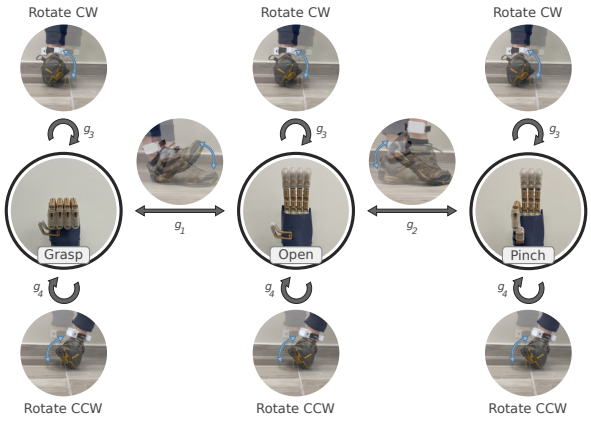


Fig. 2: State machine depicting leg gestures and corresponding hand actions. We start with an open hand, and the user can either grasp (leg gesture g_1), or pinch (leg gesture g_2). To rotate the hand clockwise (CW) or counterclockwise (CCW), the user can use leg gestures g_3 or g_4, respectively. Rotation is available regardless of the state.

C. Leg-Gesture Classification

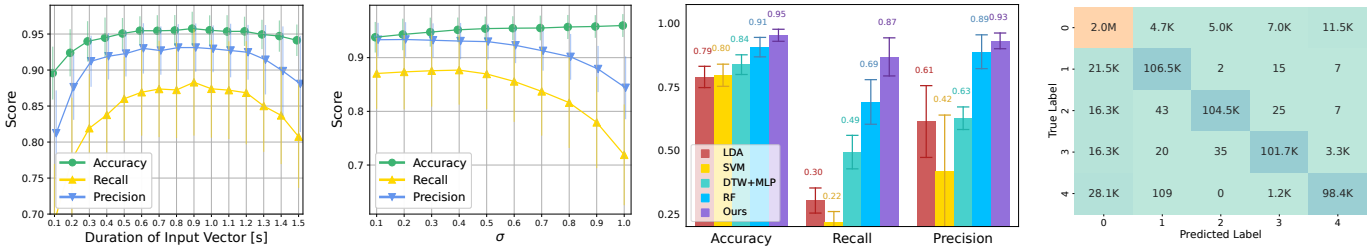
Recall that we predefine a set of leg gestures to control the robotic hand. Specifically, at time t the IMU outputs a vector $x_t = [\ddot{x}_t, \ddot{y}_t, \ddot{z}_t, \dot{\theta}^r_t, \dot{\theta}^p_t, \dot{\theta}^y_t]$. Here, $\ddot{x}_t, \ddot{y}_t, \ddot{z}_t$ are the accelerations in the x, y, and z axes, and $\dot{\theta}^r_t, \dot{\theta}^p_t, \dot{\theta}^y_t$ are the roll, pitch, and yaw angular velocities. The input $\bar{x}_t$ is a sequence of k consecutive IMU signals recorded between times t and $t+k-1$.[3] Namely $\bar{x}_t := \langle x_t, \ldots, x_{t+k-1} \rangle$. The output p_t is a vector of gesture probabilities at time t. Specifically, $p_t := \langle p^0_t, \ldots, p^4_t \rangle$, where p^0_t is the probability for no gesture and p^i_t for $i \in \{1, \ldots 4\}$ represents the probability of gesture g_i being performed at time t.

As the user can perform a leg gesture at any given time and its length varies, we need to carefully associate the input $\bar{x}_t$ to our classifier and the continuous time interval $[t^g_s, t^g_e]$ duration during which leg gesture g was performed. With a slight abuse of notation, given the time interval $[t^g_s, t^g_e]$ for leg gesture g and the input sequence $\bar{x}_t$, we define $g(\bar{x}_t, t^g_s, t^g_e)$ to be the proportion of time where gesture g lies in the interval. Namely,

$$g(\bar{x}_t, t^g_s, t^g_e) := \frac{[t^g_s, t^g_e] \cap [t, t+k-1]}{[t^g_s, t^g_e] \cup [t, t+k-1]}. \quad (1)$$

Now, given a threshold σ (chosen empirically as we detail in Sec. III-A), we label gesture g in interval $[t^g_s, t^g_e]$ if $g(\bar{x}_t, t^g_s, t^g_e) \geq \sigma$. Notice that $g(\bar{x}_t, t^g_s, t^g_e) = 1$ if g is fully present within $\bar{x}_t$ (i.e., $t \leq t^g_s$ and $t^g_e \leq t+k-1$), and $g(\bar{x}_t, t^g_s, t^g_e) = 0$ if g is not present at all (i.e., $t^g_s < t$ or $t+k-1 < t^g_e$). Our compact neural network includes a 1D convolutional layer for temporal feature extraction, followed by batch normalization to enhance performance for unseen users, and a multi-layer perceptron to generate the gesture probability vector p_t. Additional technical details can be found in the extended version of this paper [13].

