Check for updates

Latest updates: https://dl.acm.org/doi/10.1145/3788290

PDF Download
3788290.pdf
11 March 2026
Total Citations: 0
Total Downloads: 105

Published: 19 January 2026
Accepted: 03 January 2026
Revised: 06 December 2025
Received: 17 August 2025

Citation in BibTeX format

RESEARCH-ARTICLE

# SkillNet: Human Actions Assessment via Human-AI Collaboration

**INDRAJEET GHOSH**, University of Virginia, Charlottesville, VA, United States

**AVIJOY CHAKMA**, Bowie State University, Bowie, MD, United States

**MOHAMMAD SAEID ANWAR**, University of Maryland, Baltimore County (UMBC), Baltimore, MD, United States

**SREENIVASAN RAMASAMY RAMAMURTHY**, Bowie State University, Bowie, MD, United States

**NIRMALYA ROY**, University of Maryland, Baltimore County (UMBC), Baltimore, MD, United States

**Open Access Support** provided by:

**University of Maryland, Baltimore County (UMBC)**

**Bowie State University**

**University of Virginia**

ACM Transactions on Multimedia Computing, Communications and Applications
https://doi.org/10.1145/3788290
EISSN: 1551-6865

# SkillNet: Human Actions Assessment via Human-AI Collaboration

INDRAJEET GHOSH, Department of Systems and Information Engineering, University of Virginia, USA
AVIJOY CHAKMA, Department of Computer Science, Bowie State University, USA
MOHAMMAD SAEID ANWAR, Department of Information Systems, University of Maryland Baltimore County, USA
SREENIVASAN RAMASAMY RAMAMURTHY, Department of Computer Science, Bowie State University, USA
NIRMALYA ROY, Department of Information Systems, University of Maryland Baltimore County, USA

Intelligent human motion analysis is essential for developing next-generation IoT and AR/VR systems that enable automated, interpretable, and fine-grained performance assessment. Motivated by the need for real-time, explainable, and transferable skill evaluation, we propose a wearable sensing framework to assess human performance by tracking skill progression and minimizing injury risk. We use live *badminton gameplay* and *workout exercises* as representative use cases, where motion dynamics, postural stability, and limb coordination are critical to success. Both activities demand optimal posture and synchronized limb movements, while improper actions or suboptimal technique can lead to decreased performance and higher injury susceptibility. We introduce ***SkillNet***, a multi-task learning framework that extracts shared representations across all limbs while preserving limb-specific motion signatures. The architecture employs task-specific regressors to detect subtle inter-limb dissimilarities and distinctive traits, enabling collective inference in a body sensor network (BSN) environment. To holistically measure performance, we formulated a weighted performance indicator (PI) that fuses AI-driven scoring with domain-expert evaluations, providing a robust metric for both qualitative and quantitative assessment. We evaluate ***SkillNet*** on three diverse datasets Badminton Activity Recognition (BAR), Multi-Modalities Dataset of Sports (MMDOS), and Daily and Sports Activities (DSADS) capturing a broad spectrum of motion types and skill intensities. Results show that ***SkillNet*** achieves an $R^2$ score of ***86%*** and a mean squared error of ***0.0093*** in performance prediction. The integrated AI–expert scoring mechanism improves baseline performance estimation by ***14.95%***, demonstrating the advantage of combining human expertise with automated analysis. We further benchmark inference time, memory usage, and power consumption of the ***SkillNet***, validating its efficiency and feasibility for real-time, end-to-end task inference on resource-constrained embedded edge devices, Jetson Nano and Jetson Xavier NX platforms.

CCS Concepts: • **Computing methodologies → Multi-task learning**; • **Human-centered computing** → *Ubiquitous and mobile computing*; • **Computer systems organization** → Sensor networks.

