
Supplementary information

**Task-agnostic exoskeleton control via
biological joint moment estimation**

In the format provided by the
authors and unedited

Overview

This document contains supplementary methods, analysis, and discussion relevant to this study. Specifically, it includes 1) a discussion on the trade-offs of hip, knee, and ankle exoskeleton assistance, 2) an analysis of estimator accuracy during eight entirely novel tasks, 3) an analysis of user lower-limb joint kinematics during the experiments that quantified user biological joint work, 4) a description of the biomechanical modeling methods used in this study, 5) a description of the baseline method used to benchmark our joint moment estimator, 6) a description of the statistical methods used in this study, and 7) a discussion on how our joint moment estimator compared to previous offline work.

Why the Hip and Knee and not the Ankle?

During most modes of ambulation, the hip and ankle joints are the dominant power providers¹⁻³, and recent research in human-in-the-loop optimization has demonstrated that assisting the hip and ankle yields greater metabolic benefits to the user than assisting the knee⁴. Thus, most exoskeleton studies have focused on assisting the hip and ankle joints during ambulation⁵. However, the knee joint plays a critical role in many non-cyclic activities, such as squatting, jumping, and lunging¹. In fact, from our recently published dataset of human lower-limb biomechanics during cyclic and non-cyclic tasks, we found that the positive work demanded from each lower-limb joint averaged across non-cyclic activities was larger than that of any joint during cyclic tasks (i.e., during ambulation)¹. As such, assisting the knee joint is a viable approach for general-purpose human augmentation, potentially augmenting users through demanding, non-cyclic tasks while mitigating the effects of distal-borne mass during walking compared to ankle-based devices⁶.

Novel Task Analysis

As described in the Methods, 19 of the 28 total tasks were used to train the final model tested in this study. This left 9 tasks out of the training set that were then tested online in our validation experiment. The hold-out tasks were lunge, stair ascent, declined walking, start and stop, medicine ball toss, step over, squat, curb, and step up. Thus, these tasks were new to our controller. To further test the generalizability of our controller, we designed eight new tasks (Extended Data Table 2 and Supplementary Video 5) that were completely novel (i.e., never tested during previous experiments). These tasks were designed to test the absolute limits of the deep neural network controller and attempt to find edge cases where performance might deteriorate. Three participants completed these eight tasks while wearing the exoskeleton running our task-agnostic controller. The accuracy of our estimator was then compared to the ground-truth joint moments (Supplementary Fig. 1).

In general, the estimator performed well for the 9 left out tasks as well as the majority of the completely novel tasks (Supplementary Fig. 1a & b). However, three of the extreme activities showed decreased estimator performance (Supplementary Fig. 1c & d). For the layup task, R^2 remained within the distribution of the training tasks, but RMSE was much higher than the other tasks. This indicates that the shape of the estimated moment curve approximated the true joint moments; however, the magnitude was not scaled correctly (as demonstrated in Supplementary Fig. 1e). For burpees and mountain climbers, both R^2 and RMSE were worse than other tasks.

To explore the uniqueness of these novel tasks relative to the distribution of the training set, we compared the ground-truth joint moments across different tasks using two metrics. 1) We computed the average peak joint moment for each task for both the hip and knee. We then averaged the hip and the knee to give a magnitude score for each task. 2) We also broke each task into its constituent subtasks (e.g., walking has several speeds, each of which is a subtask) and created a reference table of kinetic profiles based on all of the subtasks within the training data. We then used R^2 to compare the joint moment profiles from each of the subtasks (both those inside the training set and those outside) to each of the profiles in the training data for both the hip and knee. The maximum R^2 value for each subtask (best match between shape in the training set and each task shape) was chosen (this returned 1.0 for every task in the training set since it had a perfect match whereas those outside the training set returned a score based on how similar they were to their best matching subtask in the training set). These values were then averaged across subtasks within each task group and across hip and knee to give a similarity score for each task.