[3]The value k is a hyper parameter discussed in Sec. III. It should be large enough to capture leg gestures which typically take between 0.5 to 1 seconds but small enough to be processed in real time and include at most one gesture.



(a) (b) (c) (d)

Fig. 3: (a) Performance as a function of the duration of the input vector in seconds. (b) Performance as a function of σ (the minimum percentage of gesture overlap with the input window, Sec. II-C). (c) Evaluation of popular classification methods in comparison to the chosen method. (d) Confusion matrix for the five classes: a no-gesture class and the four gestures. Rows and columns represent true labels and predictions, respectively. All experiments are averaged over 10 folds, with one subject left out for test set in each fold. Vertical segments denote one standard deviation.

III. EVALUATION

We collected data from 10 subjects (five men, five women, average age 27) free of neurological leg disorders. All subjects wore the smart ankleband on their right leg, while a Vicon motion-capture system monitored foot motion to compute label intervals after recordings. Subjects repeatedly performed each gesture for 1-2 minutes, both while sitting and standing, with each session repeated twice. In addition, we collected two minutes of "noise" data (no specific leg gesture) to improve model robustness. This process resulted in 8-16 labeled minutes of leg gestures and four minutes of regular activity per subject. The process was approved by the institution's Ethics Committee. The overall dataset includes a total of 2.5 million x_t samples. For each sample, we create the input vector $\bar{x}_t$ and label it to receive the tuple $\langle \bar{x}_t, p_t \rangle$ defined in Sec. II-C. On feedforward, our model receives a sequence of $k = 60$ IMU samples $\bar{x}_t$, and output the gesture probabilities p_t. For technical details regarding model training and architecture, please refer to the extended version of this paper [13].

A. Model Study

For all experiments, we evaluated our models using 10-fold cross-validation to simulate testing on a new, unseen user. In this setup, data from a single subject is reserved for testing, while the remaining data from the nine subjects is used for training. This process is repeated for each subject, and the results are the average of the 10 training sessions.

What is the duration for the input window? Our classifier takes a vector $\bar{x}$ of k IMU signals as input, designed to capture leg gestures lasting 0.5 to 1 second. Thus, we evaluated model performance for different durations of $\bar{x}$ ranging from 0.1 to 1.5 seconds (Fig. 3a), and found the optimal results to be between 0.6 and 1.0 seconds. Considering memory constraints, which limit the size of the tensors during inference, we set the sensor frequency to 100 Hz and chose a 0.6-second duration, resulting in $k = 60$ IMU samples.

How should we determine the existence of labels in input windows? We optimized the threshold σ, determining whether a leg gesture that partially exists in the input window should be labeled as present or not (Eq. (1)). Here, we evaluated σ values between 0.1 and 1.0 (Fig. 3b), and found that higher values significantly decreased performance. However, to ensure that each input window contains only one prominent gesture, we require $g(\bar{x}_t, t_s^g, t_e^g) \geq 0.5$. Thus, we set $\sigma = 0.5$.

How do we choose the best model? The modular design of our system allows for classifier substitution. Our model is constrained by a dynamic memory limit of 90 kB due to our limited hardware. Therefore, larger deep-learning (DL) models adopted in previous studies were not feasible [6], [7]. We evaluated classical data-driven approaches such as Linear Discriminant Analysis (LDA) [14], Support Vector Machine (SVM) [12], [15], and Random Forest (RF) [16] (Fig. 3c). Additionally, we experimented with Dynamic Time Warping (DTW) [15], [17] for feature extraction, followed by a classifier to perform gesture recognition. Consequently, our proposed DL model outperformed the other methods, despite its limited size. A possible explanation is that, unlike high-end sensors, our IMU sensor produces noisy data, requiring denoising capabilities which DL methods excel at [18], [19]. We also explored advanced optimization techniques like domain adaptation [20] and contrastive learning [21] to improve generalization to unseen subjects. However, supervised learning proved most effective in our setting.

