3D Knee and Hip Angle Estimation With Reduced Wearable IMUs via Transfer Learning During Yoga, Golf, Swimming, Badminton, and Dance

Jinxuan Li, *Student Member, IEEE*, Kezhe Zhu, *Student Member, IEEE*,
Dongxuan Li, *Student Member, IEEE*, Peiqi Kang, *Graduate Student Member, IEEE*,
and Peter B. Shull, *Member, IEEE*

Abstract—**Wearable lower-limb joint angle estimation using a reduced inertial measurement unit (IMU) sensor set could enable quick, economical sports injury risk assessment and motion capture; however the vast majority of existing research requires a full IMU set attached to every related body segment and is implemented in only a single movement, typically walking. We thus implemented 3-dimensional knee and hip angle estimation with a reduced IMU sensor set during yoga, golf, swimming (simulated lower body swimming in a seated posture), badminton, and dance movements. Additionally, current deep-learning models undergo an accuracy drop when tested with new and unseen activities, which necessitates collecting large amounts of data for the new activity. However, collecting large datasets for every new activity is time-consuming and expensive. Thus, a transfer learning (TL) approach with long short-term memory neural networks was proposed to enhance the model's generalization ability towards new activities while minimizing the need for a large new-activity dataset. This approach could transfer the generic knowledge acquired from training the model in the source-activity domain to the target-activity domain. The maximum improvement in estimation accuracy (RMSE) achieved by TL is 23.6 degrees for knee flexion/extension and 22.2 degrees for hip flexion/extension compared to without TL. These results extend the application of motion capture with reduced sensor configurations to a broader range of activities relevant to injury prevention and sports training. Moreover, they enhance the capacity of data-driven models in scenarios where acquiring a substantial amount of training data is challenging.**

Index Terms—Deep learning, prosthesis, activity recognition, reduced sensor set, transformer.

I. INTRODUCTION

WEARABLE lower-limb joint angle estimation using a reduced inertial measurement unit (IMU) sensor set facilitates the rapid and cost-effective assessment of sports injury risk [1], [2], [3], [4], [5], motion capture [6], [7], [8], [9], and improvement of motion techniques [1], [10], [11]. The IMU is a commonly used, lightweight, and affordable sensor that is often integrated with machine learning models to enable kinematic estimation [6], [10]. Traditional wearable motion capture systems often necessitate one IMU sensor per lower-limb segment to capture the 3D angles of each joint, resulting in the use of numerous sensors [12], [13]. Karatsidis et al. [12] employed a wearable system consisting of 17 IMUs. Utilizing the acceleration signals obtained from IMUs, they reconstructed the kinematics of the anatomical segments. This full IMU configuration can be intrusive, time-consuming, expensive, and susceptible to errors such as sensor swapping during mounting [6], [7]. In contrast, the reduced IMU sensor configuration offers advantages by placing IMU sensors on a subset of the total lower-limb segments [3], [14]. This configuration enhances user comfort, reduces setup time, mitigates wear-related difficulties, and decreases system costs compared to the full IMU configuration [1], [3], [6], [12]. One significant application of the reduced sensor configuration is sports performance monitoring. Athletes and fitness enthusiasts prefer using fewer IMUs, which provides greater freedom of movement and facilitates long-term daily monitoring [15]. This setup allows individuals to receive real-time information about their joint angles and movement techniques, giving them the ability to make adjustments to prevent injury risk and optimize performance [2], [15]. Furthermore, emerging human-computer interaction also benefits from a reduced sensor configuration [6]. This configuration enhances the convenience of interaction and enables the capture of diverse user motion poses, thereby improving the

overall user experience [7], [8], [9]. The most widely tested and practical sensor placement for the reduced IMU sensor configuration entails affixing three IMUs to the pelvis and shanks, utilizing one IMU per segment [1], [3], [6], [7], [8], [14], [16]. The prevalent methodology for kinematic estimation modeling involves a deep-learning approach that correlates IMU sensor data with joint angles [10]. Compared to physics-based methods, deep learning methods are favored for their superior accuracy and ability to extract sufficient information from fewer sensors [1], [2], [6], [7], [8], [9], [10], [11], [13], [16], [17].

The vast majority of existing research was implemented in only a single movement activity, typically walking [1], [2], [18], [19], [20], [21], [22], [23], [24]. Mundt et al. [1] trained a long short-term memory (LSTM) neural network to estimate 3D joint angles of the hip, knee, and ankle joints during self-selected speed-level walking. Hossain et al. [2] proposed a modular deep-learning model to estimate sagittal plane hip, knee, and ankle joint angles during various walking conditions (overground, treadmill, slope, and stair). Semwal et al. [19] approximated human gait trajectories using an LSTM model trained on human gait joint angle data generated in Opensim simulations. Given the inherent reliance of data-driven models on the training set, this model was not suitable for dynamic activities beyond gait. Semwal et al. [20] estimated joint trajectories during six walking-type activities: brisk walking, normal walking, very slow walking, moderate walking, jogging, and brisk walking. The differences in data distribution for these walking-type activities are minimal, so it is not possible to demonstrate the performance of the algorithm in different non-walking activities. Semwal et al. [22] proposed a personalized LSTM-CNN model trained on a feature set consisting of anthropometric parameters and walking speed to estimate the sagittal plane angles of hip, knee, and ankle joints. However, this model can only predict one-dimensional angles for each joint, and its accuracy significantly benefits from the highly periodic nature of walking motion. Although walking is a crucial movement in gait analysis and rehabilitation of impaired gait, various non-gait dynamic movements associated with sports, exercise, and rehabilitation also have a significant impact on injury risk and outcomes of therapeutic interventions [25], [26], [27], [28]. However, the feasibility of capturing joint angles using a reduced IMU set during complex and highly dynamic non-gait movements remain relatively unexplored. In contrast to the strong periodicity observed in walking, non-gait dynamic activities are characterized by robust non-periodic patterns, which may lead to reduced observability of joint angles under a reduced sensor configuration [6], [7], [17], thus posing challenges to accurate estimation. Huang et al. [6] proposed a bi-directional recurrent neural network that reconstructed the quaternion of all body segments using six body-worn IMUs. The model was validated on five user-defined uncertain motion categories, including controlled motion of the arms or legs, locomotion, full-body activities, and interaction tasks with objects. It achieved an angular error of 17.54° across these various movements [6].

However, it has not been tested in various common activities related to sports injury risk and exercise. Therefore, we proposed achieving three-dimensional knee and hip joint angle estimation with a reduced IMU sensor set in five non-gait activities, including yoga, golf, swimming, badminton, and dance. The five non-gait activity data we collected are highly relevant to global physical health and injury prevention and cover a range of activity levels from stationary to light, moderate, and vigorous activity [28], [29], [30], [31]. Additionally, we introduced an activity-aware hierarchical model to implement the simultaneous recognition of motion patterns and joint angle estimation across these five activities.

Another factor that limits the extension of joint angle estimation with reduced sensor configurations to a wider range of activity types, is the accuracy decline when deep-learning models test with unseen activities [3], [32], [33], [34], [35]. The process of collecting a substantial marker-based motion-tracking database for every unseen activity is time-consuming and costly [6], [36], particularly for dynamic activities characterized by higher variability. To our knowledge, current research on reduced sensor configuration has not explored improving the model's generalization ability to new activities. To tackle this challenge, we proposed a transfer learning (TL) approach that transfers the knowledge obtained from joint angle estimation of the source domain (known activity) to the target domain (unseen activity) [35], [37], [38]. TL effectively enhances the model's generalization to the unseen domain by leveraging the common knowledge of the source domain model and joint angle estimation [37], [39], [40]. In various research fields, TL has shown efficacy in reducing the required size of datasets from unseen domains. Zhang et al. [34] employed TL with existing datasets covering various movements to estimate joint torque for a new movement. They pre-trained the LSTM network to learn structural similarities between movements. Ameri et al. [38] utilized existing training data before an electrode shift and a TL method with pre-trained convolutional neural networks to address the issue of insufficient training data post-electrode shift. Although TL has been extensively explored in cross-domain activity recognition [35], [37], [40], [41], its performance in cross-activity scenarios for joint angle estimation using a reduced sensor configuration remains unclear. In addition to considering the model's generalization ability across diverse activities, we also emphasized a practical scenario where the model extends its generalization from semi-static activities to dynamic activities. Semi-static activities exhibit lower movement intensity than dynamic activities, exemplified by yoga poses [17]. This characteristic facilitates ease of execution for subjects and the quality of data obtained is relatively high, making semi-static activities suitable as source domain data sets. We hypothesized that the semi-static yoga pose dataset contains some similar features to the dynamic activity dataset. Therefore, the model's generalization ability from semi-static datasets to unseen dynamic activities can be enhanced through the TL method.

The primary contributions in this work are as follows:

- This study is the first to implement 3D knee and hip joint angle estimation using a reduced IMU sensor configuration during five non-gait activities, including yoga, golf, swimming, badminton, and dance activities.
- This is the first study to extend the generalization ability of joint angle estimation models to unseen activities. The proposed TL method utilizes a dataset of the known activity to transfer the general knowledge of mapping correlations between IMU sensing data and joint angles into a model of unseen activities. In particular, this method could reconstruct three-dimensional joint angles during dynamic activities using datasets derived from semi-static yoga poses.

In Section II, this study conducted the estimation of 3D knee and hip angles during five non-gait activities using a reduced IMU sensor configuration. An activity-aware-based hierarchical model with artificial neural networks was proposed to implement the estimation, enabling simultaneous recognition of motion patterns and joint angle estimation. This method is effective for estimating angles in known activities. In addition, we also proposed a novel TL method to enhance the model's generalization ability to unseen activities, especially unseen dynamic activities (Section II). We tested the TL method on twenty dual-activity transfer pairs to thoroughly assess its performance (Section II). The data collection process is elaborated in Section II. Section III presents the estimation results of the two proposed methods. In Section IV, we provided a comprehensive discussion of various aspects of the results, clarifying the strengths and limitations of the methods.

II. METHODS

To expand the application of motion capture with reduced sensor configurations to various activities, our study proposed two methods to estimate joint angles with a reduced sensor configuration in known and unseen activities, respectively (Fig. 1 (a)).