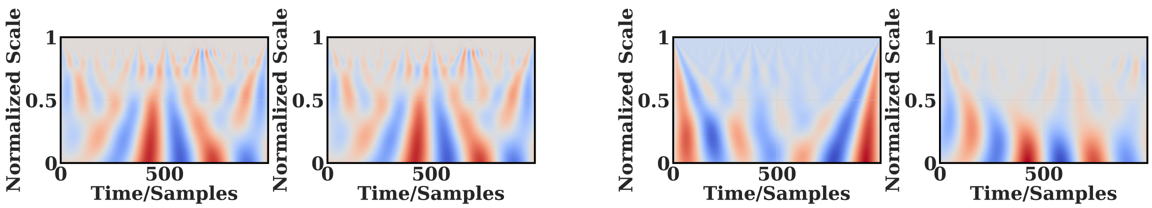
Additional Key Words and Phrases: Sports analytics, Action Recognition, Error Estimation, Performance Indicator, Human-in-the-loop, Human-AI Interaction, Skill Assessment

Authors' Contact Information: Indrajeet Ghosh, Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA, USA, vkw4ze@virginia.edu; Avijoy Chakma, Department of Computer Science, Bowie State University, Bowie, MD, USA, achakma@bowiestate.edu; Mohammad Saeid Anwar, Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD, USA, saeid.anwar@umbc.edu; Sreenivasan Ramasamy Ramamurthy, Department of Computer Science, Bowie State University, Bowie, MD, USA, sramamurthy@bowiestate.edu; Nirmalya Roy, Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD, USA, nroy@umbc.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM 1551-6865/2026/1-ART
https://doi.org/10.1145/3788290

Just Accepted




**(a)** Professional Players

**(b)** Novice Players

**Fig. 1.** (a, b) Magnitude scalograms of dominant limbs for *Smash Overhead Forehand Shot*

## 1 Introduction

Wearable devices are becoming an integral part of an individual to study and assess actions and behavior in their daily routines. Often, the devices are equipped with multiple sensors that collectively provide a vast pool of information, enabling the end-user to discover fine-grained action patterns that the human eye barely observes. This potential has resulted in a surge in the development of applications in smart homes [8, 33, 58], gait analysis [23, 38, 61], and sports analytics [42, 44, 56], healthcare [15, 16, 41, 54], sleep monitoring [24, 26, 52], etc. Sports analytics has seen extensive commercial adaptation of smart wearables during team practice sessions [30], benefiting the coach, player, and the team by providing player statistics, profiling, and supporting performance enhancement [50]. However, most commercial devices are more limited in statistical data, rather than yielding an insightful exploratory analysis, which provides a clearer picture of a player's performance. We aim to fill this gap by developing a performance assessment framework that helps to understand an individual's activity performance when motor execution involves coordinated limb movements.

The quality of an executed activity intuitively refers to an individual's continued dynamic body movements and is a crucial aspect to measure an individual's skill level. Continuous display of superior skillset contributes towards an individual's success such as in acrobatic and workout exercises. Consider, ***badminton*** a racquet sport, part of the Olympic games often played in singles and doubles format. Badminton requires performing synchronous limbs (hands, legs) maneuvering activities in the ground and air. Similarly, workout exercises, such as squats, shoulder presses, etc., need personalized assistance [22] from professional trainers or coaches to perform the exercises efficiently without muscle strains or injuries. Irrespective of the exercise type, effective and efficient maneuvering skill requires hours of training and coaching. However, continuous access to the coaching and training staff is infeasible. Therefore, it is of utmost importance to develop a framework that can measure the performance of an activity.

Moreover, we discover that the subtle difference in agility and skill can be captured and assessed to provide valuable feedback to comparatively less skilled individuals (trainee/novice). Motivated by this, we conducted a preliminary study to show that each limb is unique and has distinctive dynamics. It is also essential to understand the dynamics of those movements, and estimating such distinctive traits will enhance the individual's action performance irrespective of factors (height, dominant limb, weight, etc.). To support our proposition, the preliminary study shown in Figures 1a and 1b demonstrate the magnitude of scalograms corresponding to the ***Smash overhead forehand shots*** from the professional and novice players, respectively. Magnitude values were calculated from the 3-axis accelerometer, magnetometer and gyroscope values. The scalogram suggests that professional players perform a defined repetitive pattern that might contribute to low variations and sparse data points. Furthermore, the scalogram from the professional players in Figure 1a also reveals a steady signal power at the same frequency band throughout the time segment to complement this inference. Whereas, in *Figure* 1b, such definitive repetitive patterns are not visible in the case of the novice players. Fewer repetitive patterns suggest a