We generated the confusion matrix of the entire dataset to see how the model classifies each gesture compared to the ground truth (Fig. 3d). While gesture g_4 presented the most challenge, the model effectively predicted all four gestures. Crucially, the model avoided misclassifying gestures as other gestures, which ensures safety for tasks requiring reliable recognition. For a comprehensive evaluation of our model, please refer to the extended version of this paper [13].

B. Robotic System Study

To evaluate our system in ADL, we conducted an in-lab user study with a participant who had left transradial amputation 12 years ago and had experience with prosthetics. The experiment consisted of two tests. First, the participant performed each leg gesture ten times to familiarize himself with the smart ankleband, without using the robotic hand. Second, the participant completed the "Activities Measure for Upper-limb Amputees" (AM-ULA) test [22], which consists

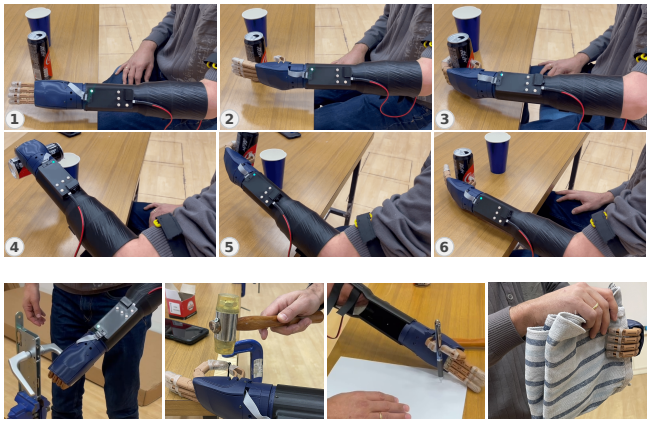


Fig. 4: Examples of tasks completed as part of the AM-ULA test. (Top) The user reaches the soda can (1) and grasps it using leg gesture g_1 (2 and 3). Then, the user lifts and rotates CW using leg gesture g_3 to pour the beverage (4). Finally, the user rotates CCW using g_4 (5) and opens the hand using g_1 to leave the soda can on the table (6). (Bottom, from left to right) pushing a door knob, hammering a nail, writing on a blank paper, and folding a towel.

of 18 listed activities requiring the operation of a prosthetic hand (Fig. 4). This test evaluates not only the functionality of our ankleband but also the user's proficiency and interaction with the robotic system.

During the first part, the participant completed 9 out of 10 repetitions for the first gesture and 10 out 10 for the remaining three gestures, yielding 97.5% accuracy, despite wearing the ankle band for the first time. It showcases the intuitive design and minimal training required to use our ankleband. For the AM-ULA test, we activated the robotic hand and ensured that the smart ankleband was connected. Then, we instructed the participant to perform the 18 tasks, encouraging them to use the robotic hand. The test revealed that the participant mostly used grasp (34 intentions) and pinch (32 intentions) actions, with only three instances of wrist rotation. The user reported the ankleband to be intuitive and responsive, and that the robotic hand did not perform only a few gestures due to communication malfunctions. Notably, throughout the two-hour experiment, our ankleband operated reliably and misclassified native movements as gestures only four times, showing promising performance for ADL.

IV. DISCUSSION & FUTURE WORK

In this paper, we introduced a "plug-and-play" robotic system that uses a smart ankleband to detect intended ankle gestures, enabling seamless control of a robotic prosthetic arm. Our results demonstrate that this system achieves subject generalization with above 95% accuracy despite hardware constraints, eliminating the need for prolonged personalization.

Future directions for this research include: (i) A large-scale study to explore alternative input gestures and application interfaces by leveraging the modularity of our system. (ii) Developing a smart ankleband for proportional control of digital or robotic devices, especially for users unable to perform continuous hand motions. (iii) Enabling typing motions through our smart ankleband. We aim to provide users with disabilities the full capability to operate digital devices.

REFERENCES

[1] P. F. Edemekong, D. Bomgaars, S. Sukumaran, and S. B. Levy, "Activities of daily living," 2019.

[2] W. Li, P. Shi, and H. Yu, "Gesture recognition using surface electromyography and deep learning for prostheses hand: state-of-the-art, challenges, and future," *Frontiers in neuroscience*, vol. 15, p. 621885, 2021.

[3] R. C. Sîmpetru, M. März, and A. D. Vecchio, "Proportional and simultaneous real-time control of the full human hand from high-density electromyography," *IEEE Trans. Neural Syst. Rehabilitation Eng.*, 2023.