These two metrics, a joint moment magnitude score and a kinetic similarity score, were used to compare all of the task groups (Supplementary Fig. 1f). These results show that the areas where the estimator performance decreased were on tasks at the extreme boundaries of the tasks that were tested. For tasks where the joint moment magnitude was excessively higher than any training data, the model had difficulty correctly scaling the estimated torque even though the shape was accurate (as shown by a high R^2 but also high RMSE). For tasks where the kinetic shape was substantially different from those tasks in the training data (possibly due to the fact that both of these tasks involve using the hands for bodyweight support), the model had difficulty accurately estimating moments. However, even in these cases, the moment estimator remained stable and defaulted to low torque estimates that did not impede the user's movements (Supplementary Fig. 1e). This effect may have also been compounded by the insole sensing breakdown when most of the force is transmitted through the ends of the toes, which is the case for both burpees and mountain climbers. These results demonstrate the ability of our approach to handle completely novel tasks, which is vital to exoskeleton performance in the real world. Even when confronted with tasks at the extreme ranges of human movement, the controller remained stable without hampering human movement.

Exoskeleton Effects on User Kinematics

To understand the potential kinematic impact of wearing the exoskeleton, we compared the joint angles at the hip, knee, and ankle under three conditions (Exo On, No Exo, and Zero Torque) during the six joint work tasks (Supplementary Fig. 2). As explained in Extended Data Table 3, environmental factors, such as squat depth and timing were controlled across conditions, allowing a fair comparison between the kinematics. These results demonstrate that there is very little difference between the kinematics when wearing the exoskeleton (powered on or off) and not wearing the exoskeleton. Even though we controlled each task with a metronome, some tasks (like squatting, lunges, and step ups) have greater variability in the exact timing of the task within each individual condition (as shown by the higher standard deviation); however, across conditions the average kinematics still follow the same profile. These results indicate that the exoskeleton was able to achieve transparency with respect to the desired range of motion of the participant, while

still providing enough assistance to significantly reduce the positive lower-limb biological joint work. This also confirms the validity of the joint work analysis by demonstrating that the improvements seen with the exoskeleton assistance are not primarily coming from a change in the way participants performed the tasks; however, we do note that upper body kinematics were not measured or investigated.

Data Processing & Biomechanical Modeling

We followed the same procedure for processing motion capture marker trajectories and GRF data as detailed in our previous work¹. Ground-truth joint kinematics and moments were computed in OpenSim using the Inverse Kinematics and Inverse Dynamics tools, respectively^{7,8}. The OpenSim gait2354 model was used as the base model for each subject-specific anthropometric model. To model the participant in the exoskeleton, the OpenSim model was first scaled based on a biological static pose and biological body mass from a neutral standing trial collected without donning the exoskeleton. This initial model ensured that the model's segment masses and sizes were accurate with those of each participant. This biological model was used to compute joint kinematics and moments during the No Exo trials.

To model the exoskeleton, cylindrical masses were placed on the shank and femur segments of the scaled biological model at locations that matched the exoskeleton actuators. The mass of each modeled actuator was 0.485 kg and was constant across participants. The mass of the exoskeleton aside from the actuators was added to the model torso segment since the additional weight was mainly from the electronics housed in the exoskeleton backpack. Finally, the markers were adjusted based on a static pose with the participant wearing the exoskeleton; however, the dimensions, mass, and inertia of each segment were held constant to maintain an accurate model of the participant's anthropometry when wearing the exoskeleton. The resulting OpenSim model was used to compute joint kinematics and moments during the Exo On, Zero Torque and model accuracy trials.

Detailed Description of Baseline Method

Because no other studies have tested an exoskeleton controller in such a wide array of tasks, we developed a best-case baseline method to compare against our joint moment estimator, implemented as an extension of previous mode-based exoskeleton controllers. Many exoskeleton controllers use an activity classifier to switch between exoskeleton control modes⁹⁻¹⁵. Within each state, different control laws can be applied¹⁵⁻¹⁹. We opted to use a 28-mode classifier approach for our best-case baseline since it matched our delineation of task groups in the rest of the paper. Although a 28-mode classifier would likely result in many mode misclassifications if trained and deployed online, substantially fewer classes would result in very poorly tuned control laws within each task group and increasing to a separate class for each subtask (66 mode classifier) would unreasonably increase possible misclassifications. This baseline is a best-case model because whenever possible we gave this approach ideal inputs that represent the ceiling of what a real system could achieve. For the classifier, this meant that we applied the model *post hoc* and assumed perfect classification accuracy, which currently would be nearly unachievable in a real scenario, especially given 28 classes to choose from.