A. Activity-Aware-Based Hierarchical Model for Known Activities

An activity-aware-based hierarchical model (AAHM) was proposed for joint angle estimation of known activities in the training set. This model can also realize simultaneous motion pattern recognition and joint angle estimation, where motion pattern recognition provides additional activity-type information. Simultaneous motion pattern recognition and joint angle estimation may be more critical in multi-sport performance and risk monitoring [20], [42], [43], [44], [45]. Activity recognition can offer additional insights into the type of activity, facilitating the computation of motion-related metrics or precise control of risk thresholds for different activities [13], [34], [42], [46], [47]. AAHM can customize the joint angle estimation process by identifying specific activity types, especially in activities with multiple movement patterns. The AAHM consists of two stages (Fig. 1 (b)). In the first stage, an activity classification model is employed to recognize the specific activity types of the input samples. Then, samples are directed to the corresponding estimation model in the second
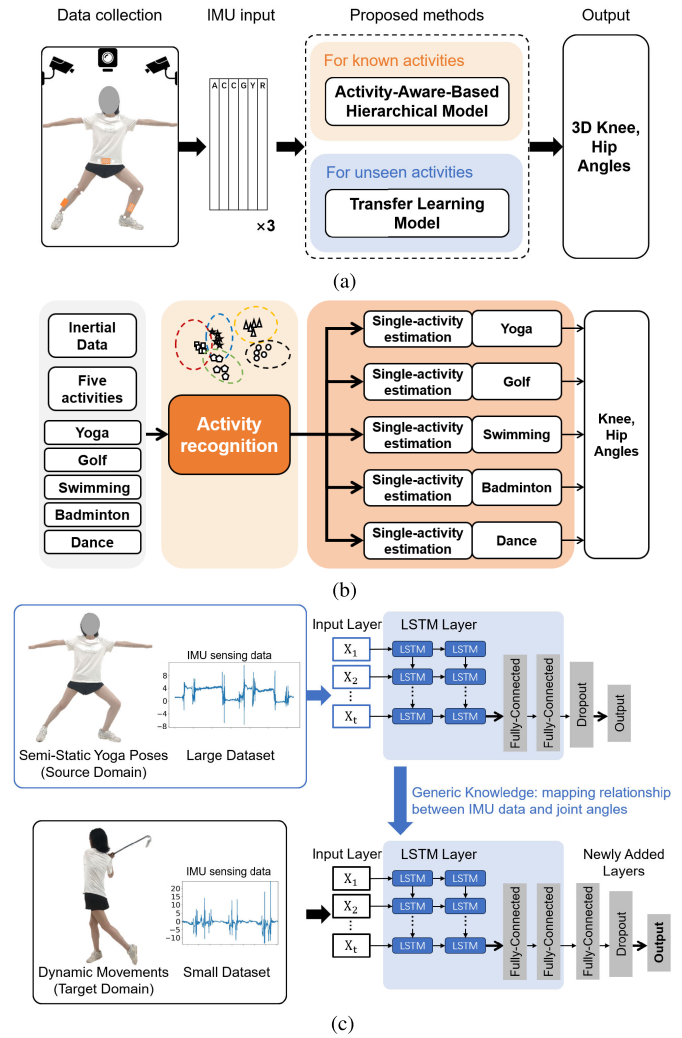

Fig. 1. (a) General flowchart for joint angle estimation of various movements with reduced sensor configurations. (b) Proposed activity-aware-based hierarchical model structure for simultaneous motion pattern recognition and joint angle estimation. The first stage is a motion pattern classification model that identifies the activity type of the input sample. Based on the identified activities, the sample is directed to the corresponding estimation model in the second stage. (c) Transfer learning model structure for generalizing between different activities, using the example of extending the yoga training model to dynamic activities. We obtained all layers preceding the dropout layer from the pre-trained source model and transferred them to the target model. The target model was integrated by the transfer layer and a newly introduced fully-connected layer, a dropout layer, and an output layer. Subsequently, we trained the target model using a minimal dynamic activity dataset.

stage based on the recognized activities. The second stage consists of five separate estimation models, each dedicated to a specific activity type and responsible for joint angle estimation (Fig. 1 (b)). By decoupling motion pattern recognition from the joint angle estimation process, AAHM incorporates motion pattern information as an additional constraint to improve joint angle observability. This is particularly beneficial for reduced sensor configurations where the information required for accurate joint angle estimation is limited [6], [9], [13], [47].

To reduce the complexity of the AAHM, we used a support vector machine (SVM) instead of a deep neural network as the

first stage of the model [47], [48]. Shallow machine learning techniques, including SVM, have a proven track record of effectively training accurate activity classification models [43]. We also tried other classification models in the first stage, such as k-nearest neighbor and random forest. Moreover, we noticed that note that the performance of the first stage has a minimal impact on the joint angle estimation of the second stage, which allows us to focus on selecting a simpler classification model without affecting the overall joint angle estimation accuracy of AAHM. We trained separate estimation models corresponding to each activity for the second stage of AAHM. The separate estimation model consists of two LSTM layers with 128 units, two fully-connected layers with 64 neurons each, and a dropout layer.

We compared the estimation accuracy between the AAHM and the approach of directly applying the five separate estimation models corresponding to each activity. The latter approach assumes 100% accuracy in activity classification for the first stage, effectively eliminating the influence of misclassification on the joint angle estimation. In this way, we can evaluate the benefits of incorporating activity awareness in the AAHM and determine the extent of improvement in joint angle estimation achieved through this hierarchical model. The activity classification model and separate estimation models in AAHM were trained and validated using the leave-one-subject-out (LOSO) cross-validation method. In this method, the activity data from one subject were used for testing, the activity data from another subject were used for parameter validation, and the activity data from the rest subjects were used for training. This method ensures that the data from the same subject do not appear in the training and test datasets, which is preferable to standard n-fold cross-validation [37].

B. Transfer Learning for Unseen Activities

To improve the model's generalization ability to unseen activities, we proposed a novel TL method that utilizes a dataset of known activities to transfer the common knowledge of the correlation mapping between IMU sensing data and joint angles to a model of unseen activities.

1) Pre-Trained Source Model on Semi-Static Yoga Poses: Initially, we required a pre-trained source model, which is exclusively trained on the inertial data of the semi-static yoga poses (Fig. 1 (c)). The pre-trained source model consists of two LSTM layers with 128 units, two fully-connected layers with 64 neurons each, and a dropout layer. The output layer is responsible for generating the joint angle estimations. The source model was trained using the LOSO cross-validation method. This method ensures rigorous evaluation of the model's performance across different subjects.

2) Knowledge Transfer Towards the Target Dynamic Activity Model: A TL technique was proposed to enhance joint angle estimation in dynamic activities, such as golf, swimming, badminton, or dance, even with a small dataset specific to each dynamic activity (Fig. 1 (c)). Note that the source model can transfer knowledge to only one dynamic activity model at a time.

According to the conclusions of Yosinski et al. [39], the initial layers of a neural network primarily learn generic features that apply to a wide range of tasks. As the network becomes deeper, subsequent layers specialize in learning task-specific features. The pre-training of the source domain model was performed using an abundant dataset of semi-static yoga poses. This allowed the model to learn to estimate joint angles from IMU data. However, since yoga movements primarily involve calibration actions on the lower limbs, although they directly incorporate the fundamental mapping relationship between IMU data and joint angles, they lack the specific motion patterns characteristic of dynamic activities (Fig. 1 (c)). To address this issue, replacing the source model's layers close to the output layer is generally recommended [34]. Then, the pre-trained model can be integrated into entirely new models for each dynamic activity (Fig. 1 (c)).

In our TL method, We obtained all layers preceding the dropout layer from the pre-trained source model and transferred them to the target model. These layers contain generic features relevant to the joint angle estimation task (Fig. 1 (c)). The target model was assembled by integrating the transferred layers with a newly introduced fully-connected layer containing 64 neurons, a dropout layer, and an output layer. During training, the weights of both the transferred layers and the newly introduced layer were updated using a minimal dataset of dynamic activity. This fine-tuning technique plays a crucial role in the TL framework. It automatically minimizes the domain discrepancy between the datasets, enabling the model to adapt to the new target activity with limited data.

3) Subject-Independent Training: The target model for the dynamic activity was also trained using the LOSO method. As a result, the TL-based model became subject-independent since its pre-training, knowledge transfer, and fine-tuning processes did not rely on any dynamic-activity data from the tested subjects. To achieve optimal training performance, the hyper-parameters of the models were tuned. The learning rate during the training process was adaptively adjusted based on the model's accuracy variation on the validation set. In the later stages of model training, a lower learning rate was utilized to avoid getting stuck in local optima and to ensure more stable and incremental changes in the trainable parameters.

4) Validation Strategies: We implemented two methods for joint angle estimation in dynamic activities: TL and a method denoted as NoTL, which directly applies the source model trained on a large amount of semi-static yoga pose data to the target dynamic activity. We then compared the performance of these two methods. Each dynamic activity served as a target domain in our study, resulting in four transfer tasks: transfer from yoga to golf, transfer from yoga to swimming, transfer from yoga to badminton, and transfer from yoga to dance.

