Article

Learning-based 3D human kinematics estimation using behavioral constraints from activity classification

Daekyum Kim [1,2,3], Yichu Jin [1,4], Haedo Cho [1,4], Truman Jones [1], Yu Meng Zhou[1], Ameneh Fadaie[1], Dmitry Popov[1], Krithika Swaminathan [1] & Conor J. Walsh [1]✉

Inertial measurement units offer a cost-effective, portable alternative to lab-based motion capture systems. However, measuring joint angles and movement trajectories with inertial measurement units is challenging due to signal drift errors caused by biases and noise, which are amplified by numerical integration. Existing approaches use anatomical constraints to reduce drift but require body parameter measurements. Learning-based approaches show promise but often lack accuracy for broad applications (e.g., strength training). Here, we introduce the Activity-in-the-loop Kinematics Estimator, an end-to-end machine learning model incorporating human behavioral constraints for enhanced kinematics estimation using two inertial measurement units. It integrates activity classification with kinematics estimation, leveraging limited movement patterns during specific activities. In dynamic scenarios, our approach achieved trajectory and shoulder joint angle errors under 0.021 m and 6.5°, respectively, 52% and 17% lower than errors without including activity classification. These results highlight accurate motion tracking with minimal inertial measurement units and domain-specific context.

Human motion capture has long been a core component of fields like biomechanics, clinical research, sports science, and entertainment[1–4]. Optical camera-based motion capture systems have been used as the gold standard for measuring human partial or whole-body kinematics, providing errors within 1-mm for movement trajectory estimation and within 1°–3° for joint angle estimation[5]. However, these camera-based systems are often limited by high costs, limited workspace, and visual occlusion issues[4,6,7].

To address these limitations, inertial measurement units (IMUs) have been widely explored as a cost-effective and wearable alternative for measuring human body movements[8]. IMUs contain accelerometers, gyroscopes, and often magnetometers, which measure linear accelerations, angular velocities, and magnetic fields, respectively[9]. By numerically integrating the accelerometer

and gyroscope measurements, human limb positions and joint angles can be estimated in varied environments. Despite being cheap, portable, and ubiquitous, IMUs exhibit a wide range of kinematic tracking accuracies compared to camera-based motion capture systems, with some reports showing shoulder joint errors close to 10°[5,10].

The main challenge with IMUs for kinematics estimation is the presence of time-varying bias and noise in the raw linear acceleration and angular velocity measurements. From numerical integration, these biases and noise result in drift errors that accumulate over time[11,12]. To address the IMU drift issue, existing approaches utilize kinematic constraints that are intrinsic to human anatomy to constrain drift accumulation. The Movella system from Xsens uses 17 IMUs placed on all body segments and leverages full-body kinematic relationships to

---

[1]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. [2]School of Mechanical Engineering, Korea University, Seoul, Republic of Korea. [3]School of Smart Mobility, Korea University, Seoul, Republic of Korea. [4]These authors contributed equally: Yichu Jin, Haedo Cho. ✉e-mail: walsh@seas.harvard.edu

mitigate drift by constraining the estimation to biomechanically plausible configurations[13]. Other approaches have aimed to leverage similar information but with a reduced number of sensors. Prior work has shown that leveraging an individual's joint range of motion can constrain knee angle estimates to a specific range using just two IMUs[14]. Another example with a single IMU leverages the relationship between the direction of the forearm and the position of the arm to estimate the trajectory of both wrist and elbow[15]. These approaches have demonstrated that kinematic constraints can effectively reduce IMU drift. However, the performance of these methods is dependent on the precision of body parameter measurements (e.g., limb length or joint range of motion, ROM) and the accuracy of the kinematics constraint formulations.

Learning-based methods can automatically infer the kinematic relationships between different body segments and generate drift-corrected kinematic estimates with a few inertial sensors. One approach trained machine learning models with synthetic datasets from optical motion capture (OMC) systems to measure full-body kinematics with only six IMUs[16,17]. Learning-based approaches have also been utilized to estimate partial-body kinematics, like wrist or elbow joint trajectory, using a single IMU on the wrist[18,19].

Extending beyond the use of kinematic constraints, there are still opportunities to further improve IMU-based motion tracking accuracy. Specifically, information about the activity a person is conducting could be used as a behavioral constraint to further reduce tracking errors from IMU drift. Such a concept has been applied during walking tasks via the zero velocity update method (ZUPT)[20]. During walking, there is a period when the foot remains stationary relative to ground; this information has been harnessed to re-initialize numerical integration per gait cycle, effectively mitigating any potential long-term drift across strides[21]. These instances of stationary body points have been estimated with heuristic[22] or machine learning algorithms[17,22] using accelerometer and/or gyroscope signals. However, despite its effectiveness, especially in the demonstrated walking task, ZUPT is mostly limited to cyclic activities with distinct zero velocity intervals, which may not always exist for more random upper body movements. Furthermore, in the context of machine learning-based kinematics estimation, ZUPT functions as a post-processing step that is applied to the output of a baseline machine learning model[17], rather than being an end-to-end model. Consequently, the accuracy of estimated kinematics is still limited by the performance of the baseline machine learning model.

In this work, we developed and evaluated an end-to-end machine learning model, Activity-in-the-loop Kinematics Estimator (AIL-KE), that incorporates human behavioral constraints within a learning-based kinematics estimation model. AIL-KE learns and leverages the behavioral constraints inherent to specific activities by integrating activity classification information with the kinematics estimation. This behavioral constraint-based model is designed based on the understanding that human motion, despite its high dimensionality, exhibits limited patterns and reduced variability within a given activity[23,24]. To maximize practicality of a wearable sensing system, we limit the maximum number of IMUs used in this work to two and focused on partial-body kinematics estimation. We evaluated the performance of AIL-KE in two dynamic functional scenarios: (i) estimating wrist and chest trajectories during various strength training exercises and (ii) estimating shoulder joint angles during simulated industrial assembly work. The approach presented in this paper aims to address the challenges of obtaining accurate movement trajectories and joint angles using IMUs over prolonged periods (3-10 minutes) with minimal sensors by leveraging behavioral constraints.

Results
Activity-in-the-loop Kinematics Estimator
The AIL-KE is composed of three sub-structures: Activity Classifier (*AC*), Kinematics Regressor (*KR*), and Feature Aggregation Network (*FAN*). Figure 1 depicts an overview of AIL-KE. *AC* classifies the activity that a user is performing at every timestep. The inputs of *AC* are 3-dimensional accelerations, 3-dimensional angular velocities, and 4-dimensional unit quaternions obtained from IMUs. *KR* estimates either trajectories and velocities of IMUs or joint angles between two body segments at every timestep by taking the same inputs as *AC*. We used stacked Dilated Convolutional Neural Networks (DCNNs) for *AC* and *KR* because these networks have been widely used in IMU-based drift and noise reduction[11].

FAN is the core part of AIL-KE that incorporates the behavioral constraint of activity classification. *FAN* passes the activity classification information to the kinematics regressor. Specifically, the final hidden layer of each stack of *AC* is processed in *FAN*, which is then fed into each stack of *KR*.
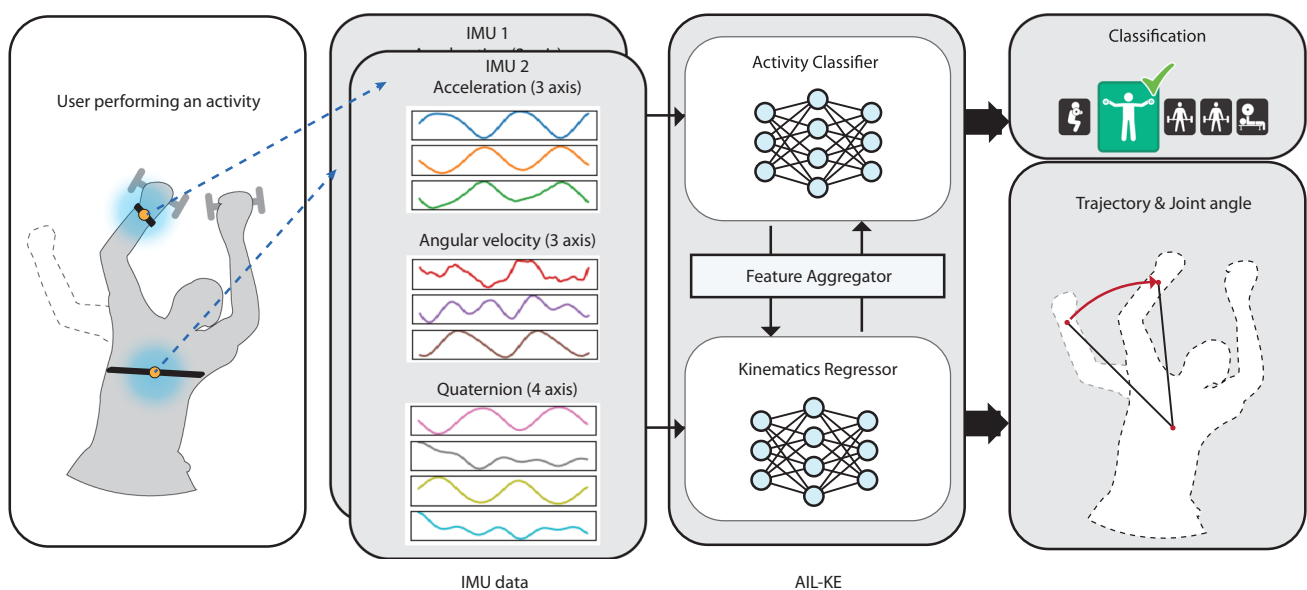