We used data from the same 15 training participants that we used to train our deep learning approach to create the baseline method. We then tested it on data from the same 10 participants that we collected to verify our deep learning-based approach. To design the control laws to apply during each mode, we used the three categories of cyclic, impedance-like, and unstructured.

During cyclic tasks, many controllers rely on gait events and gait phase percentages to look up the corresponding biological moment^{20–22}. Thus, to create the control law during cyclic tasks, the stride averaged curves from the training data (Phase 1 and Phase 2) were used to design a single moment profile for the hip and one for the knee in each of the cyclic activities. We accomplished this by first segmenting the body mass-normalized ground-truth moment data by gait cycle, normalizing time to percent cycle, and then averaging the curves across participants for each subtask. We then averaged those subtask participant-averaged curves across all the subtasks within a larger task group. This gives a single profile of the total joint moment that can be applied to any task that is classified as belonging to that task group. Although any real-time gait phase estimator will have associated errors in the estimate, we chose again to use a best-case model where we assume that we have perfect gait phase estimates when we applied these splines to the real data.

During impedance-like tasks many controllers use spring and damper parameters along with joint angles and joint velocities to compute the desired torque^{18,19,23}. To create our impedance control laws (which were only active when the foot was on the ground), we concatenated the ground-truth body mass-normalized joint moment data, joint angle data, and joint velocity data across training participants and subtasks within each larger task group. We then used an optimizer (fmincon in Matlab) to find the optimal spring stiffness and damping to minimize the error between the impedance-based estimate of the total joint moment and the actual ground-truth values. This process was performed for each task group within the impedance-like category as well as separately for the hip and the knee. Although real-time implementations of this method would likely operate on joint angles and velocities from exoskeleton encoders, which may not resemble the biological kinematics due to soft tissue deformation and sensor noise, we again chose to use a best-case model where we provide the impedance equation with the actual biological joint angles and velocities based on motion capture and inverse kinematics.

Because standard control architectures have no methods for handling highly unstructured activities, we did not compute a baseline method for the unstructured tasks.

Statistical Analyses

The statistical analyses were conducted using Minitab v19 with an alpha level of significance of 0.05. Due to the experimental design, all statistical tests were computed using within-subject differences (i.e., repeated-measures) with the subjects incorporated as a random factor; all other factors were modeled as fixed. Additionally, all tests were two-sided. Our first hypothesis was that our joint moment estimator would significantly increase R^2 and significantly reduce RMSE and normalized MAE relative to the baseline method. To test this hypothesis, a two-way ANOVA was used to test for main and interaction effects between estimators (independent variable: our approach and the baseline method) and among activity types (independent variable: cyclic and impedance-like activities) with respect to model R^2 (dependent variable), corresponding to the results in Fig. 3b. This ANOVA was also repeated, using RMSE and normalized MAE as the

dependent variable, with corresponding results in Extended Data Fig. 5a & c, respectively. Prior to computing the results, the joint moment estimates and corresponding ground-truth values were time-aligned to account for any delays in the online joint moment estimator. In the case that the ANOVA found a significant main effect between the estimators, a *post hoc* multiple comparisons test using a Bonferroni correction was also run to evaluate pairwise differences between our approach and the baseline method. Since the Bonferroni correction can substantially reduce statistical power in the presence of many pairwise comparisons and since we were only interested in comparing the performance of our approach with the baseline method within each activity type, we only evaluated a subset of the possible pairwise comparisons during this test, which were determined *a priori* (i.e., before looking at the results). Specifically, the only pairwise comparisons considered were those that compared our approach to the baseline method within the same activity type. Given that Minitab does not support this type of planned comparison, we ran a full multiple comparisons test in Minitab and subsequently corrected the p-values manually based on the reduced set of comparisons.