To evaluate the methods, we used the LOSO cross-validation method. The training set consisted of yoga activity data from nineteen subjects for the NoTL and TL methods. In addition, 20% of the dynamic activity data from the nineteen subjects were set aside for the transfer and fine-tuning process in the TL method [49]. This process was repeated for all twenty-one subjects. The average performance of the models across the twenty-one iterations was reported as the final result [37]. In addition to transferring knowledge from the semi-static yoga poses to the four dynamic activities, we fur-

Fig. 2. Subject instrumentation layout with each orange box representing one IMU. Three IMUs were attached to each subject on the pelvis and shanks, with one IMU per segment. Four 3-marker clusters were used for motion tracking to calculate the ground truth for joint angles placed on each thigh and shank, respectively. Each shank IMU was fixed together with a 3-marker cluster.
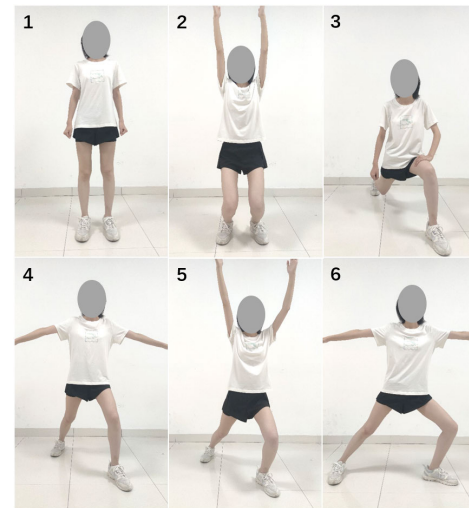

Fig. 3. Semi-static yoga poses used for reconstructing joint angles in dynamic activities. Subfigures 1 to 6 correspond to the following yoga poses: Standing Pose, Phantom Chair Pose, Dragon Pose, Initial stage of Warrior I Pose, Warrior I Pose, and Warrior II Pose. The data collection also included transitional movements for these poses.

ther explored transfer pairs involving different activities as the source and target domains. As a result, we obtained twenty transfer pairs for validation, considering the five activities in total ($5 \times 4 = 20$). These transfer pairs allowed us to thoroughly investigate the effectiveness of knowledge transfer between various activity domains and assess the model's generalization across different types of movements.

In addition to the LSTM network as the basic network for the proposed AAHM and TL methods, we also explored the application of transformer networks. The transformer consists of an encoder and a decoder, both of which are composed of stacked modules based on self-attention [50]. The bidirectional encoder representations from transformers (BERT) contains only one encoder and has been widely accepted for tasks such as text classification or named entity recognition [33]. Our transformer network is based on the BERT architecture and consists of an embedding layer, a positional encoding layer, and a stack of encoder layers. Each encoder layer comprises multi-head self-attention, position-based feedforward networks, residual connections, and layer normalization. Following this is a flattening layer and two fully-connected layers with 128 and 12 units, respectively. To prevent overfitting, a dropout layer was introduced between two fully-connected layers. We set the number of attention heads to 8, the number of encoder layers to 2, and the number of units in the hidden layer of the feedforward network to 128. The configuration of the transformer network in this study is very similar to a recently published study [33].

C. Data Collection

1) Subjects: Twenty-one subjects (12 males and 9 females; age: 22.8 ± 0.8; height: 1.71 ± 0.06 m; weight: 60.2 ± 8.2 kg) with no history of musculoskeletal disorders were recruited to participate in this study. All the subjects were healthy and given informed consent following the Declaration of Helsinki [51].

2) Markers and IMUs: Twelve reflective markers were placed on 12 anatomical landmarks to define the segments: left and right ilium anterior superior, left and right ilium posterior superior, left and right femur lateral epicondyle, left and right femur medial epicondyle, left and right fibula apex of lateral malleolus, left and right tibia apex of medial malleolus. Four 3-marker clusters were placed on each thigh and shank for motion tracking, placed on each thigh and shank, respectively (Fig. 2). The marker trajectories were captured using a ten-camera optical motion capture system (Vicon, Oxford Metrics Group, Oxford, U.K.). Three IMUs (MTw, Xsens, Netherlands) were securely strapped to the subject, with one IMU per segment on the pelvis and shanks. To reduce the influence of soft tissue and alleviate the occlusion of optical marker points, the shank IMUs were placed on the lateral side near the ankle. Although there was no rigid orientation requirement for the shank IMU in each subject, the general orientation of the shank IMU for all subjects was toward the lateral shank. The pelvis IMU was placed at the midpoint between the left and right Ilium Posterior Superior (LIPS and RIPS) on the back of the pelvis. This stable positioning indicated the subject's facing direction during the initial static phase of data collection. Each IMU outputs three-axis acceleration values and three-axis angular velocity values. Each shank IMU was affixed together with the 3-marker cluster above (Fig. 2). The IMUs and the optical motion capture system operated at a sampling rate of 100 Hz. The IMUs and optical motion capture system were electronically synchronized through a cable before starting data acquisition.

3) Experimental Procedure: Before experimenting, the researchers instructed the subjects to familiarize themselves with the activities they would be performing. Researchers also provided training on various dance and yoga movements to ensure the subjects were comfortable with the required actions. The data collection began with the subject performing a 5-second static neutral posture calibration, maintaining an upright stance with feet aligned to the marker line on the ground [52]. This ensures that all subjects are facing the same direction at the beginning of the experiment, thereby increasing the invariance of IMU data to human human-facing direction for improved model training and generalization [6]. Unlike traditional physics-based motion capture methods, this study omitted the sensor-to-segment calibration. The reason

for this decision is that deep learning models are susceptible to random errors introduced by variations in sensor placement orientation during training, ultimately enhancing the model's robustness for subject-independent applications [36]. Furthermore, it aligns with the common practice in deep-learning-based joint angle estimation studies, which typically avoid sensor-to-segment calibration [1], [2], [6], [7], [8], [9], [10], [11], [13], [16], [17].

The subjects performed various yoga movements for the yoga activity, mainly including Phantom Chair Pose, Dragon Pose, Warrior I Pose, and Warrior II Pose (Fig. 3). Each standard yoga pose was held for approximately 8 seconds to simulate the actual yoga exercise. The subjects were asked to swing an actual golf club at different amplitudes and speeds for the golf activity. During the swimming activity, subjects were required to simulate freestyle or breaststroke strokes while seated. The chair was diamond-shaped and the subject was seated on one corner of the diamond, allowing for a wide range of leg motion without contact with the chair. It was ensured that the subject's feet and legs remained in the air and did not have any contact with the ground or external supports. The pelvis (hips) functioned as a pivot point, facilitating leg movement. Additionally, subjects were permitted to support their hands backward on the two corners of the chair. This was also to meet the challenge of the common subject completing the breaststroke kick with only pelvis support. The subjects were asked to imitate their regular movements for the badminton activity, including swinging the racket in various directions and with different intensities. As a result, the magnitude and intensity of the activity varied among the subjects. The dance activity involved performing five distinct fitness dance movements, some selected from Zumba routines. The duration of each activity was two minutes. The reference knee and hip joint angles were calculated using the optical motion capture system and Visual3D software (C Motion, MD, USA).

4) Pre-Processing: The Vicon-captured optical data was processed in Visual3D (C-Motion, MD, USA) following the CAST procedure [53] to calculate the ground truth joint angles. This optical motion capture system determines ground truth joint angles through the following steps: a) calibration of the constant transformation between marker cluster coordinate systems and segments' anatomical coordinate systems during the initial N-pose (upright stance); b) tracking segment movement using three-marker cluster coordinate systems during activities; c) calculating relative Euler angles from the proximal segment to the distal segment in a flexion-abduction-rotation sequence. Additionally, joint angles during the initial static N-pose period are regarded as zero, so the offsets during this period are removed [54]. Specifically, the anatomical coordinate systems of segments are established using markers affixed to bony landmarks during static calibration. Concurrently, three-marker clusters are utilized to monitor segment movement, assuming a rigid connection with the clusters [53], [55], [56].

The raw time-series data from each IMU, including three-axis acceleration and three-axis angular velocity, were used as inputs for the data-driven models. The data was transformed into time slices, with each time slice consisting of three hundred time steps and eighteen features. The min-max normalization technique was used to normalize each axis of acceleration and angular velocity and normalize each dimension of the output joint angles.

D. Data Analysis

The performance of the TL method and AAHM was evaluated using the root-mean-square error (RMSE), which is calculated as the difference between the ground truth and the estimated joint angles. We calculated the RMSE for each subject and determined the average RMSE across all subjects. A paired t-test was performed on the RMSE values of the NoTL and TL methods with a significance level of $p = 0.05$. For the classification stage of AAHM, supplementary metrics such as accuracy, sensitivity, specificity, and F1-score were employed for performance evaluation [43].