Fig. 1 | Overview of activity-in-the-loop kinematics estimator (AIL-KE). AIL-KE consists of activity classifier (AC), feature aggregator network (FAN), and kinematics regressor (KR).

Trajectory estimation during strength training exercises (Exer)

Data were collected from fifteen healthy participants (3 females; 28.1 ± 5.6 years) with two IMUs, one placed on their chest and one on their right wrist, to measure their 3D movement velocities and trajectories (Supplementary Fig. S1). The participants were asked to perform four sets of 11 different strength training exercises with 12 repetitions per set (full list shown in Supplementary Fig. S2). The four sets were performed at four different self-selected movement speeds: normal, slow, fast, and variable. Kinematics data from IMUs and a ground truth OMC system were time-synchronized and frame-aligned to ensure proper evaluation[25] (more detail in Methods). We used data from 11 randomly selected participants for training and one participant for validating the classification model. Data from the remaining three participants were used as the test dataset to evaluate the performance of the model. For kinematics estimation, we compared our method against two learning-based methods:

- Long short-term memory (LSTM): a commonly used deep learning structure for time-series data analysis.
- DCNN: the same model architecture as our proposed method, but without *AC* and *FAN*.

AC achieved an overall classification accuracy of 99.6%. It demonstrated 100% accuracy across all the exercise labels except for the triceps extension exercise (Fig. 2a). Triceps extensions were confused with Biceps curls approximately 6% of the time.

Overall, AIL-KE achieved a velocity error (in Root Mean Squared Error, RMSE) of $0.020 m/s$ versus DCNN with $0.040 m/s$ and LSTM with $0.063 m/s$ (Fig. 2b, c). The errors of AIL-KE were 48% and 67% lower than the errors of DCNN and LSTM, respectively. In addition, for trajectory estimation, AIL-KE achieved an RMSE of $0.02 m$ while RMSEs from DCNN and LSTM were $0.044 m$ and $0.050 m$, respectively. Each of these errors was calculated from the three test participants by averaging across both the chest and wrist IMU sensors. The average RMSE across the chest and wrist IMUs for AIL-KE was 52% and 58% lower than the errors of DCNN and LSTM, respectively. We further found that the improvement of AIL-KE over DCNN was consistent across different numbers of stacks of DCNNs at both the chest and wrist (Supplementary Table S4). Example time-series data from a participant performing Bench Press Exercise is depicted in Fig. 2d. Details of the performance of all models tested are tabulated in Supplementary Tables S1 and S2; we also include the performance of a Transformer-based model as a comparison. A movie containing exercise demonstrations and the corresponding estimated trajectories is shown in Supplementary Movie 1. In addition to time-series comparisons, we also compared the true and estimated mean and peak velocities, which are important metrics in strength training[26,27], across bench press repetitions in the test set (Supplementary Figs. S5 and S6). We find that both point metrics show strong correlations with the ground truth, with correlation coefficients (r) greater than 0.8[28].

We observed that AIL-KE had a lower RMSE of $0.017 m$ from the chest IMU compared to an RMSE of $0.023 m$ from the wrist IMU. It is worth noting that the wrist undergoes larger ranges of movements and velocities compared to the chest in most exercises. In strength training, movement velocity is self-selected and significantly differs depending on an individuals' workout strategy and level of fatigue[29,30]. Therefore, it is important to validate the performance of the models at different movement speeds.

We conducted a comparative analysis of movement speeds and assessed the corresponding effect on method performance (Fig. 2e, f). For AIL-KE, trajectory and velocity errors for fast speed were higher than the errors for the other movement speeds, with RMSEs of $0.022 m$ and $0.024 m/s$ for trajectory and velocity estimation, respectively. Still, these errors for the fast speed were only $0.002 m$ and $0.003 m/s$ higher than the average errors of AIL-KE across all the speeds. At the fast

speed, the trajectory error of AIL-KE was 55.1% and 63.8% lower than those of DCNN and LSTM, respectively, and the velocity error of AIL-KE was 45% and 70.1% lower than those of DCNN and LSTM respectively. Overall, AIL-KE outperformed the other two methods across all speeds. AIL-KE had a trajectory error standard deviation of $0.0007 m$ across all movement speeds. This value was seven times lower than those for DCNN and LSTM, indicating that AIL-KE had lower variability in performance across different speeds (more detail in Supplementary Tables S6 and S7).

We further analyzed the errors of the estimated trajectory based on different strength training exercises, depicted in Fig. 2g. Within these exercises, the barbell lunge had the largest error ($0.034 m$) across all models. Still, this error was lower than the errors of $0.069 m$ for DCNN and $0.088 m$ for LSTM. For all the other strength training exercises, AIL-KE had errors lower than $0.030 m$. For trajectory estimation, the RMSE when Triceps Extension was misclassified was $0.0172 m$, compared to the overall RMSE of Triceps Extension, which was $0.0165 m$. This small difference is likely due to the similarity in Triceps Extension and Biceps Curl in our participants as many biceps curl motions were performed with the dumbbells oriented vertically, similar to in triceps extensions. We also observed that the error of DCNN for Shoulder Press was unexpectedly higher than the errors for the other two models; further research is needed to systematically analyze these errors by collecting additional data and identifying their sources.

To assess the sensitivity of AIL-KE to individual variability, we tested its performance across the test set participants. The standard deviation values of the errors of our generalized model across the three test participants for the wrist IMU trajectory were as low as $6.6 \cdot 10^{-5} m$ (more detail in Supplementary Table S1). Similarly, to evaluate sensitivity to potential drift in the sensor, we conducted an additional experiment that simulated rotational shifts in sensor data. We found that the average error remains within $0.03 m$ for up to $7°$ of artificially added random sensor rotation (Supplementary Fig. S9).

We further performed the Mann–Whitney U test to determine the statistical significance of the errors in the peak-to-peak distance, i.e., the distance between the maximum and the minimum peaks, for LSTM, DCNN, and AIL-KE (more detail in Methods). We did not observe statistical significance between the peak-to-peak RMSEs of DCNN and AIL-KE when considering data from the entire exercise (Supplementary Fig. S8a). This was likely due to the high variance introduced by including all exercises. However, for each individual exercise, we observed statistical significance ($p<0.001$) between AIL-KE and DCNN (Supplementary Fig. S8b).

Orientation estimation in simulated industrial assembly work (Ind)

Six participants (all males; 30 ± 5.5 years) wore three IMUs, one on their chest and another two on their right and left upper arms to measure shoulder joint angles. They were asked to perform three tasks, simulating a typical industrial assembly workflow: overhead drilling, desk work, and treadmill walking. Each task was completed in three-minute intervals, totaling more than 10 minutes for each trial, with breaks and transitions included. Each participant performed five trials of activities for approximately one hour of data collected in total.