Within each type of activity (cyclic and impedance-like), three additional two-way ANOVAs were used to test for main and interaction effects between estimators (independent variable) and among each activity (independent variable), which corresponds to the results presented in Extended Data Fig. 4 for R^2 , RMSE, and normalized MAE (dependent variables corresponding to the ANOVAs). In the case that an ANOVA found a significant effect between the estimators, a *post hoc* multiple comparisons test using a Bonferroni correction was also run to test for differences between our approach and the baseline method during each task. As described above, we again used a planned comparisons approach to evaluate the pairwise comparisons between the results of our approach and the baseline method only when each result pertained to the same task.

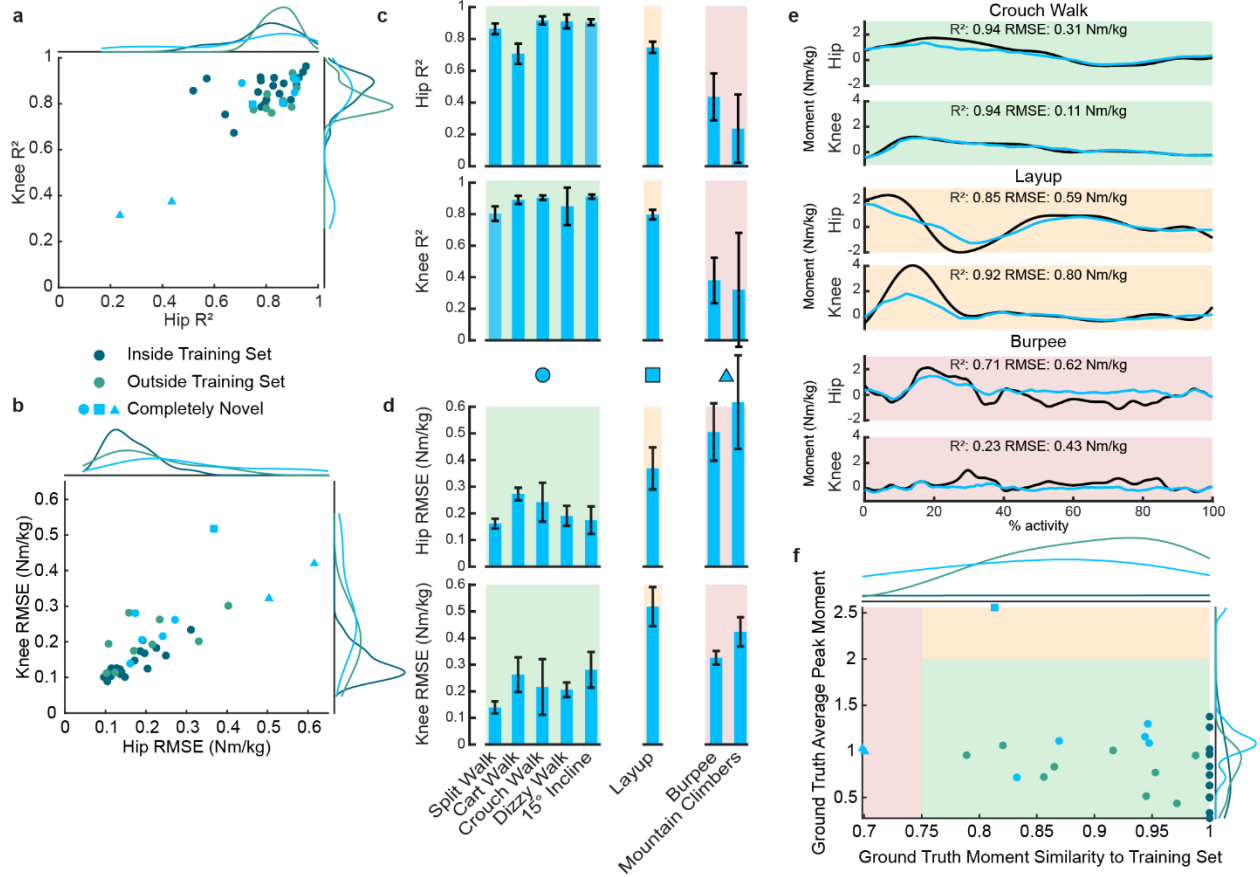
Additionally, we hypothesized that our task-agnostic controller would significantly reduce user metabolic cost and lower-limb positive biological joint work relative to the Zero Torque condition for each evaluated activity. To analyze metabolic cost (Fig. 4a) and lower-limb joint work (Fig. 4c), a one-way ANOVA was used to test for a significant main effect among the exoskeleton controllers (independent variable: Exo On, No Exo, and Zero Torque) for each activity (dependent variable was metabolic cost or lower-limb joint work depending on the activity). Following a significant effect, a multiple comparisons tests with a Bonferroni correction was used to determine pairwise differences within each activity. We did not test for interaction effects on metabolic cost or lower-limb joint work between exoskeleton controller and user activity.