III. RESULTS

The proposed TL method improved estimation accuracy for swimming, badminton, dance, and golf activities in inter-activity generalization scenarios with limited target training data (Fig. 4 (a)(c)(e)(g)). If the basic network was LSTM, the RMSE accuracy improvements brought by TL ranged from 3.7° to 23.6° in knee flexion/extension and from 3.2° to 22.2° in hip flexion/extension (all $p < 0.0001$). Specifically, when the model generalized from yoga to swimming activity, the TL method significantly outperformed the NoTL method (Fig. 4 (a)). For knee angle estimation, TL achieved RMSE accuracy improvements of 22.2°, 2.4°, and 1.4° in flexion/extension, adduction/abduction, and internal/external rotation, respectively. For the hip joint angle estimation, TL also achieved RMSE accuracy improvements of 23.6°, 3.6°, 10.4° in flexion/extension, adduction/abduction, and internal/external rotation, respectively (Fig. 4 (a)). When the model generalized from yoga to badminton activity, the TL method significantly outperformed the NoTL method. TL resulted in improved RMSE accuracy by 4.3 degrees in knee flexion/extension and 4.9 degrees in hip internal/external rotation (Fig. 4 (c)). When the model generalized from yoga to dance activity, the TL method significantly outperformed the NoTL method, resulting in accuracy improvements of 4.8° in knee joint flexion/extension. For hip angle estimation, TL also brought accuracy improvements of 5.3° in flexion/extension and 2.6° in adduction/abduction (Fig. 4 (e)). When the model generalized from yoga to golf activity, the TL method significantly outperformed the NoTL method, resulting in accuracy improvements of 4.5° in knee joint flexion/extension. For hip angle estimation, TL also brought accuracy improvements of 4.3° in flexion/extension and 1.7° in adduction/abduction (Fig. 4 (g)). If the base network was a transformer, the RMSE accuracy improvements brought by TL ranged from 6.37° to 15.07° in knee flexion/extension and from 5.98° to 22.05° in hip flexion/extension when the model generalized from yoga to other four activities (all $p < 0.0001$) (TABLE III). Compared to the TL method using LSTM as the base network, the minimum accuracy of joint flexion/extension of the TL
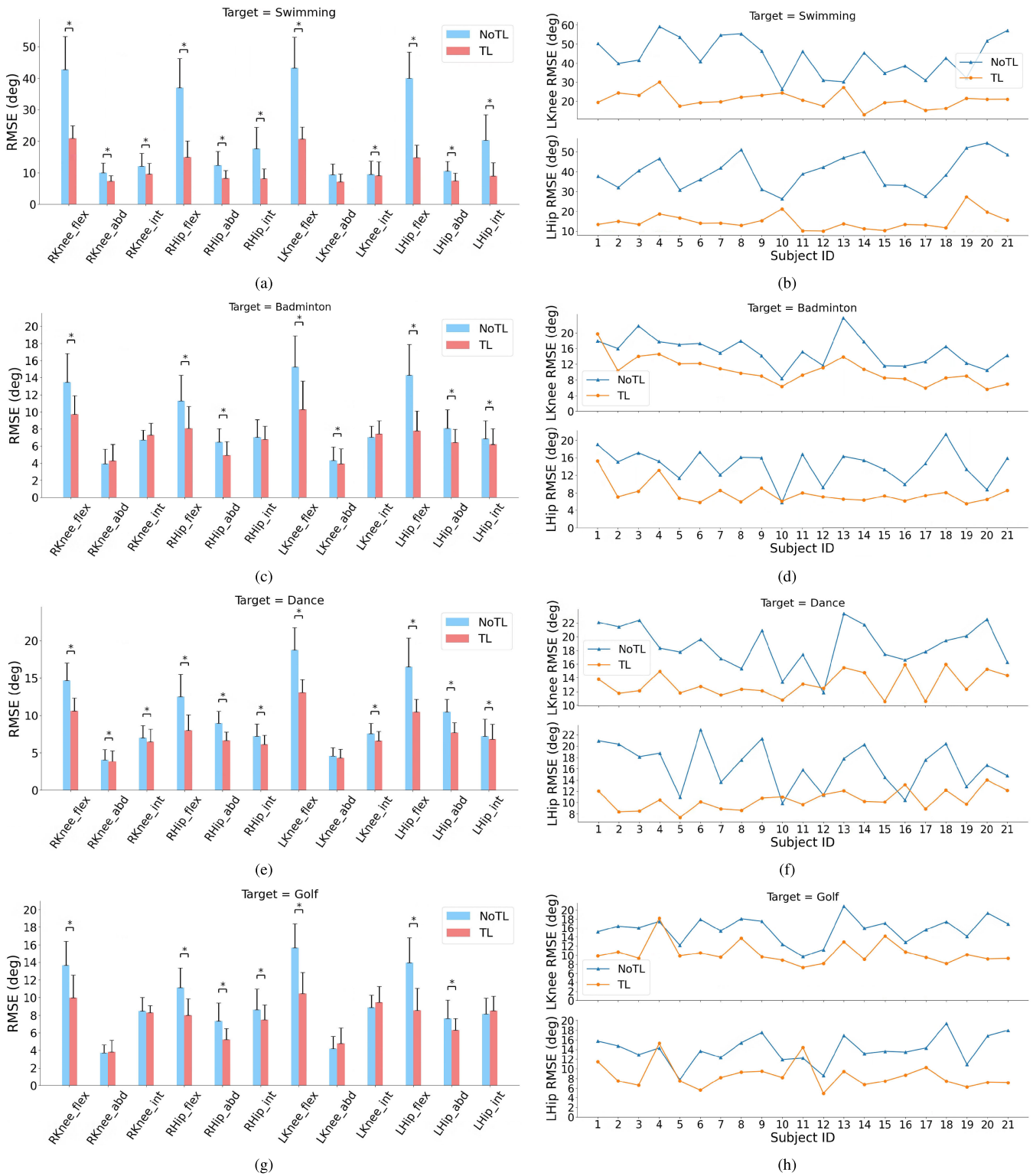
Fig. 4. Accuracy comparison between NoTL and the proposed TL when transferring from the semi-static yoga poses to the four dynamic activities (swimming, badminton, dance, and golf) (basic network was LSTM). (a)(c)(e)(g) The mean RMSEs in each DoF for the methods. (b)(d)(f)(h) The RMSEs in the left knee and hip flexion/extension for individual subjects. The subject-independent TL-based model was robust to individual differences in the dynamic activity domain. NoTL represents directly applying the source model to the target activity (without TL). * denotes a significant difference between the RMSEs of NoTL and TL ($p < 0.001$ for flexion/extension; $p < 0.05$ for other dimensions). R: Right, L: Left, flex: flexion/extension, abd: adduction/abduction, int: internal/external rotation.

method using the transformer network is improved when tested on unseen activities, which suggests that the transformer network can perform better activity generalization after adding the fine-tuning mechanism of the TL method. This may be because the transformer model itself (without TL) is more sensitive to test samples from unseen activities compared to the

LSTM model. Therefore, TL is more beneficial for improving the transformer's accuracy on unseen activities, which also reflects the necessity of TL to build generalized and robust transformer models in the future. We compared the accuracy of the transformer and the LSTM model after applying the TL method, and there was no significant difference between the two models in all 12-dimensional angle values (TABLE I, TABLE III).

The subject-independent TL achieved higher accuracy than NoTL when transferring knowledge from each subject's data of the yoga domain to the swimming domain (Fig. 4 (b)). In addition, subject-independent TL generally achieved higher accuracy than NoTL for badminton, dance, and golf activities, except for a few subjects (Fig. 4 ((d)(f)(h))). This indicates that the subject-independent TL-based model was robust to individual differences for the dynamic activity (Fig. 4 (b)(d)(f)(h)).

TL consistently outperformed NoTL when the source or target domains included swimming and when the source domain was yoga, especially in the flexion/extension dimension (Fig. 4, TABLE I). It is also noteworthy that in transfer pairs such as golf-badminton, badminton-golf, badminton-dance, dance-yoga, dance-golf, and dance-badminton, TL, and NoTL exhibited similar performance due to the similarity in the data distribution of these activities (TABLE I). Golf, badminton, and dance activities could easily be substituted for each other in terms of joint angle estimation (Fig. 5 (b)).

In the task of simultaneous motion pattern recognition and joint angle estimation using a reduced IMU configuration, the average RMSEs of knee joint angle estimation by the AAHM model (the base network is LSTM) were 10.20° ($\pm$1.61°) in flexion/extension, 4.21° ($\pm$1.21°) in adduction/abduction, and 6.34° ($\pm$1.03°) in internal/external rotation across five activities (Fig. 5 (a)). The average RMSEs of hip joint angle estimation were 8.91° ($\pm$1.58°) in flexion/extension, 5.38° ($\pm$1.03°) in adduction/abduction, and 6.24° ($\pm$1.05°) in internal/external rotation, also across five activities. Similarly, the average RMSEs of knee joint angle estimation by the AAHM model (the base network is transformer) were 11.64° ($\pm$1.69°) in flexion/extension, 4.38° ($\pm$1.23°) in adduction/abduction, and 6.92° ($\pm$1.07°) in internal/external rotation across five activities (TABLE II). The average RMSEs of hip joint angle estimation were 9.59° ($\pm$1.58°) in flexion/extension, 5.92° ($\pm$1.03°) in adduction/abduction, and 6.64° ($\pm$1.05°) in internal/external rotation, also across five activities. The estimation accuracy of the transformer model for these five non-gait activities was comparable to or slightly lower than the LSTM model.

The first activity classification stage of AAHM achieved an accuracy of 81.8%, an F1-score of 82%, a sensitivity of 81.7%, and a specificity of 95.4% (Fig. 5 (b)). Note that directly applying the five separate estimation models corresponding to each activity would be equivalent to a virtual AAHM with 100% accuracy of activity classification in the first stage. Surprisingly, there was no significant difference between the real AAHM and virtual AAHM (with perfect activity classification) in all twelve knee and hip angle axes (Fig. 5 (a)). This observation indicates that the first classifica-

tion stage has a negligible effect on the subsequent estimation stage, despite the AAHM achieving a classification accuracy of only 81.8%. Badminton and golf were the most easily confused with each other among the activities, followed by badminton, golf, and dancing, making these three activities prone to confusion (Fig. 5 (b)). Additionally, there was confusion between yoga and badminton/golf/dancing, indicating that certain activities exhibited similar motion patterns (Fig. 5 (b)). However, swimming had the highest classification accuracy among all the activities (Fig. 5 (b)). The degree of similarity between activities reflects the performance of transfer learning (TABLE I).

Based on the comparison between the ground truth and the estimated knee and hip angles for each activity, yoga, swimming, badminton, and dance showed a strong correlation in knee joint flexion/extension compared to other axes (Fig. 6). On the other hand, hip adduction/abduction and hip internal/external rotation were the two dimensions that exhibited lower estimation accuracy across all the activities (Fig. 6), consistent with the results obtained in the previous study [3]. We presented the allowable ranges of ground truth values and estimated values for five activities (TABLE IV), along with the normalized RMSE of joint angle estimates concerning the allowable range of ground truth (TABLE V).

IV. DISCUSSION

For the known activities in the training set, the proposed AAHM with an LSTM neural network achieved satisfactory three-dimensional joint angle estimation accuracy during yoga, golf, swimming, badminton, and dance, using a reduced IMU configuration (Fig. 5 (a)). To further enhance the model's generalization ability to unseen activities, the proposed TL method improved the estimation accuracy of the model trained on the yoga activity when tested on unseen dynamic activities. The maximum improvement in estimation accuracy (RMSE) achieved by TL is 23.6° for knee flexion/extension and 22.2° for hip flexion/extension compared to the NoTL baseline (Fig. 4(a)(c)(e)(g), TABLE I). Notably, these improvements were achieved using only 20% of the original size of the dynamic activity dataset (Section II-B.4).