Just Accepted

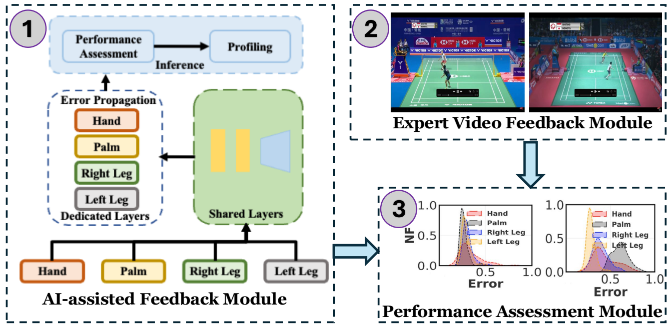



**Fig. 2.** High overview of **SkillNet** framework comprising three key modules: (i) *AI-assisted personalized feedback*, (ii) *expert video-based feedback* and (iii) *performance assessment (where NF denotes the normalized frequency).*

higher error occurrence in the novice player than the professional player. Figures 1a and 1b suggests that the IMU data can capture the differences between the professional players' and the novice players' proficiency.

Figure 2 presents an overview of the ***SkillNet*** framework, developed to assess and enhance the quality of limb maneuvering. The framework consists of three core components: (i) an *AI-assisted module* that leverages deep learning to detect limb-specific errors from sensor data and incorporates error propagation analysis to characterize intra-limb variance and model distinctive player traits for personalized skill profiling; (ii) an *expert video-based feedback module* that provides clip-based corrective guidance using the BadmintonDB dataset [4], focusing on the most error-prone limb identified by the AI module; and (iii) a *performance assessment module* that quantifies improvement using task-relevant performance metrics, validating the impact of feedback integration. The inference phase generates a comprehensive performance summary to ensure real-world feasibility and profiles the utilization of computational resources of the system. This profiling is essential to benchmark SkillNet's efficiency against state-of-the-art (SOTA) approaches.

***Key contributions*** of the paper are summarized as follows:

- **Generalized Multi-task Learning Framework**: We propose a novel approach to learn both generalized and unique features by employing a multi-task learning pipeline. Shared layers are used to extract common representations from each limb of the players, followed by task-specific regression layers that preserve limb-specific differences. To foster reproducibility and encourage engagement from the research community, we have open-sourced **SkillNet**[1].
- **Algorithm Evaluation**- We evaluate the *SkillNet* framework using three publicly available Badminton Activity Recognition (BAR), Daily and Sports Activities (DSADS) and Multi Modalities Dataset of Sports (MMDOS) datasets comprises of activities of daily living (ADLs), instrumental activities of daily living (iADLs), workout and sports activities. Next, we introduce a ***cross-person validation strategy*** to assess the scalability of the *SkillNet* framework. We observe that *SkillNet* achieves $R^2$ Score of $\approx$ ***86%*** for predicting the players' performance performance improvement via human and AI collaborative feedback. Finally, we devise a weighted ***performance indicator (PI)*** to quantify players' performance improvement by integrating expert-based video analysis with AI collaborative feedback, demonstrating performance improvement ranging from ***1.90%*** to ***14.95%*** in players' skill progression.

[1]https://github.com/indrajeetghosh/SkillNet-Framework

Just Accepted

- **System Evaluation**- We propose ***layer-wise time complexity*** profiling of the ***multi-task learning*** pipeline to demonstrate the reliability and robustness for real-time inference. We quantify the hardware resource consumption (*time*) of the PyTorch operations for CPU and GPU processors, respectively and benchmark the *SkillNet* framework on resource-constrained edge devices: ***Jetson Nano*** and ***Jetson NX Xavier*** with *81.80* seconds of total inference time with *45.8%* of total memory and *6.42 W* total power consumed during the experiment.

## 2 Related Work

This paper builds on summarizing the previous work applied methodologies into three categories: *(i)* wearable sensors in sports analytics, activities of daily living (ADLs) and instrumental activities of daily living (iADLs), *(ii)* multi-task learning for activity recognition and *(iii)* performance assessment. We mainly focus on the aspect of our approach that differs from SOTA methods.

***Wearables sensors*** are making swift progress in developing various applications and research works in different sports such as tennis [43], soccer [31], cricket [19, 20], badminton [47, 55], swimming [7, 35], skiing [6], etc. This literature study mainly discusses the research works focusing on Badminton as discussing other works is beyond this paper's scope. In [51], the authors attempted to classify nine badminton shots and the associated body movements by leveraging Convolution Neural Networks (CNNs). The authors placed sensors on three body positions (the wrist, the bottom of the racket's grip and the upper arm) and collected badminton data from two right-handed players during the data collection. However, the work shortfall focuses on recognizing a limited number of badminton strokes using only upper limb movements data.