Since each sensor modality provided unique information to the joint moment estimator (joint 1D positions and velocities from encoders, 3D linear acceleration and 3D angular rate from IMUs, and ground reaction force information from pressure insoles), we also hypothesized that retraining the model with a reduced sensor suite would significantly reduce R^2 . Similarly, since the model was trained on a specific sensor suite (thigh, shank, and foot IMUs, hip and knee encoders, and pressure sensing insoles), we hypothesized that the loss of a sensor during deployment would significantly reduce the R^2 of the model. To compare the effects of each sensor combination (independent variable) on overall estimator performance (Extended Data Fig. 1a & b), we used two separate one-way ANOVAs to test for a main effect in R^2 (dependent variable) at the hip and

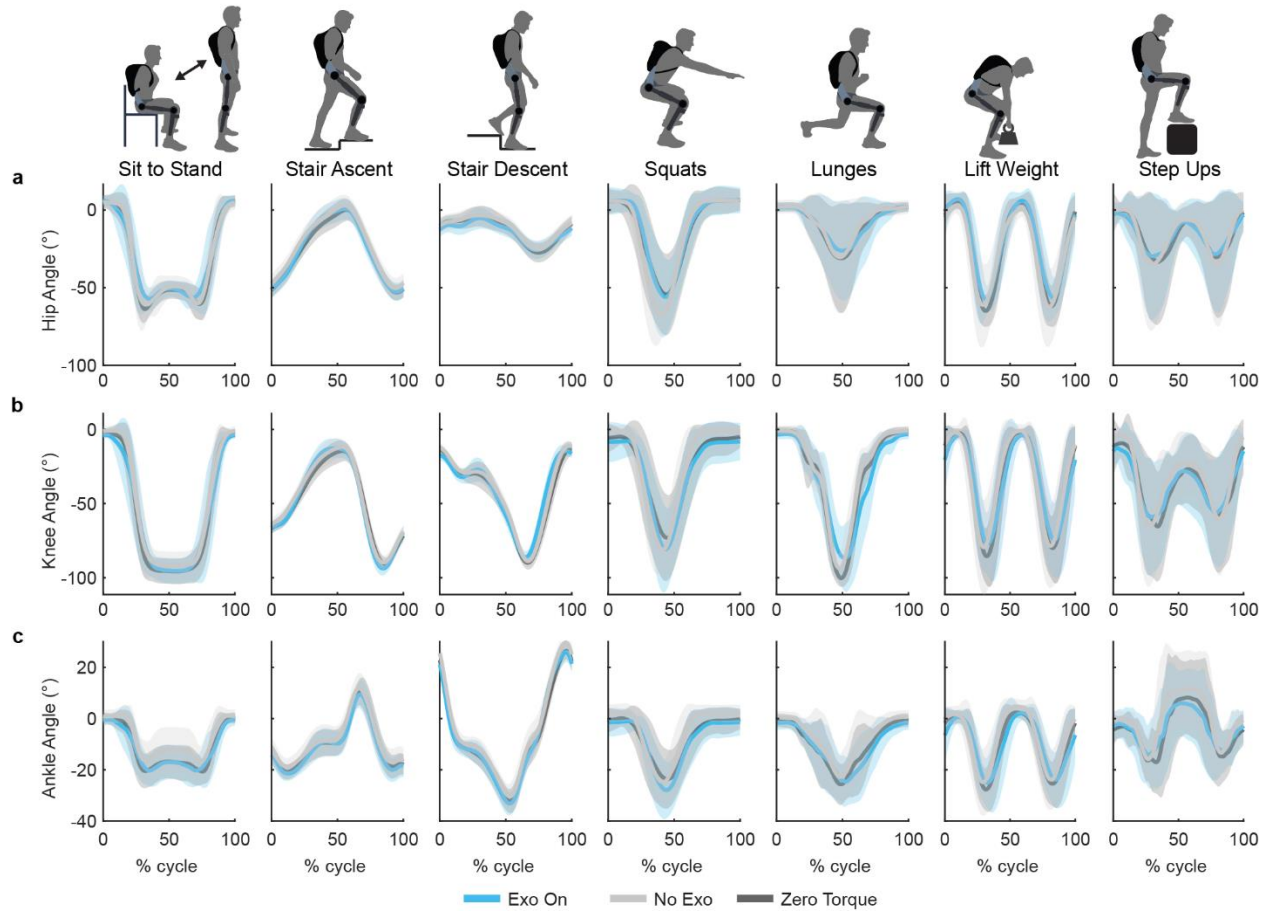
knee. The same approach was used to test the effect of loss of sensor data (independent variable) on R^2 (dependent variable), with corresponding results in Extended Data Fig. 1c & d. To evaluate pairwise comparisons between the All condition and each sensor ablation or dropout condition we used a planned comparisons approach with a Bonferroni correction as explained above.