A. Joint Angle Estimation in Various Activities

We compared the joint angle estimation accuracy with the most related motion capture studies that also evaluated various semi-static and dynamic movements using reduced sensor configuration [6], [8], [17], [57]. However, note that those studies did not specifically focus on validating the accuracy of a particular activity, and the subjects only performed a few repetitions for the specific movements. In our study, the joint angle estimation accuracy for golf, badminton, and dance ranged from 3.83 to 13.07° in all dimensions (Fig. 4 (g)(c)(e), TABLE I). This level of accuracy was higher compared to the RMSEs reported in previous studies [6] (17.54°), [17] (10-15°), and [8] (15.02°). Additionally, our results were consistent with the RMSE reported in the study [57]. The joint angle estimation accuracy for the highly dynamic swimming activity was approximately 7.15-20.71° (Fig. 4 (g)(c)(e),

TABLE I

RMSE COMPARISON BETWEEN TL AND NoTL IN ALL THE SOURCE-TARGET TRANSFER PAIRS (BASIC NETWORK WAS LSTM)

| Source Activity	Target Activity	Method	Right Knee			Right Hip			Left Knee			Left Hip		
			flex	abd	int	flex	abd	int	flex	abd	int	flex	abd	int
Yoga	Golf	NoTL	13.68	3.7	8.47	11.13	7.33	8.61	15.69	4.18	8.86	13.97	7.64	8.14
		TL	9.98*	3.83	8.27	7.99*	5.2*	7.49*	10.42*	4.78	9.46	8.52*	6.29*	8.5
	Swimming	NoTL	42.75	9.93	11.98	36.96	12.27	17.64	43.26	9.31	9.49	39.96	10.48	20.35
		TL	20.9*	7.27*	9.61*	14.88*	8.21*	8.16*	20.71*	7.15	9.02*	14.77*	7.39*	8.99*
	Badminton	NoTL	13.45	3.94	6.75	11.26	6.46	7.03	15.26	4.32	7.04	14.28	8.08	6.9
		TL	9.74*	4.3	7.3	8.07*	4.94*	6.81	10.3*	3.97*	7.45	7.78*	6.42*	6.19*
	Dance	NoTL	14.66	4.07	7.01	12.49	8.95	7.22	18.69	4.59	7.59	16.51	10.48	7.22
		TL	10.6*	3.85*	6.45*	7.98*	6.62*	6.12*	13.07*	4.33	6.6*	10.47*	7.71*	6.8*
Golf	Yoga	NoTL	21.05	3.58	6.69	17.78	9.49	8.77	28.53	5.32	8.64	26.23	12.46	11.66
		TL	14.2*	3.75	6.01	13.38*	9.04	8.75	17.38*	5.95	7.35*	13.8*	8.84*	6.35*
	Swimming	NoTL	48.63	8.92	13.58	38.88	9.3	32.03	49.35	8.13	13.46	37.48	8.66	26.74
		TL	18.56*	7.06*	9.01*	13.57*	7.69*	8.33*	17.97*	7.16*	9.03*	13.45*	7.14*	8.74*
	Badminton	NoTL	8.75	3.64	6.18	7.86	4.27	5.25*	9.2	3.26*	6.69	6.86*	4.97*	5.64
		TL	8.4	3.97	6.53	7.77	4.8	6.16	9.01	3.83	6.95	8.13	6.22	5.59
	Dance	NoTL	10.37	3.93	7.49	8.3	7.65	5.99	12.88	3.98	7.16	11.02	9.15	7.37
		TL	9.31*	3.77	6.06*	7.86	6.95	5.99	11.1*	4.13	6.07*	10.24*	7.29*	6.6
Swimming	Yoga	NoTL	21.06	5.51	6.51	30.52	10.22	14.62	22.98	6.02	6.64	21.42	10.75	10.45
		TL	17.58*	4.06	6.91	16.65*	9.45	8.69*	21.49	6.05	7.84	17.12*	8.21*	5.96*
	Golf	NoTL	25.27	5.35	9.08	26.43	6.57	11.01	25.6	4.74	9.62	26.07	6.1	9.46
		TL	11.6*	3.77*	8.44*	8.12*	5.38*	7.53*	12.16*	4.96	9.31	9.12*	6.63	9.02
	Badminton	NoTL	22.24	5.74	7.32	24.84	5.83	9.54	23.21	4.87	7.61	25.19	5.31*	7.88
		TL	11.58*	4.08*	7.4	8.51*	5.3	6.61*	12.79*	3.91*	7.85	9.53*	6.83	6.16*
	Dance	NoTL	22.8	6.23	7.3	27.1	9.72	10.71	23.03	5.5	7.2	25.25	9.74	9.09
		TL	11.65*	3.98*	6.59	8.53*	8.07*	6.11*	13.83*	4.51*	6.61*	11.27*	8.52*	6.99*
Badminton	Yoga	NoTL	19.13	3.8	6.14	18.54	9.53	9.22	24.15	5.19	5.88	25.88	11.06	8.54
		TL	14.69*	3.8	6.23	14.67*	8.94	8.09	14.72*	6.32	6.65	13.95*	8.26*	5.76*
	Golf	NoTL	7.87	3.3	7.36	7.1	4.89	6.62	8.52	3.33	7.66	7.3*	5.07*	6.79
		TL	8.14	3.67	7.18	7.41	5.03	6.76	8.68	4.81	7.83	8.83	5.66	7.46
	Swimming	NoTL	40.15	9.11	11.09	35.42	9.55	27.12	43.31	8.47	10.12	41.53	10.51	25.71
		TL	18.72*	7.23*	8.9	13.58*	8.43	8.33*	18.45*	7.16*	9.05	13.47*	7.3*	8.87*
	Dance	NoTL	9.05	3.56	5.92	8.75	7.03	5.81	11.17	4.2	6.06	10.58	7.56	6.39
		TL	9.02	3.54	6.21	8.04	6.19*	5.87	11.04	4.19	5.89	10.88	6.7	6.35
Dance	Yoga	NoTL	19.53	3.56	6.43	16.57	8.91	8.43	17.59	5.08	6.17	15.9	8.39	6.78
		TL	14.99*	3.83	6.68	13.6	9.01	8.44	15.06	6.07	6.97	13.41	7.93	6.37
	Golf	NoTL	8.44	3.41	8.2	7.84	4.97	7.03	9.89	3.47	8.16	8.81	5.44	7.08
		TL	8.39	3.86	7.56	7.72	4.8	6.64	9.44	4.6*	8.07	8.97	5.82	7.57
	Swimming	NoTL	45.79	9.38	10.74	40.82	9.45	21.53	45.68	8.93	9.53	44.92	9.88	22.98
		TL	20.38*	7.31*	8.86*	14.81*	8.03*	7.99*	19.56*	7.05*	8.97	14.74*	7.08*	8.86*
	Badminton	NoTL	8.89	3.55	6.21	8.22	4.21	5.6	9.51	3.33	6.43	7.91	4.69	5.72
		TL	8.7	4.15	6.99	8.05	4.37	6.54	8.78	3.88	6.88	7.7	5.86	6.1

The blue color indicates the cases where the proposed TL method outperformed NoTL. TL consistently outperforms NoTL in 14 out of 20 source-target transfer pairs, especially when the source or target domains included swimming and when the source domain was yoga. NoTL represents directly applying the source model to the target activity (without TL). flex: flexion/extension, abd: adduction/abduction, int: internal/external rotation. * denotes a significant difference between the RMSEs of NoTL and TL (all $p < 0.05$).

TABLE I), which aligns with the results reported in the study [6]. Semwal et al. [20] proposed an algorithm for calculating joint angles of the robot based on the 3-link principle. However, a human knee joint cannot be simplified into a one-dimensional hinge or a three-dimensional ball hinge for robots because the knee joint can move in all three dimensions, and the range of motion in the coronal plane is deeply affected by the motion in the sagittal plane. During walking, the flexion of the knee joint changes greatly, and the other two-dimensional angle changes are small. However, during non-walking dynamic activities, the three-dimensional angle changes may be large. Therefore, even if the error of this algorithm is small during walking-type activities, it may not be effective during dynamic activities. We also compared our results with a study that estimated lower-limb 3D joint angles during walking using the same reduced sensor set [3]. Sy et al. [3] reported sagittal knee and hip joint angle RMSEs of $10.0\pm2.9°$ and $9.9\pm3.2°$, respectively, similar to our results.
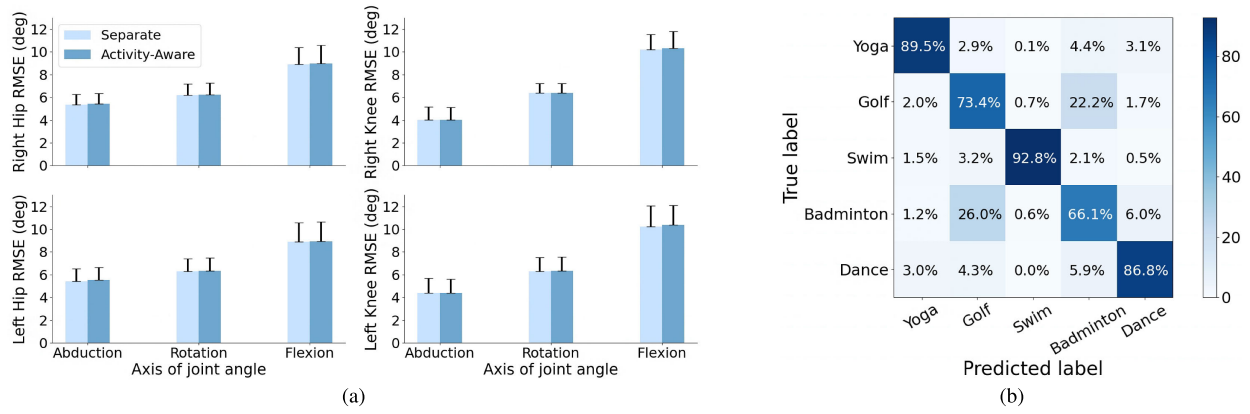
Fig. 5. Simultaneous motion pattern recognition and joint angle estimation using a reduced IMU configuration (basic network was LSTM). (a) Overall accuracy comparison between the activity-aware-based hierarchical model (AAHM) and directly applying the five separate estimation models corresponding to each activity. Note that directly applying the five separate estimation models would be equivalent to a virtual AAHM with 100% accuracy of activity classification in the first stage. There were no significant differences between the AAHM and the five separate estimation models (all less than 0.1°). (b) Confusion matrix for the first classification stage of AAHM.
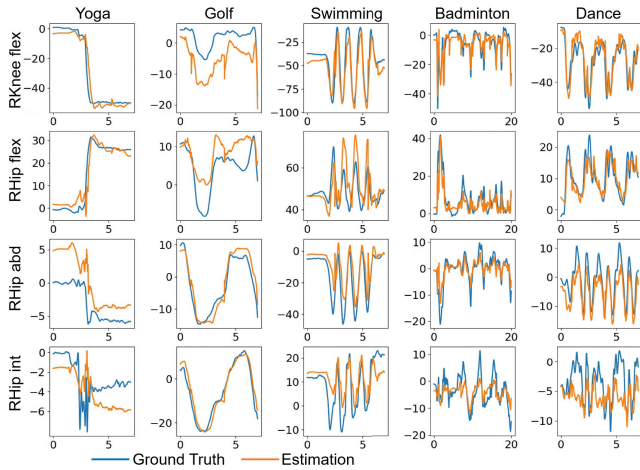
Fig. 6. Comparison of the ground truth and the estimated knee and hip angles corresponding to each activity (basic network was LSTM). X-axis of each subfigure: time (s), y-axis: angle (degrees). R: Right, L: Left, flex: flexion/extension, abd: adduction/abduction, int: internal/external rotation.