The inputs to AIL-KE were data from two IMUs (chest-right arm or chest-left arm), and the corresponding ground truth data were from motion capture cameras. Like in the strength training experiment, we aligned the motion capture system and IMU coordinate frames[25]. We performed Leave-One-Out Cross Validation (LOOCV) to assess the generalizability of AIL-KE across participants: data from each participant was used as a test dataset to evaluate the model's performance. The training data included four participants, while the validation data included one person. *AC* achieved an overall classification accuracy of 99.8% (Fig. 3a).
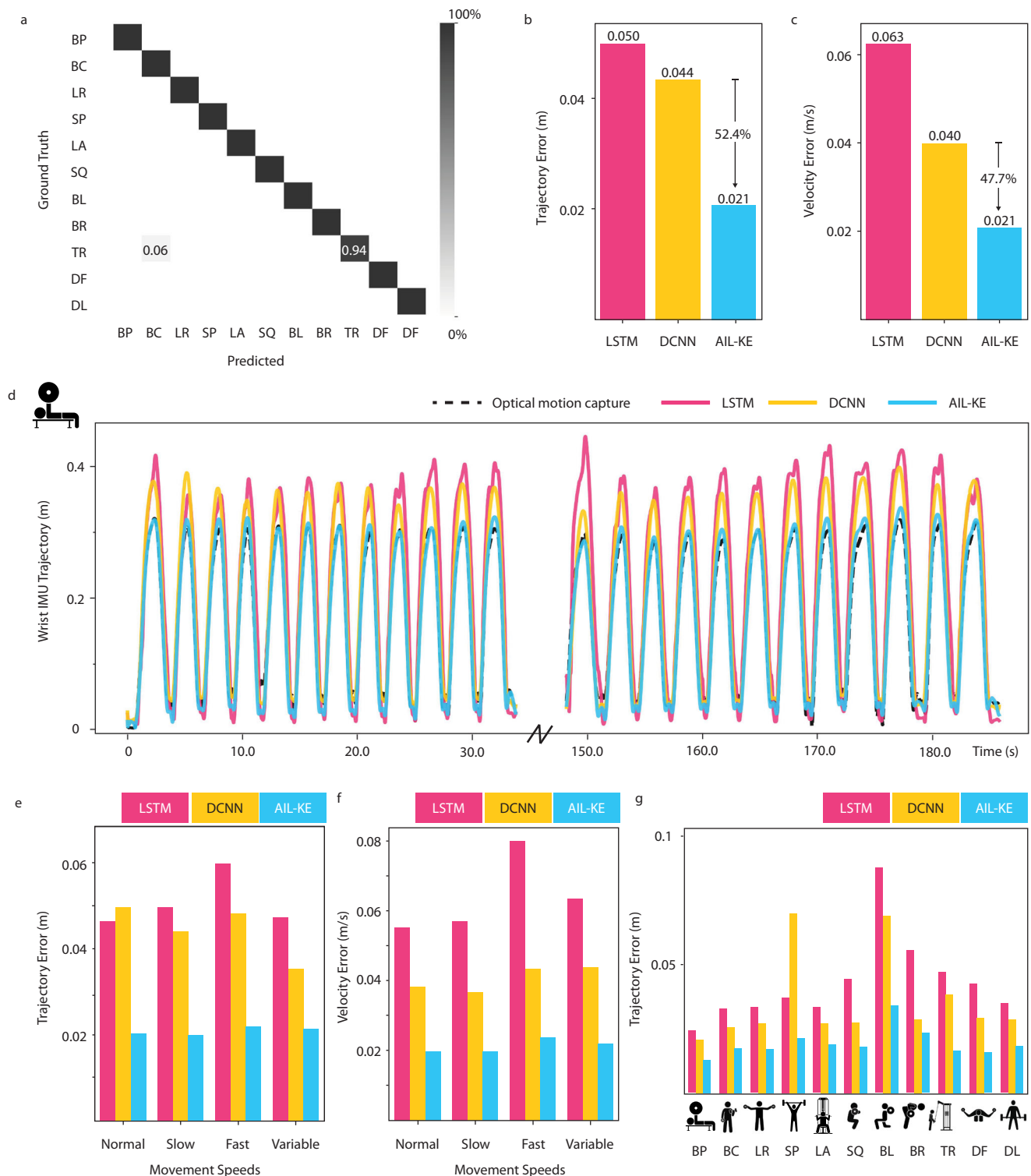
Fig. 2 | Trajectory estimation during strength training. a Confusion matrix across the following strength training exercise classes: Bench Press (BP), Biceps Curl (BC), Side Lateral Raise (LR), Shoulder Press (SP), Lat Pull Down (LA), Squat (SQ), Barbell Lunge (BL), Barbell Row (BR), Triceps Curl (TR), Dumbbell Fly (DF), Deadlift (DL). Unmarked black boxes indicate 100% accuracy. **b** Overall trajectory error across exercises in Root Mean Squared Error (RMSE). LSTM stands for Long Short-Term Memory, DCNN stands for Dilated Convolutional Neural Network, and AIL-KE stands for Activity-In-the-Loop Kinematics Estimator. **c** Overall velocity error across exercises in RMSE. **d** Example trajectory plot in the global z-direction for Biceps curl at a normal to variable speeds. **e** Trajectory error of wrist IMU for different movement speeds in RMSE. **f** Velocity error for different movement speeds in RMSE. **g** Trajectory error for different strength training exercises.

We evaluated the estimation results, which represent the average RMSE across participants for the 3D joint angles of the right and left upper arms (see Eq. 8 in Methods: Orientation calculation for more detail about how errors were computed) (Fig. 3b).

We evaluated the estimation results for 3D joint angles of the right and left upper arms (see Eq. 8 in Methods: Orientation calculation for more detail about how errors were computed) (Fig. 3b). We compared AIL-KE against three methods: LSTM, DCNN, and Xsens. In Xsens, we
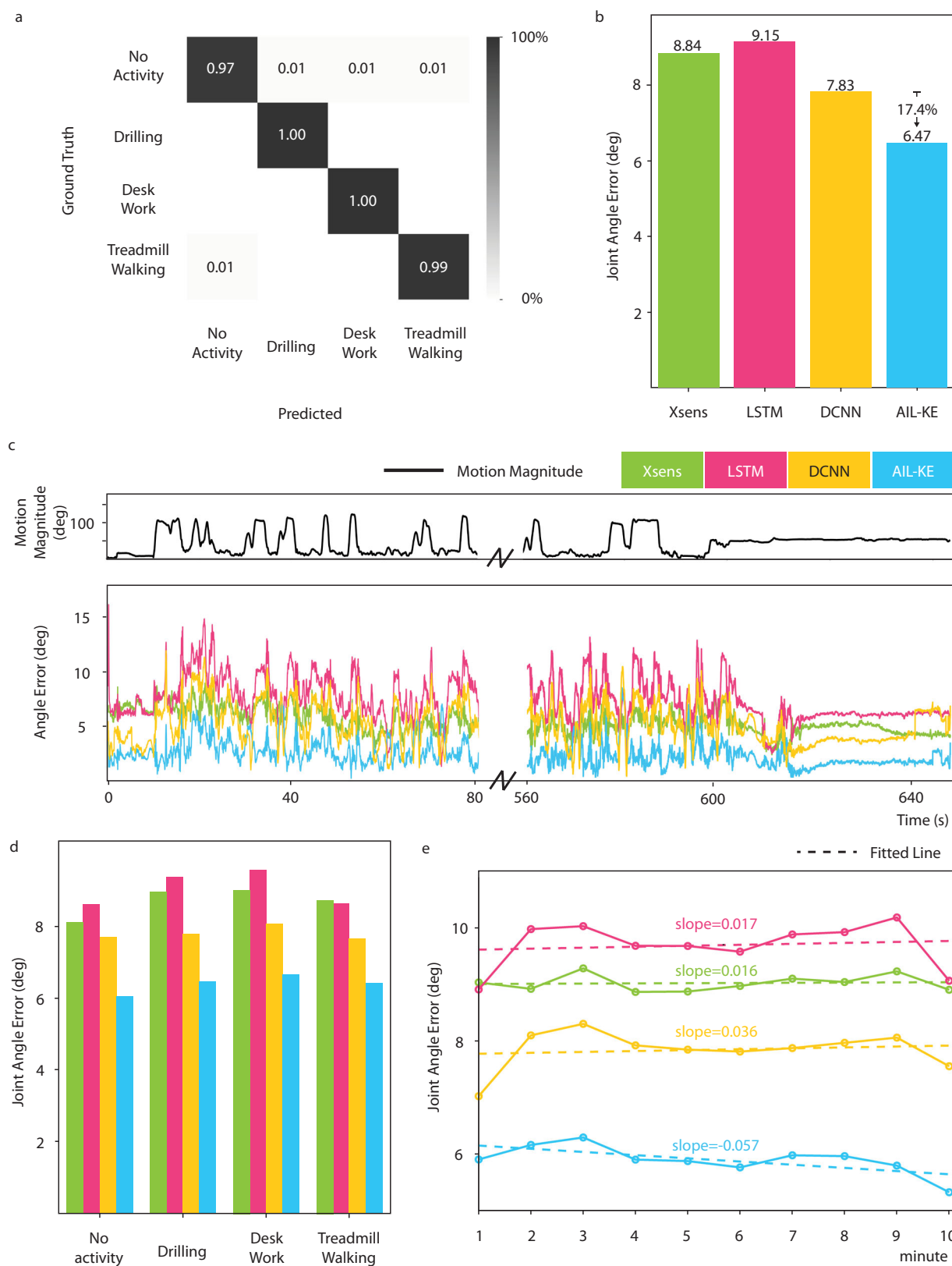
Fig. 3 | Shoulder angle estimation during functional movement. a Confusion matrix. **b** Overall angle error in Root Mean Squared Error. Xsens represents joint angle output from its proprietary sensor fusion algorithm. LSTM stands for Long Short-Term Memory, DCNN stands for Dilated Convolutional Neural Network, and AIL-KE stands for Activity-In-the-Loop Kinematics Estimator. **c** Plot of motion changes (in Motion Magnitude) and angle errors for different approaches. **d** Angle error for different activities. **e** Root Mean-Square Errors for IMU-based shoulder joint angles over 10 minutes. The fitted lines (dotted lines) represent linear fits across 10 minutes of the trial.