[4] J. McIntosh, A. Marzo, M. Fraser, and C. Phillips, "Echoflex: Hand gesture recognition using ultrasound imaging," in *CHI*, 2017, pp. 1923–1934.

[5] D. Zadok, O. Salzman, A. Wolf, and A. M. Bronstein, "Towards predicting fine finger motions from ultrasound images via kinematic representation," in *ICRA*, 2023, pp. 12 645–12 651.

[6] C. Uyanik, S. F. Hussaini, E. Erdemir, E. Kaplanoglu, and A. Sekmen, "A deep learning approach for final grasping state determination from motion trajectory of a prosthetic hand," *Procedia Comput. Sci.*, vol. 158, pp. 19–26, 2019.

[7] J. Cui, Z. Li, H. Du, B. Yan, and P. Lu, "Recognition of upper limb action intention based on IMU," *Sensors*, vol. 22, no. 5, p. 1954, 2022.

[8] M. N. Castro and S. Dosen, "Continuous semi-autonomous prosthesis control using a depth sensor on the hand," *Front. neurorobot.*, vol. 16, p. 814973, 2022.

[9] M. A. B. Sarker, J. P. S. Sola, A. Jones, E. Laing, E. Sola-Thomas, and M. H. Imtiaz, "Vision controlled sensorized prosthetic hand," *arXiv*, vol. abs/2407.12807, 2024.

[10] M. H. Asyali, M. Yilmaz, M. Tokmakci, K. Sedef, B. H. Aksebzeci, and R. Mittal, "Design and implementation of a voice-controlled prosthetic hand," *Turk. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 1, pp. 33–46, 2011.

[11] P. C. Maibam, D. Pei, P. Olikkal, R. K. Vinjamuri, and N. M. Kakoty, "Enhancing prosthetic hand control: A synergistic multi-channel electroencephalogram," *Wearable Technologies*, vol. 5, p. e18, 2024.

[12] C. Zhu, L. Luo, J. Mai, and Q. Wang, "Recognizing continuous multiple degrees of freedom foot movements with inertial sensors," *IEEE Trans. Neural Syst. Rehabilitation Eng.*, vol. 30, pp. 431–440, 2022.

[13] D. Zadok, O. Salzman, A. Wolf, and A. M. Bronstein, "Smart ankleband for plug-and-play hand-prosthetic control," *arXiv*, vol. abs/2503.17846, 2025.

[14] Y. Wang and H. Ma, "Real-time continuous gesture recognition with wireless wearable IMU sensors," in *Healthcom*, 2018, pp. 1–6.

[15] G. V. Tchane Djogdom, M. J.-D. Otis, and R. Meziane, "Dynamic time warping–based feature selection method for foot gesture cobot operation mode selection," *Int. J. Adv. Manuf. Technol.*, vol. 126, no. 9, pp. 4521–4541, 2023.

[16] A. Nadaf and S. Pardeshi, "Classifying sign language gestures using decision trees: A comparison of semg and imu sensor data," in *INCET*, 2023, pp. 1–8.

[17] B. K. Iwana and S. Uchida, "Time series classification using local distance-based features in multi-modal fusion networks," *Pattern Recognit.*, vol. 97, p. 107024, 2020.

[18] M. Brossard, S. Bonnabel, and A. Barrau, "Denoising IMU gyroscopes with deep learning for open-loop attitude estimation," *IEEE Robotics Autom. Lett.*, vol. 5, no. 3, pp. 4796–4803, 2020.

[19] M. Elad, B. Kawar, and G. Vaksman, "Image denoising: The deep learning revolution and beyond - A survey paper," *SIAM J. Imaging Sci.*, vol. 16, no. 3, pp. 1594–1654, 2023.

[20] F. M. Carlucci, L. Porzi, B. Caputo, E. Ricci, and S. R. Bulò, "Autodial: Automatic domain alignment layers," in *IEEE ICCV*, 2017, pp. 5077–5085.

[21] S. Abbaspourazad, O. Elachqar, A. C. Miller, S. Emrani, U. Nallasamy, and I. Shapiro, "Large-scale training of foundation models for wearable biosignals," in *ICLR*, 2024.

[22] L. Resnik, L. Adams, M. Borgia, J. Delikat, R. Disla, C. Ebner, and L. S. Walters, "Development and evaluation of the activities measure for upper limb amputees," *Arch. Phys. Med. Rehabil.*, vol. 94, no. 3, pp. 488–494, 2013.