TABLE II
RMSEs OF SIMULTANEOUS MOTION PATTERN RECOGNITION AND
JOINT ANGLE ESTIMATION USING A REDUCED IMU CONFIGURATION
WITH DIFFERENT MODELS

Model	Knee			Hip		
	flex	abd	int	flex	abd	int
Separate (LSTM)	10.20 (1.61)	4.21 (1.21)	6.34 (1.03)	8.91 (1.58)	5.38 (1.03)	6.24 (1.06)
Activity-aware (LSTM)	10.35 (1.59)	4.19 (1.18)	6.37 (1.02)	8.96 (1.66)	5.47 (1.01)	6.3 (1.08)
Separate (Transformer)	11.64 (1.69)	4.38 (1.23)	6.92 (1.07)	9.59 (1.56)	5.92 (1.1)	6.64 (0.99)
Activity-aware (Transformer)	11.75 (1.65)	4.37 (1.21)	6.95 (1.07)	9.65 (1.69)	5.96 (1.12)	6.69 (1.03)

Activity-aware: activity-aware-based hierarchical model (AAHM) for five activities. Separate: directly applying the five separate estimation models corresponding to each activity. The RMSE of the left and right legs was averaged because the difference in accuracy was very small. flex: flexion/extension, abd: adduction/abduction, int: internal/external rotation.

However, they observed that during turning movements, the RMSEs for sagittal knee and hip joint angles could reach 15-20°, and the RMSE for the hip joint angle of internal/external rotation could reach 30-45°. The reason for these large errors is that the constraints designed by the algorithm are mainly based on walking movement patterns, which may not be applicable or require additional constraints for movements other than walking, such as turning. Our activities include many dynamic movements, such as golf swings that involve movements similar to turning, so joint angle estimation in dynamic activities may achieve lower accuracy than walking. Their study also reported an error of 15° in hip adduction/abduction at the beginning of the movement, which could be attributed to the model considering previous actions and performing event detection. There is more movement variability in non-periodic dynamic activities such as badminton, which could be one reason for the larger errors observed in dynamic activities. In addition, estimating hip adduction/abduction and hip internal/external rotation angles posed greater challenges than other axes (Fig. 6). These insights can inform future improvements in joint angle estimation models and help focus on addressing the challenges associated with specific joint angle dimensions and activity types.

Swimming exhibited significant differences in movement patterns compared to the other three dynamic activities (golf, badminton, and dance), so swimming was less likely to be confused with the other activities (Fig. 5(b)). In swimming, the subject's feet could not touch the ground, and their legs could not rely on external support. Therefore, the pelvis served as the pivot point to support the legs, allowing for lifting off and completing movements (Section II-C.3). Whereas in other activities, the pivot point is the foot in contact with the ground. Furthermore, other body segments will move around the pivot point. Therefore, the relationship between the movement of the shank IMU and the pelvis IMU in swimming is different from that in other dynamic activities.

TABLE III
RMSE COMPARISON BETWEEN TL AND NOTL IN ALL THE YOGA-DYNAMIC ACTIVITY TRANSFER PAIRS (BASIC NETWORK WAS TRANSFORMER)

Source Activity	Target Activity	Method	Right Knee			Right Hip			Left Knee			Left Hip		
			flex	abd	int	flex	abd	int	flex	abd	int	flex	abd	int
Yoga	Golf	NoTL	17.33	3.81	6.97	13.28	7.74	7.43	18.8	3.89*	7.42	16.74	8.99	7.22
		TL	10.57*	4.19	7.52	8.75*	5.11*	7.22*	11.61*	4.31	7.56	7.93*	6.39*	6.45
	Swimming	NoTL	17.92	4.03	7.06	14.63	10.92	7.83	20.53	4.45	7.08	18.5	10.76	7.14
		TL	11.55*	3.93*	6.58*	8.41*	7.94*	6.1*	14.17*	4.4	6.45	11.74*	8.79*	6.87*
	Badminton	NoTL	19.65	3.6*	8.42	12.71	8.07	8.57	19.82	3.8*	8.82	15.74	8.71	8.18
		TL	10.71*	3.81	8.32	8.02*	5.4*	7.52	11.52*	5	9.07	8.47*	5.97*	8.41
	Dance	NoTL	38.27	9.84	9.98	38.64	9.9	20.2	39.28	8.39	9.37	36.15	10.55	19.46
		TL	23.3*	7.56*	9.01*	15.59*	8.03*	7.97*	24.11*	7.28	9.02*	15.1*	7.18*	9.1*

The blue color indicates the cases where the proposed TL method outperformed NoTL. TL consistently outperforms NoTL in all four yoga-dynamic activity transfer pairs. NoTL represents directly applying the source model to the target activity (without TL). flex: flexion/extension, abd: adduction/abduction, int: internal/external rotation. * denotes a significant difference between the RMSEs of NoTL and TL (all $p < 0.05$).

The transformer excels in handling sequence-to-sequence mapping problems through attention mechanisms, such as translation tasks in natural language processing [50]. However, joint angle estimation with reduced sensor configuration is not a simple mapping problem, it mainly involves the challenge of time series estimation [19], [20], [21], [22], [58]. Furthermore, due to the weak periodicity of the five non-gait dynamic activities investigated in this study, finding the motion patterns of human body segments is more challenging compared to typical time series modeling problems. The reduction in the number of sensors further reduces the observability of the joint angle estimation problem [6]. Some recent studies reported that RNNs (or LSTMs) and MLP-type models may outperform transformers in certain time series tasks [59], [60]. One reason for this is that transformers are less effective in handling time series data compared to vision-related information [61]. Moreover, RNN-type models might still be a viable option if the temporal nature of the data is crucial for the task. Furthermore, we did not explore the sensitivity of the transformer model to longer time windows of data. Using longer historical sequence information tends to gradually enhance the performance of the transformer [33]. However, training the transformer model with a longer sequence reduces its feasibility in potential real-time applications.

In a recent study, Geissinger and Asbeck [17] estimated the orientation of all body segments using a reduced set of sensors (five or six for the full body). The study compared two RNN models and two transformer models. They collected a special test set that included physical exercises with significantly higher accelerations, such as Frankensteins, burpees, push-ups, high jumps, and jogging, as well as stationary periods and some low-acceleration movements. This special test set exhibits non-periodic and unpredictable behavioral patterns that are more similar to the dataset collected in our study. The overall joint angle estimation results showed that the RNN-type model performs best, but the accuracy difference among the four different models is within 2°. No single model outperformed the others in all four reduced-sensor configurations, and the transformer (mainly encoder) outperformed the transformer under some configurations. These findings suggest that in more dynamic and non-periodic activities, models of the RNN class may be slightly superior. However, the differences between transformers and RNNs are small, so it is difficult to conclude which of these models is superior across different datasets and reduced-sensor configurations. Sharifi-Renani et al. [33] employed a transformer that utilizes four IMUs for sagittal angle estimation at the hip, knee, and ankle joints. The transformer architecture comprised encoder layers and fully connected layers. The study reported comparable or slightly superior performance of the transformer and LSTM models. However, it is noteworthy that the problem addressed in this study relates to a normal sensor configuration, not our reduced sensor configuration. Furthermore, the study emphasized caution in applying transformer models to new datasets, as variations in sensor placement and sensor accuracy could adversely affect model estimations [33]. Future improvements in transformer architecture may demonstrate better results in the domain of reduced sensor configurations. Our study did not include all types of activities, and for other activities, the transformer may still be a favorable candidate model.

B. Transfer Learning

When transferring knowledge from the semi-static yoga poses to the swimming activity, TL substantially improved accuracy compared to NoTL (Fig. 4 (a)). However, when the target domain was badminton, dancing, or golf, TL showed a relatively smaller improvement over NoTL (Fig. 4 (c)(e)(g)). It could be attributed to the differences between yoga and the dynamic activities of golf, badminton, and dance being smaller than the differences between yoga and swimming (Fig. 5(b)). Transferring knowledge between highly similar activity domains, such as badminton, golf, and dance, may result in no significant effect or even slightly negative transfer (TABLE I). It suggests that the information contained in the source domain data was already sufficient for estimating samples from the target domain [32].

The NoTL method displayed notable variability in RMSEs across subjects. Specifically, when applying the model trained on the yoga activity to the swimming activity, the RMSE values for some subjects showed a difference of more than 20° (Fig. 4(b)). The error of the NoTL method includes not only individual differences but also large differences in data distribution between activities, such as swimming and yoga activities that are significantly different (Fig. 5(b)).