5 Comparison to Offline Moment Estimators

While current biological moment estimators in the literature have only been validated on a small subset of activities and typically rely on offline analyses, our joint moment estimator performed similarly to or outperformed previous subject-independent hip and knee moment estimators^{12,24–27}. For instance, Wang *et al.* introduced a moment estimator evaluated on four non-locomotion tasks in addition to level ground walking and running with performance of 0.85 Nm/kg RMSE at the hip and 0.37 Nm/kg RMSE at the knee²⁸. Our results (hip RMSE 0.17 Nm/kg and knee RMSE 0.14 Nm/kg) substantially outperform this approach and were validated on more than 20 additional tasks, demonstrating the breadth of our approach. Averaging Hossain *et al.*²⁹ Pearson's correlation coefficient (r) across treadmill, ramps, and stair modes results in a r of 0.904 for hip flexion moment and 0.919 for knee moment. Our results for those modes (Pearson's correlation of 0.904 for the hip and 0.928 for the knee) are similar; however, the comparison isn't direct because different input sensor data were used in that study, and their approach estimated both moments and GRFs. For only ambulation modes (level walking, ramps, and stairs), Molinaro *et al.* reported offline RMSE results of 0.13 Nm/kg at the hip³⁰ with follow up work reporting an RMSE of 0.14 Nm/kg when deployed onboard a hip exoskeleton²⁵. We found similar performance during these activities (average RMSE of 0.15 Nm/kg); however, our model simultaneously estimated knee joint moments and was capable of estimating joint moments over a much broader range of tasks.



Supplementary Fig. 1. Novel task analysis. The (A) R^2 and (B) RMSE of the hip and knee are compared for three categories (19 tasks matching those inside the training set, 9 tasks matching those withheld from the training set, and 8 completely novel tasks). Each symbol represents a 3-subject average across the subjects that performed the novel tasks. The (C) R^2 and (D) RMSE of the hip and knee for the eight novel tasks are grouped for easier visualization based on the estimator's performance on the tasks. The bars represent the average across the three subjects and the error bars represent ± 1 standard deviation. (E) A single representative time series is shown from three selected activities where the ground-truth joint moments are presented in black behind our approach's estimate. The exact RMSE and R^2 for each specific time-series trace is included at the top of each graph. (F) A comparison of the ground-truth joint moments between the training set tasks and those outside the training set is presented in terms of moment magnitude and shape similarity. The colored backgrounds are based on the performance of the estimator and are included to give additional intuition that the estimator didn't perform as well on tasks where ground truth moments were far outside the bounds of the training set.