calculated the angular difference between the chest and left/right upper arms using the angles directly output from the Xsens' proprietary filters.

Overall, AIL-KE achieved an RMSE of 6.5 °, which was averaged across all participants through LOOCV from both shoulders, compared to DCNN with 7.83 °, LSTM with 9.15 °, and Xsens with 8.84 °. AIL-KE generated the best performance, with the angular error being 17.4%, 29.3%, and 26.8% lower than DCNN, LSTM, and Xsens, respectively, highlighting the effectiveness of our approach in reducing angle estimation errors. The RMSE in Euler angle representation is also tabulated in Supplementary Table S8. We further found that the improvement of AIL-KE over DCNN was consistent across different numbers of stacks for both the left and right shoulders (Supplementary Table S5). Numerical details regarding the performance of all the models tested are tabulated in Supplementary Table S3, which also includes the results of another popular machine learning model, Transformers. Figure 3c depicts a time-series error plot from a representative participant. The first row in Fig. 3c shows a time-series motion magnitude profile, which is calculated by finding the angular distance[25,31] between the shoulder kinematics in the first time frame and those in consecutive time frames (see Methods: Orientation magnitude calculation for more detail).

We further analyzed the errors in joint angles during different functional activities (Fig. 3d). The RMSE when the misclassification happened was 6.36 °, compared to the overall RMSE, which was 6.50 °. AIL-KE demonstrated the lowest error for all functional activities compared to the other approaches with a standard deviation of 0.25 ° across activities.

Further analysis on how the model's estimation performance changes over time, known as long-term drift, is presented in Fig. 3e. Our approach demonstrated a negative trendline slope from the first minute to the last minute of −0.057 °/min with the lowest joint angle errors across all minutes. Other approaches demonstrated positive trendline slopes smaller than 0.04 °/min, but with higher joint angle errors.

Like the strength training data, we assessed the sensitivity of AIL-KE to individual variability for joint angle estimation. The standard deviation values of the errors of our generalized model across participants were as low as 0.24 ° (more detail in Supplementary Table S3). In an additional experiment investigating the effect of shifts in sensor location during simulated industrial assembly work, we found that the average error remains below 6.8 ° with up to 13° of sensor rotation (Supplementary Fig. S9).

Discussion

This paper presents a behavioral constraint-based machine learning model, AIL-KE, which aggregates activity classification information to improve kinematics estimation accuracy. AIL-KE outperformed other learning-based approaches used for comparison, including an equivalent model architecture without the feature aggregation network, for applications in strength training (Exer) and industrial work (Ind).

Our approach achieved enhanced kinematics tracking performance by incorporating activity classification features as additional behavioral constraints. The strategy of using additional information to improve the performance of a machine-learning model has been widely adopted in various studies[32,33]. Within the field of motion kinematics estimation, studies used additional sensory modalities, such as full-body IMUs or visual information, to enhance model performance[32,33]. While effective, this approach requires adding additional sensors, which limits their practical use in the real world. The main advantage of our approach is that the additional information does not come from extra sensory inputs. Rather, classification information was derived using data from a minimal number of IMUs (two

IMUs in this case). We also show that expanding the size of the DCNN by making the DCNN layer deeper did not enhance model performance, suggesting that the additional classification information helps improve model performance (see Supplementary Tables S4 and S5 for more information).

The results suggest that aggregating classification information also helps reduce long-term drift, which is an active challenge in the field[34,35]. Our results showed Root Mean Squared Differences of <1° between the first and the last minutes, with a net negatively sloped RMSE trendline, representing near-zero drift over 10 minutes. Previous studies have explored traditional filtering-based approaches including, complementary filter and Kalman filter to reduce long-term drift but mainly focused on lower-limb joint angle estimation[34,35], or simulations using a robotic arm[36]. A previous study on lower-limb joint angle estimation conducted 10-minute trials and obtained linear fits to RMSEs over time with slopes of −0.14 to +0.17 °/ min[33]. The study reported that this result was on par with the result obtained from the proprietary filter from Xsens. In our case, the proprietary filter from Xsens also had a trendline slope less than 0.1 °, which aligns with results from the previous study. However, it demonstrated error values more than twice as large as AIL-KE across the trial. These results suggest that our method is both accurate and robust to drift over the span of 10 minutes, but further work is needed to understand the performance over hours or days. For example, while a robotic arm is not sensitive to sources of error inherent to a human arm such as the relative movement of anatomical structures (e.g., skin-to-bone displacement), this approach may allow for rapid characterization and iteration of IMU-based estimation methods under idealized conditions[36].

Accurate estimation of human movement is challenging because the same activity can be done with different movement patterns[37]. The standard deviations of AIL-KE's errors across test set participants, which were $6.65 \cdot 10^{-5} m$ for Exer and 0.24 ° for Ind, were lower than the other learning-based approaches used for comparison (see Supplementary Tables S1 and S3 for more detail). The lower standard deviations imply that there is less variability in estimation performance across unseen participants. The Normalized Root Mean-Square Deviation (NRMSD) across participants on test data, which evaluates the dispersion of data across participants, was the lowest with AIL-KE for both trajectory and angle estimation (detail in Supplementary Information Tables S1 and S3). In particular, during Exer, the wrist IMU NRMSD was less than 4% for trajectory and velocity estimates. Similarly, during Ind, the NRMSD averaged across both shoulders for joint angle estimates was also less than 4%. An NRMSD value closer to 0 indicates that the errors across participants are similar. Previous studies considered NRMSDs values of less than 4% as acceptable against individual variability for joint angle estimation[38,39]. Moreover, the NRMSDs for AIL-KE estimates were less than half of those from DCNN. Furthermore, prior work on angle estimation on the same participants across days using IMUs reported an NRMSD of 10% for the simple flexion/extension tasks and slightly under 20% for complex tasks[40]. Compared to this, the NRMSD value of AIL-KE across participants is considered low; albeit examining different joint angles (shoulder angles in this study vs. thorax and lumbar spine angles in Graham et al[40].). While these results support the potential use of the AIL-KE across individuals without concerns of sensor-to-segment misalignment, we expect there is a possibility to further improve accuracy by using sensor-to-segment calibration approaches proposed by other studies[41,42].