TABLE IV
GROUND TRUTH AND ESTIMATED ALLOWABLE RANGE (DEGREES) FOR FIVE ACTIVITIES (BASIC NETWORK WAS LSTM)

Activity		Right Knee			Right Hip			Left Knee			Left Hip		
		flex	abd	int	flex	abd	int	flex	abd	int	flex	abd	int
Yoga	Truth	64.14	14.61	26.85	52.34	33.2	32.16	66.51	16.98	30.6	53.37	31.67	29.1
	Estimation	57.56	7.95	15.46	48.86	28.69	23.65	61.92	12.62	19.15	51.86	23.81	17.33
Golf	Truth	38.99	14.19	43.26	37.81	24.33	32.91	37.29	14.79	43.31	36.08	23.09	33.34
	Estimation	38.23	7.72	29.87	29.02	18.55	23.98	35.6	8.51	31.49	25.15	14.47	25.81
Swimming	Truth	106.19	36.9	45.78	86.79	41.69	39.77	106.52	33.59	46.62	87.26	40.86	39.87
	Estimation	85.33	17.77	27.26	48.38	30.18	22.52	85.3	16.76	30.42	47.27	28.09	22.52
Badminton	Truth	59.54	18.9	36.27	49.87	27.16	27.05	58.58	19.19	37.75	44.79	26.73	29.41
	Estimation	49.33	9.71	20.43	38.31	18.03	14.8	46.38	11.15	23.43	24.36	15.87	17.74
Dance	Truth	59.43	21.12	37.97	45.63	45.52	31.3	63.21	23.25	37.96	56.63	45.41	32.49
	Estimation	49.76	12.78	22.63	30.74	37.21	16.57	52.58	12.94	22.97	40.45	38.6	16.64

flex: flexion/extension, abd: adduction/abduction, int: internal/external rotation.

TABLE V
NORMALIZED RMSE (%) OF JOINT ANGLE ESTIMATES CORRESPONDING TO EACH ACTIVITY (BASIC NETWORK WAS LSTM)

	Yoga	Golf	Swimming	Badminton	Dance
RKnee flex	12.53	6.89	11.73	9.26	9.05
RHip flex	12.1	6.29	12.02	7.73	8.68
RHip abd	12.05	8.84	9.31	7.88	6.72
RHip int	12.14	9.48	10.35	10.85	7.15

R: Right, flex: flexion/extension, abd: adduction/abduction, int: internal/external rotation.

Consequently, individual differences in the test set may be amplified by the activity differences, resulting in considerable RMSE variability among specific subjects. This implies that the model may be less robust across subjects if not adapted to new activities. In contrast, our proposed TL method effectively minimizes the data distribution differences between activities. Therefore, its RMSE variability between subjects appears reasonable and significantly smaller than that of the NoTL method (Fig. 4(b)).

Using approximately 20% of the target domain data for transfer is a commonly employed practice. Kang et al. [49] used 20% of the target domain data for transfer and observed that the accuracy reached a stable level when the amount of target domain data involved in the transfer exceeded this threshold. Similarly, Ameri et al. [38] utilized 25% of the target domain data to minimize the length of retraining sessions while ensuring accurate estimation accuracy.

C. Simultaneous Motion Pattern Recognition and Joint Angle Estimation

Despite the lower classification accuracy of the first stage of AAHM, the final estimation accuracy of AAHM remained comparable to directly applying the five separate estimation models corresponding to each activity (Fig. 5(a)). In cases where the classification model misclassified badminton samples as golf activity, which occurred with a probability of 22.2% (Fig. 5(b)), the estimation model for golf was used to test the badminton samples in the second stage of AAHM. The scenario corresponds to the NoTL method, where the source domain is golf, and the target domain is badminton (TABLE I). The accuracy of NoTL did not significantly decrease, remaining below 10° (TABLE I). Therefore, even when there are misclassifications in the first stage, they have a relatively small impact on the final estimation accuracy of the AAHM. On the other hand, the first stage of AAHM demonstrated a high correct classification accuracy of 92.8% for swimming (Fig. 5(b)). It indicates that AAHM rarely confused swimming with other activities due to the significant differences between swimming and other activities. Therefore, samples were effectively prevented from entering the wrong estimation model in the second stage, which helped avoid a severe decrease in accuracy (TABLE I).

D. Future Possibilities

In future work, the proposed TL method can be extended to knowledge transfer between any domains with similarities in data distribution or feature space (Section II-B.2). Therefore, this approach can enhance the model's generalization ability in various dimensions, including different populations, and tolerate IMU placement errors, not limited to different activities alone. The TL method enables fine-tuning of estimation models originally trained on healthy populations for application to osteoarthritis and total arthroplasty populations. This adaptation uses only a small subset of observations from these specific groups, reducing the need for extensive data collection and retraining of models.

Our approach can be further extended to include more dynamic activities such as football and skiing in the future. This expansion holds the potential to make wearable injury prevention and sports training systems more widespread and economically viable. Moreover, the TL method enables the model to acquire the ability to estimate joint angles in dynamic activities using existing semi-static yoga data sets. This will contribute to the development of joint angle estimation models for more dynamic activities. Importantly, collecting data from semi-static activities is easier than collecting data sets from dynamic activities, thus significantly saving data collection efforts.

Future work should seek to extend these findings to further investigate the effectiveness of our proposed method

in estimating upper body joint angles with reduced sensor configurations. This method has the potential to integrate dependencies between the upper and lower bodies, allowing for the joint angle estimation of the upper body during activities involving the entire body, particularly in situations such as golf and swimming. The currently proposed activity-aware-based hierarchical model can only recognize known activities, though future work should focus on exploring additional techniques such as anomaly detection to achieve activity recognition for unknown activities.

E. Limitations

One limitation of this study is that the swimming data was collected in a simulated seated position rather than in actual water. In real swimming situations, the forces on the joints of the whole body are significantly different compared to common activities on land. Therefore, although we instructed subjects to simulate swimming movements by driving their legs as if lifting off in water without attaching any object, the absence of water resistance and other aquatic factors might have somewhat affected the dynamics and joint angles.

V. Conclusion

This study performed accurate 3D knee and hip angle estimations using a reduced IMU sensor set across five activities, including yoga, golf, swimming, badminton, and dance. This validates the feasibility of the reduced sensor configuration beyond gait-type activities. Simultaneous motion pattern recognition and joint angle estimation revealed remarkable similarities in models for activities prone to confusion. This suggests a practical strategy: when developing estimation models for new activities with data distributions akin to known activities, the datasets of known activities can be directly employed, thus substantially reducing the need for extensive data collection. Additionally, we proposed a TL-based approach to estimate joint angles in unseen activities with limited training data, thereby enhancing the model's generalization ability across diverse activities. This is particularly beneficial for activities where the data distribution differs significantly from known data sets. Leveraging the known semi-static yoga poses dataset, our proposed approach significantly improved joint angle estimation in dynamic activities compared to without TL, utilizing only 20% of the original dataset size. These findings not only extend the applicability of motion capture with reduced sensor configurations to a wider spectrum of activities relevant to injury prevention and sports training but also offer a concrete TL-based solution for augmenting the model's adaptability to unseen activities using minimal training data.

References

[1] M. Mundt, A. Koeppe, F. Bamer, S. David, and B. Markert, "Artificial neural networks in motion analysis—Applications of unsupervised and heuristic feature selection techniques," *Sensors*, vol. 20, no. 16, p. 4581, Aug. 2020.

[2] M. S. B. Hossain, H. Choi, and Z. Guo, "Estimating lower extremity joint angles during gait using reduced number of sensors count via deep learning," in *Proc. 14th Int. Conf. Digit. Image Process. (ICDIP)*, Oct. 2022, pp. 1116–1123.

[3] L. Sy et al., "Estimating lower limb kinematics using a reduced wearable sensor count," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 4, pp. 1293–1304, Apr. 2021.

[4] J. S. Pedley et al., "Utility of kinetic and kinematic jumping and landing variables as predictors of injury risk: A systematic review," *J. Sci. Sport Exerc.*, vol. 2, no. 4, pp. 287–304, Nov. 2020.

[5] G. Hughes, "A review of recent perspectives on biomechanical risk factors associated with anterior cruciate ligament injury," *Res. Sports Med.*, vol. 22, no. 2, pp. 193–212, Apr. 2014.

[6] Y. Huang, M. Kaufmann, E. Aksan, M. J. Black, O. Hilliges, and G. Pons-Moll, "Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time," *ACM Trans. Graph.*, vol. 37, no. 6, pp. 1–15, Dec. 2018.

[7] X. Yi, Y. Zhou, and F. Xu, "TransPose: Real-time 3D human translation and pose estimation with six inertial sensors," *ACM Trans. Graph.*, vol. 40, no. 4, pp. 1–13, Aug. 2021.

[8] X. Yi et al., "Physical inertial poser (PIP): Physics-aware real-time human motion tracking from sparse inertial sensors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 13157–13168.

[9] Z. Zheng, H. Ma, W. Yan, H. Liu, and Z. Yang, "Training data selection and optimal sensor placement for deep-learning-based sparse inertial sensor human posture reconstruction," *Entropy*, vol. 23, no. 5, p. 588, May 2021.

[10] F. J. Wouda et al., "Estimation of vertical ground reaction forces and sagittal knee kinematics during running using three inertial sensors," *Frontiers Physiol.*, vol. 9, p. 218, Mar. 2018.

[11] V. Hernandez, D. Dadkhah, V. Babakeshizadeh, and D. Kulic, "Lower body kinematics estimation from wearable sensors for walking and running: A deep learning approach," *Gait Posture*, vol. 83, pp. 185–193, Jan. 2021.

[12] A. Karatsidis, G. Bellusci, H. Schepers, M. de Zee, M. Andersen, and P. Veltink, "Estimation of ground reaction forces and moments during gait using only inertial motion capture," *Sensors*, vol. 17, no. 12, p. 75, Dec. 2016.

[13] J.-S. Tan et al., "Predicting knee joint kinematics from wearable sensor data in people with knee osteoarthritis and clinical considerations for future machine learning models," *Sensors*, vol. 22, no. 2, p. 446, Jan. 2022.

[14] T. von Marcard, B. Rosenhahn, M. J. Black, and G. Pons-Moll, "Sparse inertial poser: Automatic 3D human pose estimation from sparse IMUs," *Comput. Graph. Forum*, vol. 36, no. 2, pp. 349–360, May 2017.

[15] D. R. Seshadri et al., "Wearable sensors for monitoring the internal and external workload of the athlete," *npj Digit. Med.*, vol. 2, no. 1, p. 71, Jul. 2019.