Supplementary Fig. 2. Lower limb joint angle comparison during joint work activities. Hip (A), knee (B), and ankle (C) angles are shown for three conditions: wearing the exoskeleton using the task-agnostic controller (Exo On), without wearing the exoskeleton (No Exo), and while wearing the exoskeleton without assistance (Zero Torque). The solid lines represent the subject-average joint angle and the shaded fill represents the standard deviation.

References:

1. Scherpereel, K., Molinaro, D., Inan, O., Shepherd, M. & Young, A. A human lower-limb biomechanics and wearable sensors dataset during cyclic and non-cyclic activities. *Sci. Data* **10**, 924 (2023).
- 5 2. Camargo, J., Ramanathan, A., Flanagan, W. & Young, A. A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions. *J. Biomech.* **119**, 110320 (2021).
3. Nuckols, R. W. *et al.* Mechanics of walking and running up and downhill: A joint-level perspective to guide design of lower-limb exoskeletons. *PLOS ONE* **15**, e0231996 (2020).
- 10 4. Franks, P. W. *et al.* Comparing optimized exoskeleton assistance of the hip, knee, and ankle in single and multi-joint configurations. *Wearable Technol.* **2**, e16 (2021).
5. Sawicki, G. S., Beck, O. N., Kang, I. & Young, A. J. The exoskeleton expansion: improving walking and running economy. *J. NeuroEngineering Rehabil.* **17**, (2020).
- 15 6. Browning, R. C., Modica, J. R., Kram, R. & Goswami, A. The Effects of Adding Mass to the Legs on the Energetics and Biomechanics of Walking. *Med. Sci. Sports Exerc.* **39**, 515 (2007).
7. Delp, S. L. *et al.* OpenSim: Open-Source Software to Create and Analyze Dynamic Simulations of Movement. *IEEE Trans. Biomed. Eng.* **54**, 1940–1950 (2007).
- 20 8. Seth, A. *et al.* OpenSim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. *PLOS Comput. Biol.* **14**, e1006223 (2018).
9. Kim, J. *et al.* Reducing the metabolic rate of walking and running with a versatile, portable exosuit. *Science* **365**, 668–672 (2019).
- 25 10. David Li, Y. & Hsiao-Weeksler, E. T. Gait mode recognition and control for a portable-powered ankle-foot orthosis. in *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)* 1–8 (2013). doi:10.1109/ICORR.2013.6650373.
11. Laschowski, B., McNally, W., Wong, A. & McPhee, J. Environment Classification for Robotic Leg Prostheses and Exoskeletons Using Deep Convolutional Neural Networks. *Front. Neurorobotics* **15**, (2022).
- 30 12. Kang, I., Molinaro, D. D., Choi, G., Camargo, J. & Young, A. J. Subject-Independent Continuous Locomotion Mode Classification for Robotic Hip Exoskeleton Applications. *IEEE Trans. Biomed. Eng.* **69**, 3234–3242 (2022).
13. Camargo, J., Flanagan, W., Csomay-Shanklin, N., Kanwar, B. & Young, A. A Machine Learning Strategy for Locomotion Classification and Parameter Estimation Using Fusion of Wearable Sensors. *IEEE Trans. Biomed. Eng.* **68**, 1569–1578 (2021).
- 35 14. Qian, Y. *et al.* Predictive Locomotion Mode Recognition and Accurate Gait Phase Estimation for Hip Exoskeleton on Various Terrains. *IEEE Robot. Autom. Lett.* **7**, 6439–6446 (2022).
15. Medrano, R. L., Thomas, G. C., Keais, C. G., Rouse, E. J. & Gregg, R. D. Real-Time Gait Phase and Task Estimation for Controlling a Powered Ankle Exoskeleton on Extremely Uneven Terrain. *IEEE Trans. Robot.* **39**, 2170–2182 (2023).