The approach introduced in this paper has a broad range of practical applications with the potential for utilization in commercial wearable devices. As an example, the range of motion and movement velocity of a body part lifting weights are important in strength training as they provide information regarding injury risk and muscle

development[19,30]. While other groups have studied wearable IMUs to measure movement velocity during strength training, there are challenges due to inaccurate velocity estimates. For example, in one study, moderate to weak correlations of $r = 0.62$ for mean velocity and $r = 0.49$ for peak velocity compared to ground truth during bench press exercise were found[43]. Here, we showed that AIL-KE results in strong correlations of $r = 0.81$ and $r = 0.88$ for mean and peak velocity, respectively, during bench press exercise across movement speeds (see Supplementary Figs. S5 and S6). Velocity measures are also important for estimating muscle strength, which is closely related to physical function, risk of injury, and neuromuscular fatigue[19,29,44]. The improved estimates from AIL-KE may enable future work to accurately estimate muscle strength changes using IMUs. Future work should include rigorous biomechanical analysis[45] to evaluate AIL-KE for sports-related applications.

Another application investigated in this paper was estimating joint kinematics and posture during overhead industrial work. Overhead tasks in which the arm is elevated for extended durations are known to be a significant contributing factor to work-related musculoskeletal disorders, such as shoulder disorder[46,47]. We evaluated the performance of AIL-KE for longer than 10-minutes and found that shoulder angle estimation accuracy during the last minute was at least 20% better than with the other approaches we investigated (see Supplementary Table S3 for more detail). Overall, the RMSE of AIL-KE at the shoulder joint was less than 6 °. Given that the range of joint angles for typical hand/tool positionings during overhead work is reported to be 70 °[47], this performance corresponds to less than 10% error across the range of motion. Our method provides accurate information on shoulder elevation angles of an individual against long-term drift, which is essential for ergonomics applications, such as risk assessment and injury prevention[46,47]. This information could further be incorporated into wearable assistive robots[47].

Our paper has several directions for future work. First, the effect of the complexity of the AC architecture on the performance of AIL-KE has not yet been evaluated. Further investigation is needed to determine whether a smaller AC model architecture can achieve the same level of accuracy. Second, we did not evaluate AIL-KE on IMUs from different vendors. Because each IMU has unique characteristics, such as sensor bias and noise[48], applying a pre-trained AIL-KE model to data from different IMUs may result in degraded performance. Future research should evaluate the performance of AIL-KE across IMUs from different manufacturers. If performance degradation is observed with different IMU products, transfer learning methods could be a promising approach to mitigate this issue[49], by pretraining AIL-KE with one type of IMU and fine-tuning with IMUs from a different vendor.

This paper presents an approach, the AIL-KE, that accurately estimates human kinematics using two IMUs. It consists of an end-to-end machine learning model incorporating human behavioral constraints for enhanced kinematics estimation by leveraging limited patterns and reduced variability in motion during specific activities. Our results show that by incorporating human activity information, AIL-KE could estimate the movement kinematics and 3D joint angles more accurately than the same model without activity information. We expect that AIL-KE will be also compatible with other learning-based partial-body[18,19] and full-body[16,17] kinematics estimation approaches to further enhance estimation performance.

Methods
Participant & data collection for Exer
The IMUs (Bosch BNO0030, Bosch, Germany) were connected to the Beaglebone Black (Texas Instrument, USA) to measure 3D acceleration, 3D angular velocity, and 3D orientation (represented as 4D unit quaternions) data at 100 Hz. The quaternion values were obtained from the internal Kalman filter of the IMUs. Each IMU was mounted on a custom 3D printed case with four motion capture markers on each of

the corners to determine the orientation of the IMU (Supplementary Fig. S1). IMU and OMC data were time-synchronized using a $5V$ analog trigger signal. A $5V$ signal was also used to obtain start and end times for each exercise, which were used for labeling the dataset prior to classification.

Data were collected on fifteen healthy participants (3 females; 28.1 ± 5.6 years) with two IMUs, one placed on their chest and one on their right wrist, to measure their 3D movement velocities and trajectories (Supplementary Fig. S1). Participants performed the following strength training exercises, each for 12 repetitions, in randomized order: Bench Press, Biceps Curl, Side Lateral Raise, Shoulder Press, Lat Pull Down, Squat, Barbell Lunge, Barbell Row, Triceps Curl, Dumbbell Fly, and Deadlift (Supplementary Fig. S2). The four sets were performed at four different self-selected movement speeds, normal, slow, fast, and variable. Six participants had one to three years of experience in strength training, another six had about one year or less of experience, and the remaining three had no prior strength training experience. While we provided instructions on performing the exercises before data collection, we did not verify whether the participants executed the strength training exercises with the correct form. We asked participants to place the sensors themselves and place them tightly to minimize movement during activities. Data were collected in accordance with Harvard Institutional Review Board (Protocol IRB-20-1847). We used data from 11 randomly selected participants for training and one participant for validation of the classification model. Data from the remaining three participants were used as the test dataset to evaluate the performance of the model.

Participant & data collection for Ind
Six participants (all males; 30 ± 5.5 years) wore three IMUs, one on their chest and another two on their right and left upper arms, to measure shoulder joint angles. Each participant performed 5 sets of overhead drilling, desk work (such as typing and note-taking), and treadmill walking (Supplementary Fig. S4). Each task was 3 minutes. We have "no action" as an additional label to indicate any activities performed transitioning among the three tasks. The time duration of "No action" between tasks was decided by each participant for each trial, ranging between 60 seconds to 90 seconds. We used Xsens MTI-3 IMUs collected at $100Hz$. Each IMU was mounted on a custom 3D printed case with four motion capture markers on each of the corners (Supplementary Fig. S3). IMU and OMC data were time-synchronized using a $5V$ analog trigger signal. This trigger also provided times for the start and end of a functional activity, which were used for classification. Data were collected in accordance with the Harvard Institutional Review Board (Protocol IRB19-1321). We performed LOOCV to assess the generalizability across participants. The training data included four participants, while the validation data included one person.

IMU and OMC coordinate frame definition
To ensure a fair comparison between IMU and OMC measurements, it is crucial to understand and align the coordinate frames of the two systems. As illustrated in Supplementary Fig. S10, the IMU sensor frame (SF) is defined by the physical placement of the sensing chip within the IMU, while its inertial frame (IF) is defined by the direction of gravity and the Earth's magnetic North. Conversely, the body frame of OMC (BF) is defined by four markers rigidly mounted on the IMU case, and its lab frame (LF) is defined using an OMC L-frame calibration tool that was placed flat on the ground at the start of the data collection.

The relationship between the sensor and inertial frames of the IMU, and the body and LFs of OMC can be mathematically expressed as follows:

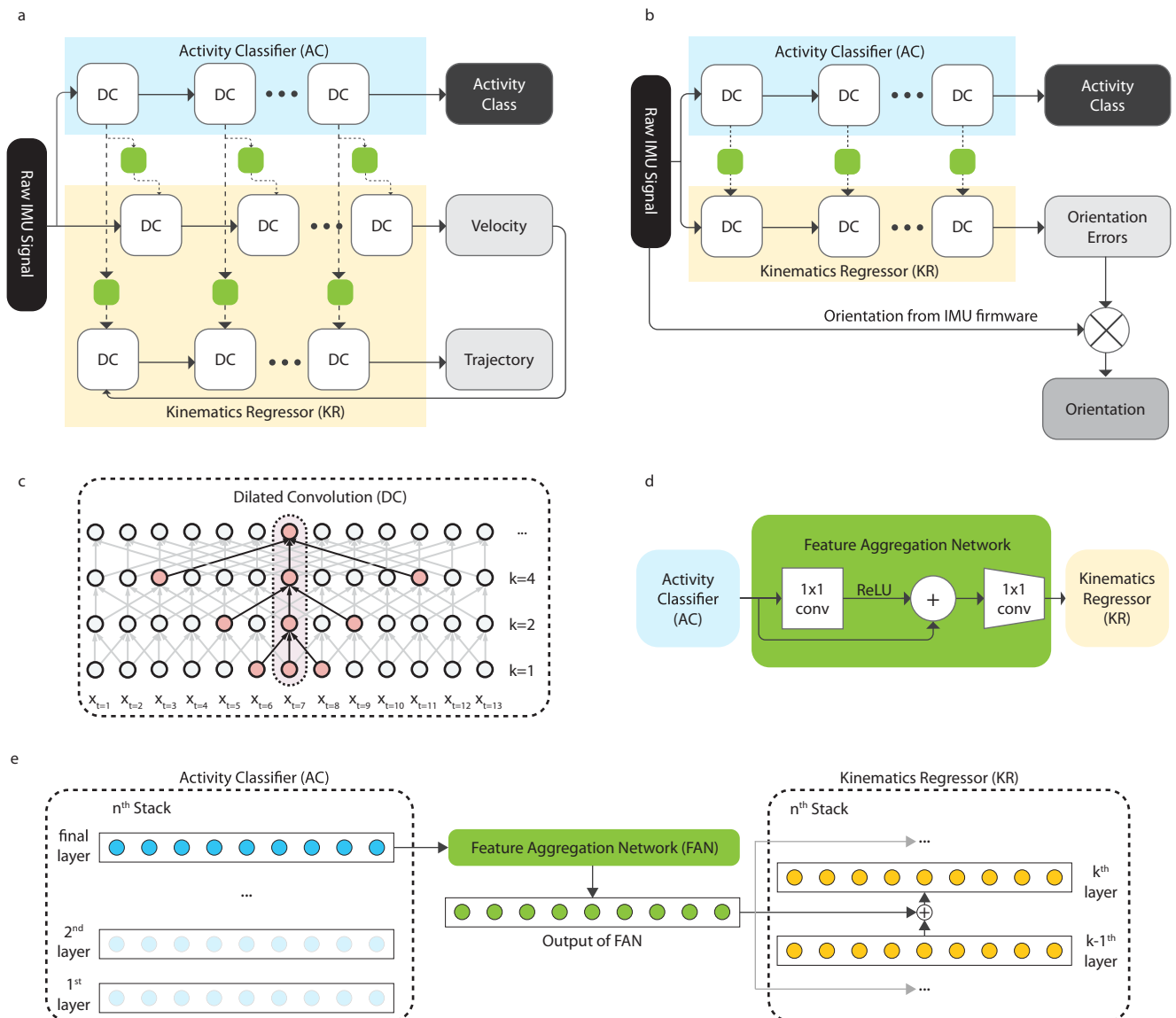$$q_{BF}^{LF} = q_{IF}^{LF}\, q_{SF}^{IF}\, q_{BF}^{SF} \qquad (1)$$

Fig. 4 | Detailed view of our model. a Overview schematic of how information from Activity Classifier (AC) is incorporated with Kinematics Regressor (KR) for velocity and trajectory estimation. **b** Overview schematic of how information from Activity Classifier (AC) is incorporated with Kinematics Regressor (KR) for joint angle estimation. **c** Detail view of Dilated Convolution layers (DC). The highlighted region in red shows how the features are operated in dilated convolution. **d** Detailed view of Feature Aggregation Network (FAN). 1×1 conv represents the one-by-one convolution layer, and ReLU represents the Rectified Linear Unit. **e** Detailed view of how features from AC are incorporated with KR through FAN.

where q represents the unit quaternion of the coordinate frame in subscript, expressed in the coordinate frame in superscript. Specifically, q_{BF}^{IF} and q_{SF}^{IF} correspond to the OMC and IMU orientation measurements, respectively. The term q_{IF}^{LF} and q_{BF}^{SF} are the unknown misalignments between the IMU and OMC coordinate frames. These misalignments were determined using an optimization-based frame alignment method presented in our prior work[25].

For Ind, the 3D shoulder orientations are calculated using the following equation:

$$q_{shoulder} = q_{arm}^{torso} = (q_{torso})^*q_{arm} \qquad (2)$$

Where $(q)^*$ denotes the conjugate of quaternion q, and q_{torso} and q_{arm} represent the unit quaternions of the torso and upper arm, respectively, as measured by either the IMU or OMC. This equation assumes that q_{torso} and q_{arm} are expressed in the same coordinate frame (IMU Inertial Frame in this case).

Detailed description of AIL-KE
We present an end-to-end machine learning model incorporating human behavioral constraints for enhanced kinematics estimation using IMU sensors. In this study, we used two IMU sensors, but the number of IMU sensors for AIL-KE is not limited. We study two applications for AIL-KE: velocity and trajectory estimation (Fig. 4a) and 3-dimensional joint angle estimation (Fig. 4b). Although in this paper, we used separate models for trajectory and joint angle for specific purposes (i.e., Exer and Ind), these models can be merged to estimate all metrics in an end-to-end manner. The trajectory estimation model (Fig. 4a) used global accelerations, angular velocities, and quaternions from IMUs - one on the chest and the other on the wrist−as an input to predict 1) AC: exercise class $\{c_1, ..., c_t\}$ and 2) KR: velocity $V = \{v_1, ..., v_t\}$ and trajectory $\Phi = \{\varphi_1, ..., \varphi_t\}$ in each of the IMU global frames for every time frame $t = 1, ..., T$, where T is the time length of each trial. The joint angle estimation model (Fig. 4b) used global accelerations, gyroscopes, and quaternions from the IMUs on the chest and each

shoulder to predict activity class $\{c_1, ..., c_t\}$ for AC, and 2) quaternion angle errors $\{e_1, ..., e_t\}$ for KR at every time frame $t = 1, ..., T$. The output of KR is then multiplied by quaternions obtained through initial IMU calibration. AIL-KE is composed of stacked Dilated Convolutional Neural Networks, shown in Fig. 4a, b with DC (Fig. 4c)[50,51] and Feature Aggregation Network or FAN (Fig. 4d). Each Dilated Convolutional Neural Network, depicted in Fig. 4c, was composed of dilated 1-d convolutions[52] with a dilation rate of $2^0, 2^1, 2^2, ..., 2^d$ and kernel size 3. This was followed by the Rectified Linear Unit (ReLU) activation function and a 1×1 convolution. The output of the 1×1 convolution is then summed with the input as a means of skip connection. The stacked dilated convolution structure allows the model to take temporal data with variable time lengths while the maximum dilation rate, i.e., 2^d must be smaller than the total time length of the one data sample, T.

FAN is a structure that provides activity classification information to KR (Fig. 4d). The last hidden layer (h_i) of each DCNN in AC was processed using FAN, which was summed with the output of each dilated convolutional neural network in KR. FAN was composed of point-wise convolution blocks and ReLU. Each h_i was fed into a one-by-one convolution layer, followed by ReLU. The output was summed with h_i as a residual structure, such that $F(h_i) + h_i$, where $F(h_i)$ is a 1×1 convolution. This was further processed by an additional 1×1 convolution layer to reduce depth size to fit into each DCNN in KR.

For every layer of DCNN, we used 1. $\mathcal{L}_{AC}$: the Categorical Cross-entropy loss to minimize the classification error between ground truth and predicted for AC, and 2. $\mathcal{L}_{KR}$: the Mean Squared Error function to minimize the error between estimated and ground truth velocities and trajectories or joint angles for KR. The integrated loss equation is shown as follows:

$$\mathcal{L}_{AIL-KE} = \sum_{s=1}^{S} (\mathcal{L}_{AC} + \mathcal{L}_{KR}) \tag{3}$$

$$\mathcal{L}_{AC} = -\sum_{i=1}^{C} y_i \log(\widehat{y}_i) \tag{4}$$

$$\mathcal{L}_{KR} = \frac{1}{N} \sum_{i=1}^{N} \left(V - \widehat{V} \right)^2 + \frac{1}{N} \sum_{i=1}^{N} (\Phi - \widehat{\Phi})^2 \tag{5}$$

Where s is s^{th} stack given that $s \in \{1, 2, ..., S\}$, and N is the number of samples. V and Φ are the ground truth velocity and trajectory that can be obtained from motion capture cameras, and $\widehat{V}$ and $\widehat{\Phi}$ are the predicted velocity and trajectory based on IMU data.