[16] M. Mundt, W. R. Johnson, W. Potthast, B. Markert, A. Mian, and J. Alderson, "A comparison of three neural network approaches for estimating joint angles and moments from inertial measurement units," *Sensors*, vol. 21, no. 13, p. 4535, Jul. 2021.

[17] J. H. Geissinger and A. T. Asbeck, "Motion inference using sparse inertial sensors, self-supervised learning, and a new dataset of unscripted human motion," *Sensors*, vol. 20, no. 21, p. 6330, Nov. 2020.

[18] S. K. Challa, A. Kumar, V. B. Semwal, and N. Dua, "An optimized-LSTM and RGB-D sensor-based human gait trajectory generator for bipedal robot walking," *IEEE Sensors J.*, vol. 22, no. 24, pp. 24352–24363, Dec. 2022.

[19] V. B. Semwal et al., "Development of the LSTM model and universal polynomial equation for all the sub-phases of human gait," *IEEE Sensors J.*, vol. 23, no. 14, pp. 15892–15900, Jul. 2023.

[20] V. B. Semwal, N. Gaud, P. Lalwani, V. Bijalwan, and A. K. Alok, "Pattern identification of different human joints for different human walking styles using inertial measurement unit (IMU) sensor," *Artif. Intell. Rev.*, vol. 55, no. 2, pp. 1149–1169, Feb. 2022.

[21] V. B. Semwal, P. Tokas, M. Nilesh, and N. Khare, "Dynamic balancing of bipedal robot on inclined terrain using polynomial trajectories," in *Proc. 7th Int. Conf. Graph. Signal Process.*, Jun. 2023, pp. 72–77.

[22] V. B. Semwal, R. Jain, P. Maheshwari, and S. Khatwani, "Gait reference trajectory generation at different walking speeds using LSTM and CNN," *Multimedia Tools Appl.*, vol. 82, no. 21, pp. 33401–33419, Sep. 2023.

[23] V. B. Semwal, "Data driven computational model for bipedal walking and push recovery," 2017, *arXiv:1710.06548*.

[24] M. Raj, V. B. Semwal, and G. C. Nandi, "Bidirectional association of joint angle trajectories for humanoid locomotion: The restricted Boltzmann machine approach," *Neural Comput. Appl.*, vol. 30, no. 6, pp. 1747–1755, Sep. 2018.

[25] M. L. Baker, D. R. Epari, S. Lorenzetti, M. Sayers, U. Boutellier, and W. R. Taylor, "Risk factors for knee injury in golf: A systematic review," *Sports Med.*, vol. 47, no. 12, pp. 2621–2639, Dec. 2017.

[26] K. S. Rudolph, M. J. Axe, and L. Snyder-Mackler, "Dynamic stability after ACL injury: Who can hop?" *Knee Surg., Sports Traumatol., Arthroscopy*, vol. 8, no. 5, pp. 262–269, Sep. 2000.

[27] M. R. Hutchinson and M. L. Ireland, "Knee injuries in female athletes," *Sports Med.*, vol. 19, no. 4, pp. 288–302, Apr. 1995.

[28] M. Bourgain, P. Rouch, O. Rouillon, P. Thoreux, and C. Sauret, "Golf swing biomechanics: A systematic review and methodological recommendations for kinematics," *Sports*, vol. 10, no. 6, p. 91, Jun. 2022.

[29] L. Fu, F. Ren, and J. S. Baker, "Comparison of joint loading in badminton lunging between professional and amateur badminton players," *Appl. Bionics Biomechanics*, vol. 2017, pp. 1–8, Jun. 2017.

[30] K. K. Yamakawa, H. Takagi, and Y. Sengoku, "Three-dimensional analysis of hip and knee joint movements during dolphin kicking and butterfly swimming," in *Proc. 13th Biomechanics Med. Swimming. Tokyo, Impress RD*, 2018, pp. 92–187.

[31] E. Whissell, L. Wang, P. Li, J. X. Li, and Z. Wei, "Biomechanical characteristics on the lower extremity of three typical yoga manoeuvres," *Appl. Bionics Biomechanics*, vol. 2021, pp. 1–7, Aug. 2021.

[32] S. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, pp. 1345–1359, Nov. 2010.

[33] M. Sharifi-Renani, M. H. Mahoor, and C. W. Clary, "BioMAT: An open-source biomechanics multi-activity transformer for joint kinematic predictions using wearable sensors," *Sensors*, vol. 23, no. 13, p. 5778, Jun. 2023.

[34] L. Zhang, D. Soselia, R. Wang, and E. M. Gutierrez-Farewik, "Lower-limb joint torque prediction using LSTM neural networks and transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 600–609, 2022.

[35] J. Wang, V. W. Zheng, Y. Chen, and M. Huang, "Deep transfer learning for cross-domain activity recognition," in *Proc. 3rd Int. Conf. Crowd Sci. Eng.*, 2018, pp. 1–8.

[36] M. Sharifi Renani, A. M. Eustace, C. A. Myers, and C. W. Clary, "The use of synthetic IMU signals in the training of deep learning models significantly improves the accuracy of joint kinematic predictions," *Sensors*, vol. 21, no. 17, p. 5876, Aug. 2021.

[37] J. Li, P. Kang, T. Tan, and P. B. Shull, "Transfer learning improves accelerometer-based child activity recognition via subject-independent adult-domain adaption," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 2086–2095, May 2022.

[38] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "A deep transfer learning approach to reducing the effect of electrode shift in EMG pattern recognition-based control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 370–379, Feb. 2020.

[39] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[40] J. Wang, Y. Chen, L. Hu, X. Peng, and P. S. Yu, "Stratified transfer learning for cross-domain activity recognition," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2018, pp. 1–10.

[41] D. Cook, K. D. Feuz, and N. C. Krishnan, "Transfer learning for activity recognition: A survey," *Knowl. Inf. Syst.*, vol. 36, no. 3, pp. 537–556, Sep. 2013.

[42] A. Gautam, M. Panwar, D. Biswas, and A. Acharyya, "MyoNet: A transfer-learning-based LRCN for lower limb movement recognition and knee joint angle prediction for remote monitoring of rehabilitation progress from sEMG," *IEEE J. Transl. Eng. Health Med.*, vol. 8, pp. 1–10, 2020.

[43] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, no. 12, pp. 31314–31338, Dec. 2015.

[44] M. Janidarmian, A. Roshan Fekr, K. Radecka, and Z. Zilic, "A comprehensive analysis on wearable acceleration sensors in human activity recognition," *Sensors*, vol. 17, no. 3, p. 529, Mar. 2017.

[45] Y. Ahn, D. Jayalath, and A. Oloyede, "A framework for modularised wearable adaptive biofeedback devices," in *Proc. IEEE 18th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Sep. 2016, pp. 1–6.

[46] N. Nithya and G. Nallavan, "Role of wearables in sports based on activity recognition and biometric parameters: A survey," in *Proc. Int. Conf. Artif. Intell. Smart Syst. (ICAIS)*, Mar. 2021, pp. 1700–1705.

[47] T. Sztyler, H. Stuckenschmidt, and W. Petrich, "Position-aware activity recognition with wearable devices," *Pervas. Mobile Comput.*, vol. 38, pp. 281–295, Jul. 2017.

[48] Y. Geng, O. W. Samuel, Y. Wei, and G. Li, "Improving the robustness of real-time myoelectric pattern recognition against arm position changes in transradial amputees," *BioMed Res. Int.*, vol. 2017, pp. 1–10, Jan. 2017.

[49] P. Kang, J. Li, B. Fan, S. Jiang, and P. B. Shull, "Wrist-worn hand gesture recognition while walking via transfer learning," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 3, pp. 952–961, Mar. 2022.

[50] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[51] H. Xia, J. Xu, J. Wang, M. A. Hunt, and P. B. Shull, "Validation of a smart shoe for estimating foot progression angle during walking gait," *J. Biomechanics*, vol. 61, pp. 193–198, Aug. 2017.

[52] M. Nazarahari, A. Noamani, N. Ahmadian, and H. Rouhani, "Sensor-to-body calibration procedure for clinical motion analysis of lower limb using magnetic and inertial measurement units," *J. Biomechanics*, vol. 85, pp. 224–229, Mar. 2019.

[53] A. Cappozzo, F. Catani, U. D. Croce, and A. Leardini, "Position and orientation in space of bones during movement: Anatomical frame definition and determination," *Clin. Biomechanics*, vol. 10, no. 4, pp. 171–178, Jun. 1995.

[54] P. M. Mills, S. Morrison, D. G. Lloyd, and R. S. Barrett, "Repeatability of 3D gait kinematics obtained from an electromagnetic tracking system during treadmill locomotion," *J. Biomechanics*, vol. 40, no. 7, pp. 1504–1511, Jan. 2007.

[55] M. Donati, V. Camomilla, G. Vannozzi, and A. Cappozzo, "Enhanced anatomical calibration in human movement analysis," *Gait Posture*, vol. 26, no. 2, pp. 179–185, Jul. 2007.

[56] S. Ceccon et al., "Motion analysis of front crawl swimming applying CAST technique by means of automatic tracking," *J. Sports Sci.*, vol. 31, no. 3, pp. 276–287, Feb. 2013.

[57] F. Wouda, M. Giuberti, G. Bellusci, and P. Veltink, "Estimation of full-body poses using only five inertial sensors: An eager or lazy learning approach?" *Sensors*, vol. 16, no. 12, p. 2138, Dec. 2016.

[58] V. B. Semwal and G. C. Nandi, "Generation of joint trajectories using hybrid automate-based model: A rocking block-based approach," *IEEE Sensors J.*, vol. 16, no. 14, pp. 5805–5816, Jul. 2016.

[59] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 9, pp. 11121–11128.

[60] A. Das, W. Kong, A. Leach, R. Sen, and R. Yu, "Long-term forecasting with tide: Time-series dense encoder," 2023, *arXiv:2304.08424*.

[61] L. Yang, X. Fan, and Z. Zhang, "Your time series is worth a binary image: Machine vision assisted deep framework for time series forecasting," 2023, *arXiv:2302.14390*.