16. Kang, I., Kunapuli, P. & Young, A. J. Real-Time Neural Network-Based Gait Phase Estimation Using a Robotic Hip Exoskeleton. *IEEE Trans. Med. Robot. Bionics* **2**, 28–37 (2020).
- 5 17. Shepherd, M. K., Molinaro, D. D., Sawicki, G. S. & Young, A. J. Deep Learning Enables Exoboot Control to Augment Variable-Speed Walking. *IEEE Robot. Autom. Lett.* **7**, 3571–3577 (2022).
18. Huo, W. *et al.* Impedance Modulation Control of a Lower-Limb Exoskeleton to Assist Sit-to-Stand Movements. *IEEE Trans. Robot.* **38**, 1230–1249 (2022).
- 10 19. Shepherd, M. K. & Rouse, E. J. Design and Validation of a Torque-Controllable Knee Exoskeleton for Sit-to-Stand Assistance. *IEEEASME Trans. Mechatron.* **22**, 1695–1704 (2017).
20. Kang, I. *et al.* Real-Time Gait Phase Estimation for Robotic Hip Exoskeleton Control During Multimodal Locomotion. *IEEE Robot. Autom. Lett.* **6**, 3491–3497 (2021).
- 15 21. Seo, K., Lee, J., Lee, Y., Ha, T. & Shim, Y. Fully autonomous hip exoskeleton saves metabolic cost of walking. in *2016 IEEE International Conference on Robotics and Automation (ICRA)* 4628–4635 (2016). doi:10.1109/ICRA.2016.7487663.
22. Seo, K., Lee, J. & Park, Y. J. Autonomous hip exoskeleton saves metabolic cost of walking uphill. in *2017 International Conference on Rehabilitation Robotics (ICORR)* 246–251 (2017). doi:10.1109/ICORR.2017.8009254.
- 20 23. Yang, X. *et al.* Spine-Inspired Continuum Soft Exoskeleton for Stoop Lifting Assistance. *IEEE Robot. Autom. Lett.* **4**, 4547–4554 (2019).
24. Dorschky, E. *et al.* CNN-Based Estimation of Sagittal Plane Walking and Running Biomechanics From Measured and Simulated Inertial Sensor Data. *Front. Bioeng. Biotechnol.* **8**, (2020).
- 25 25. Molinaro, D. D., Kang, I. & Young, A. J. Estimating human joint moments unifies exoskeleton control, reducing user effort. *Sci. Robot.* **9**, eadi8852 (2024).
26. Molinaro, D. D., Park, E. O. & Young, A. J. Anticipation and Delayed Estimation of Sagittal Plane Human Hip Moments using Deep Learning and a Robotic Hip Exoskeleton. in *2023 IEEE International Conference on Robotics and Automation (ICRA)* 12679–12685 (2023).
- 30 doi:10.1109/ICRA48891.2023.10161286.
27. Forner-Cordero, A., Koopman, H. J. F. M. & van der Helm, F. C. T. Inverse dynamics calculations during gait with restricted ground reaction force information from pressure insoles. *Gait Posture* **23**, 189–199 (2006).
28. Wang, H., Basu, A., Durandau, G. & Sartori, M. A wearable real-time kinetic measurement sensor setup for human locomotion. *Wearable Technol.* **4**, e11 (2023).
- 35 29. Hossain, M. S. B., Guo, Z. & Choi, H. Estimation of Lower Extremity Joint Moments and 3D Ground Reaction Forces Using IMU Sensors in Multiple Walking Conditions: A Deep Learning Approach. *IEEE J. Biomed. Health Inform.* 1–12 (2023) doi:10.1109/JBHI.2023.3262164.

30. Molinaro, D. D., Kang, I., Camargo, J., Gombolay, M. C. & Young, A. J. Subject-Independent, Biological Hip Moment Estimation During Multimodal Overground Ambulation Using Deep Learning. *IEEE Trans. Med. Robot. Bionics* **4**, 219–229 (2022).