For joint angle estimation, we used the following loss function for KR to minimize angular error, based on the quaternion inner product.

$$\mathcal{L}_{KR} = \frac{1}{N} \sum_{i=1}^{N} \arccos\left(\left| q_{gt} \cdot q_{pred} \right| \right) \tag{6}$$

Where q_{gt} is the ground truth quaternion obtained from the OMC and q_{pred} is the quaternion obtained after normalizing the model-predicted orientation. This loss is reported to have numerical issues as there is a discontinuous gradient in the interval $(-1, 1)$ at point 0, which results in extreme values at the points where $\arccos(|q_{gt} \cdot q_{pred}|) \rightarrow 0$. Therefore, we used a gradient clipping approach, where the error derivative is clipped to a threshold during backpropagation through the deep learning network, and the clipped gradients are used to update the weights.

Model training strategy
We first trained AC for the first 500 epochs. Once AC was trained, we fixed the weights of AC and then trained KR for 1000 epochs. Then, AC

and KR were trained together for another 500 epochs. We used 4 stacks of AC and KR. The hidden dimension was set to 64 for all layers, including DC and FAN. The maximum dilation rate for each stack was set to $2^9 = 512$. We used the Adam optimizer with a learning rate of 10^{-4} and weight decay of 10^{-7}. These parameters were determined by grid searches.

Existing models for comparison
We compared the performance of the AIL-KE approach against the following models:

- *DCNN*: We used the same DCNN structure without *FAN*, i.e., we only used *KR*. The model architecture is shown in Supplementary Fig. S7.
- Long short-term memory (LSTM): LSTM is a Recurrent Neural Network architecture that has input, forget, and output gates in each of its nodes. The forget gate determines what information to retain or discard by applying a sigmoid function, which either multiplies by a factor of 1 or 0. These gates allow the network to handle long-range dependencies that arise from the vanishing and exploding gradient issues. LSTM structures are extensively utilized for time-series data processing[53–55], particularly for estimating position and angle based on IMU data. Hyperparameters of the LSTM, such as the number of layers and feature size, were found by grid searches. We used a 3-layer LSTM with a hidden feature size of 128, followed by two linear layers, each with a hidden feature size of 128.
- Transformer: The Transformer architecture has been widely used for training large language models[56]. It is based on the scaled dot-product attention and self-attention mechanism, offering an alternative to traditional temporal models such as Recurrent Neural Networks. The hyperparameters, including the number of layers and feature size, were determined through grid searches. We used the encoder part of the Transformer, consisting of a two-layer Transformer with a hidden feature size of 128 and an attention head size of 8. Following the Transformer encoder, we added two fully connected layers with a hidden feature size of 256. The estimation results were tabulated in Supplementary Tables S1–S3.
- Xsens proprietary filter (Xsens): For angle estimation, we compared results from our model with the joint angle output from Xsens' proprietary sensor fusion algorithm. Xsens is one of the world's leading IMU companies and its proprietary algorithm is generally considered as the state-of-the-art. Joint angles were obtained by calculating the rotation matrices between chest and shoulder IMUs.

Orientation calculations
In the simulated industrial work experiment, the time-series motion magnitude was defined as the angular distance between the shoulder's orientation at the initial time frame, $q_{arm,0}^{torso}$, and its orientation at subsequent time frames, $q_{arm,t}^{torso}$. Specifically, the motion magnitude at a specific time frame, θ_t, is calculated using

$$\theta_t = 2 \cdot \arccos(Re(q_{arm,t}^{torso}(q_{arm,0}^{torso})^*)) \tag{7}$$

where $Re(q)$ denotes the real part of the quaternion q.

Similarly, the time-series error profile for shoulder orientation was calculated as the angular distance between the estimated shoulder orientation and the ground truth. Specifically, the orientation error at a specific time frame, ψ_t, is calculated as

$$\psi_t = 2 \cdot \arccos(Re(q_{est,t}(q_{OMC,t})^*)) \tag{8}$$

where $q_{est,t}$ represents the shoulder orientation estimated by the machine learning model at time frame, t, and $q_{OMC,t}$ represents the

ground truth orientation captured by the OMC system. This equation differs from Eq. 6 as it calculates angular distance at specific time frames, while Eq. 6 is a loss function for model training, leveraging numerical simplifications like the absolute operation for stability.

Statistical analysis on peak-to-peak errors

We calculated the peak-to-peak distances, i.e., the distance between the maximum and the minimum peaks, of the ground truth and AIL-KE, followed by calculating the RMSE between the positions of the ground truth and AIL-KE predictions. Then, the same operation was applied for DCNN and LSTM. We conducted the Mann–Whitney U test to determine statistical significance between the models, i.e., AIL-KE vs. DCNN, LSTM vs. DCNN, and AIL-KE vs. LSTM, using a significance level of 0.05.

Data availability

All data supporting the findings of this study are available within the article and its supplementary files. Any additional requests for information can be directed to and will be fulfilled by the corresponding author. Source data are provided with this paper.

Code availability

All code for this work will be made available from the corresponding author upon request.

References

1. Wang, L., Hu, W. & Tan, T. Recent developments in human motion analysis. *Pattern Recognit.* **36**, 585–601 (2003).
2. Carling, C., Bloomfield, J., Nelsen, L. & Reilly, T. The role of motion analysis in elite soccer. *Sport. Med.* **38**, 839–862 (2008).
3. Menache, A. Understanding motion capture for computer animation and video games. Ch. 1 (Morgan Kaufmann, 2000).
4. Van der Kruk, E. & Reijne, M. M. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *Eur. J. Sport Sci.* **18**, 806–819 (2018).
5. Morrow, M. M., Lowndes, B., Fortune, E., Kaufman, K. R. & Hallbeck, M. S. Validation of inertial measurement units for upper body kinematics. *J. Appl. Biomech.* **33**, 227–232 (2017).
6. Johansson, D., Malmgren, K. & Alt Murphy, M. Wearable sensors for clinical applications in epilepsy, Parkinson's disease, and stroke: a mixed-methods systematic review. *J. Neurol.* **265**, 1740–1752 (2018).
7. Chen, X. & Davis, J. Camera placement considering occlusion for robust motion capture. *Comput. Graph. Lab. Stanf. Univ. Tech. Rep.* **2**, 2 (2000).
8. Porciuncula, F. et al. Wearable movement sensors for rehabilitation: a focused review of technological and clinical advances. *PMR* **10**, S220–S232 (2018).
9. Ahmad, N., Ghazilla, R. A. R., Khairi, N. M. & Kasi, V. Reviews on various Inertial Measurement Unit (IMU) sensor applications. *Int. J. Signal Process. Syst.* **1**, 256–262 (2013).
10. El-Gohary, M. & McNames, J. Shoulder and elbow joint angle tracking with inertial sensors. *IEEE Trans. Biomed. Eng.* **59**, 2635–2641 (2012).
11. Brossard, M., Bonnabel, S. & Barrau, A. Denoising IMU gyroscopes with deep learning for open-loop attitude estimation. *IEEE Robot. Autom. Lett.* **5**, 4796–4803 (2020).
12. Jiménez, A. R., Seco, F., Prieto, J. C. & Guevara, J. Indoor pedestrian navigation using an INS/EKF framework for yaw drift reduction and a foot-mounted IMU. In: 2010 *IEEE 7th workshop on positioning, navigation and communication*, pp. 135–143 (2010, March).
13. Schepers, M., Giuberti, M. & Bellusci, G. Xsens MVN: consistent tracking of human motion using inertial sensing. *Xsens Technol.* **1**, 1–8 (2018).
14. Hidalgo, A. F., Lora-Millán, J. S. & Rocon, E. *IMU-based knee angle estimation using an extended Kalman filter*. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology *Society (EMBC)*, pp. 570–573 (2019, July).
15. Shen, S., Wang, H. & Roy Choudhury, R. *I am a smartwatch and I can track my user's arm*. In: Proceedings of the 14th annual international conference on Mobile systems, applications, and services, pp. 85–96 (2016, June).
16. Yi, X. et al. *Physical Inertial Poser (PIP): physics-aware real-time human motion tracking from sparse inertial sensors*. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13167–13178 (2022).
17. Jiang, Y. et al. Transformer inertial poser: real-time human motion reconstruction from sparse IMUs with simultaneous terrain generation. In: *SIGGRAPH Asia 2022 Conference Papers*, pp. 1–9 (2022, November).
18. Feigl, T. et al. RNN-aided human velocity estimation from a single IMU. *Sensors* **20**, 3656 (2020).
19. Wei, W., Kurita, K., Kuang, J. & Gao, A. Real-time limb motion tracking with a single IMU sensor for physical therapy exercises. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 7152–7157 (2021, November).
20. Fischer, C., Talkad Sukumar, P. & Hazas, M. Tutorial: implementing a pedestrian tracker using inertial sensors. *IEEE Pervasive Comput.* **12**, 17–27 (2013).
21. Arens, P. et al. Real-time gait metric estimation for everyday gait training with wearable devices in people poststroke. *Wearable Technol.* **2**, e2 (2021).
22. Wagstaff, B., & Kelly, J. LSTM-based zero-velocity detection for robust inertial navigation. In: 2018 IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–8 (2018, September).
23. Lee, C. S. & Elgammal, A. Coupled visual and kinematic manifold models for tracking. *Int. J. Comput. Vis.* **87**, 118–139 (2010).
24. Yao, A., Gall, J. & Van Gool, L. Coupled action recognition and pose estimation from multiple views. *Int. J. Comput. Vis.* **100**, 16–37 (2012).
25. Jin, Y. et al. A data-based approach to simultaneously align local and global frames between an Inertial Measurement Unit (IMU) and an optical motion capture system. In: 2022 *9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pp. 1–8 (2022, August).
26. Thompson, S. W. et al. A novel approach to 1RM prediction using the load-velocity profile: a comparison of models. *Sports* **9**, 88 (2021).
27. García-Ramos, A., Pestaña-Melero, F. L., Pérez-Castilla, A., Rojas, F. J. & Haff, G. G. Mean velocity vs. mean propulsive velocity vs. peak velocity: which variable determines bench press relative load with higher reliability? *J. Strength Cond. Res.* **32**, 1273–1279 (2018).
28. Akoglu, H. User's guide to correlation coefficients. *Turk. J. Emerg. Med.* **18**, 91–93 (2018).
29. Halson, S. L. Monitoring training load to understand fatigue in athletes. *Sports Med.* **44**, 139–147 (2014).
30. Weakley, J. et al. Velocity-based training: from theory to application. *Strength Condition. J.* **43**, 31–49 (2021).
31. Huynh, D. Q. Metrics for 3D rotations: comparison and analysis. *J. Math. Imaging Vis.* **35**, 155–164 (2009).
32. Von Marcard, T., Henschel, R., Black, M. J., Rosenhahn, B. & Pons-Moll, G. Recovering accurate 3d human pose in the wild using IMUs and a moving camera. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 601–617 (2018).
33. Trumble, M., Gilbert, A., Malleson, C., Hilton, A. & Collomosse, J. Total capture: 3d human pose estimation fusing video and inertial sensors. *In Proceedings of 28th British Machine Vision Conference*, pp. 1–13 (2017, September).

34. Al Borno, M. et al. OpenSense: an open-source toolbox for inertial-measurement-unit-based measurement of lower extremity kinematics over long durations. *J. Neuroeng. Rehabil.* **19**, 22 (2022).

35. Robert-Lachaine, X., Mecheri, H., Larue, C. & Plamondon, A. Validation of inertial measurement units with an optoelectronic system for whole-body motion analysis. *Med. Biol. Eng. Comput.* **55**, 609–619 (2017).

36. El-Gohary, M. & McNames, J. Human joint angle estimation with inertial sensors and validation with a robot rrm. *IEEE Trans. Biomed. Eng.* **62**, 1759–1767 (2015).

37. Fetters, L. Perspective on variability in the development of human action. *Phys. Ther.* **90**, 1860–1867 (2010).

38. Hou, J. et al. A pilot study of individual muscle force prediction during elbow flexion and extension in the neurorehabilitation field. *Sensors* **16**, 2018 (2016).

39. Sheng, Y., Liu, J., Zhou, Z., Chen, H. & Liu, H. Musculoskeletal joint angle estimation based on isokinetic motor coordination. *IEEE Trans. Med. Robot. Bionics* **3**, 1011–1019 (2021).

40. Graham, R. B., Dupeyron, A. & van Dieën, J. H. Between-day reliability of IMU-derived spine control metrics in patients with low back pain. *J. Biomech.* **113**, 110080 (2020).

41. Zhu, K., Li, J., Li, D., Fan, B. & Shull, P. B. IMU shoulder angle estimation: effects of sensor-to-segment misalignment and sensor orientation error. *IEEE Trans. Neural Syst. Rehabil. Eng.* **31**, 4481–4491 (2023).

42. Picerno, P. et al. Upper limb joint kinematics using wearable magnetic and Inertial Measurement Units: an anatomical calibration procedure based on bony landmark identification. *Sci. Rep.* **9**, 14449 (2019).

43. Van den Tillaar, R. & Ball, N. Validity and reliability of kinematics measured with PUSH band vs. linear encoder in bench press and push-ups. *Sports* **7**, 207 (2019).

44. García-Ramos, A. et al. Feasibility of the 2-point method for determining the 1-repetition maximum in the bench press exercise. *Int. J. Sports Physiol. Perform.* **13**, 474–481 (2018).

45. Pataky, T. C., Vanrenterghem, J. & Robinson, M. A. The probability of false positives in zero-dimensional analyses of one-dimensional kinematic, force and EMG trajectories. *J. Biomech.* **49**, 1468–1476 (2016).

46. Seeberg, K. G. V., Andersen, L. L., Bengtsen, E. & Sundstrup, E. Effectiveness of workplace interventions in rehabilitating musculoskeletal disorders and preventing its consequences among workers with physical and sedentary employment: systematic review protocol. *Syst. Rev.* **8**, 1–7 (2019).

47. Zhou, Y. M. et al. A portable inflatable soft wearable robot to assist the shoulder during industrial work. *Sci. Robot.* **9**, eadi2377 (2024).

48. De Agostino, M., Manzino, A. M., & Piras, M. Performances comparison of different MEMS-based IMUs. *In: IEEE/ION Position, Location and Navigation Symposium*, pp. 187–201 (2010).

49. Nuckols, R. W. et al. Design and evaluation of an independent 4-week, exosuit-assisted, post-stroke community walking program. *Ann. N.Y. Acad. Sci.* **1525**, 147–159 (2023).

50. Farha, Y. A. & Gall, J. MS-TCN: multi-stage temporal convolutional network for action segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3575–3584 (2019).

51. Park, J., Kim, D., Huh, S. & Jo, S. Maximization and restoration: action segmentation through dilation passing and temporal reconstruction. *Pattern Recognit.* **129**, 108764 (2022).

52. Oord, A. V. D. et al. Wavenet: a generative model for raw audio. *arXiv* http://arxiv.org/abs/1609.03499 (2016).

53. Kang, B. B. et al. Learning-based fingertip force estimation for soft wearable hand robot with tendon-sheath mechanism. *IEEE Robot. Autom. Lett.* **5**, 946–953 (2020).

54. Kim, D. et al. Eyes are faster than hands: a soft wearable robot learns user intention from the egocentric view. *Sci. Robot.* **4**, eaav2949 (2019).

55. Hur, B., Baek, S., Kang, I. & Kim, D. Learning based lower limb joint kinematic estimation using open source IMU data. *Sci. Rep.* **15**, 5287 (2025).

56. Wolf, T. Transformers: state-of-the-art natural language processing. *arXiv* https://arxiv.org/abs/1910.03771 (2020).

Acknowledgements

Author contributions
Conceptualization: D.K., H.C., and C.J.W. Methodology: D.K., Y.J., H.C., T.J., Y.M.Z., D.P., and C.J.W. Investigation: D.K., H.C., Y.J., T.J., Y.M.Z., and C.J.W. Analysis: D.K., Y.J., H.C., K.S., and C.J.W. Visualization: D.K., Y.J., H.C., A.F., K.S., Writing—original draft: D.K., Y.J., H.C., Y.M.Z., D.P., K.S., and C.J.W., Writing—review and editing: D.K., Y.J., H.C., K.S., and C.J.W., Supervision: C.J.W. Funding acquisition: C.J.W.

Competing interests

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-58624-6.

Correspondence and requests for materials should be addressed to Conor J. Walsh.

Peer review information *Nature Communications* thanks Frank Wouda, who co-reviewed with Alessandro Castellaz, Mathias Blandeau, and the other anonymous reviewer for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.