In contrast, the lower limb movement data were not considered in their study. In [2], the authors developed a classification model to detect five different badminton shots. The dataset was collected by fixing a sensor (Magnetic Pickup Unit 6050) at the racket head, which is very impractical and substantially affects the shot quality. Prior additional research works mainly focused on analyzing the smash shot [28, 45], where the authors proposed approaches to measure the acceleration and the movement of the upper and lower arms of the player. All these literature works focus on classifying and recognizing various shots, but the feedback mechanism for beginner-level players is still missing. We aim to extend the existing research direction from the classification task to develop a feedback-producing mechanism, enabling any entry-level player to be informed of errors, make corrections, and improve their performance. Additionally, wearable sensors are not limited to the sports analytics domain but have expanded from kitchen tasks; *ActionSense* [12], daily living activities [36], smart home health [48, 60], etc. and are still expanding.

***Multi-task learning*** [9] is an inductive transfer learning methodology that accomplishes to increases the generalization performance of a proposed approach by exploiting the commonalities and differences among various tasks. In [5] leverages a multi-task long short-term memory (LSTMs) model to accomplish two tasks - accurately classify the activity types and estimate the intensity of each activity. Both tasks play a substantial role in healthcare applications, such as fitness tracking and patient monitoring. Furthermore, in [63], the authors investigated recognizing collective ADLs and sports activities using a single model. The authors experimented with several state-of-the-art approaches and found that multi-task variants result in increased performance. Further, the authors collected the C-Sports dataset, which comprised eleven different sports with five different activities and used the evaluation dataset. METIER [10] proposed a deep multi-task learning approach to perform two tasks jointly - activity and user recognition. The proposed method focuses on alleviating the data distribution heterogeneity caused due to the user (kinetic variation).

In RubyBot [25], the authors adapted a multi-task learning approach for making different Rugby game predictions such as outcome, possession and win probability of the game. [63], the authors investigated on recognizing collective activity and sports activity using a single model. The authors experimented with several

Just Accepted

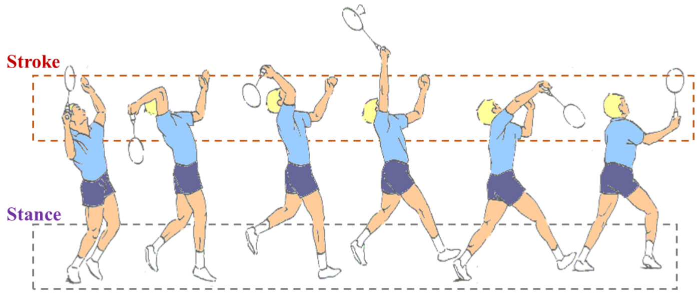



**Fig. 3.** Figure demonstrates the importance of coordinating the player's upper (stroke) and lower (stance) limbs to play a successful shot shown in [17].

SOTA approaches and found that multi-task variants result in increased performance. Further, the authors collected C-Sports dataset, which comprised eleven different sports with five different activities for each sport and used the evaluation dataset. In our proposed work, we aim to borrow the benefit of using multi-task learning to model to formulate an end to end data-driven pipeline to estimate the difference between the distinctive traits of the players' limbs.

***Action Quality and Performance Assessment*** is an essential aspect of activity detection for long-term monitoring and training purposes. Recent surveys [62, 66] emphasize that action quality assessment (AQA) quantitatively evaluates the quality of human actions across domains such as sports, skill learning, and medical care, enabling standardized and unbiased performance evaluation. In the broader literature, several studies have focused on assessing human performance and physiological well-being, including sleep quality [34], spinal movement [40], walking stability in Parkinson's disease [27], AI-generated video AQA [11], and multimodal AQA [64]. In [57], the authors propose Uni-FineParser, a human-centric fine-grained Action Quality Assessment framework that parses actions across semantic, spatial, and temporal dimensions for evaluating diverse sports actions such as diving, gym vault, and big air skiing. However, such assessments are often carried out by coaches or computer-vision-based approaches that rely on prior experience, making these judgments prone to subjectivity and limited in temporal understanding. Only a few studies have explored wearable-based action quality assessment, where body-worn sensors are used to evaluate execution quality and provide quantitative feedback for performance improvement.

In [27] studied the available literature on the wearable usage in parkinson's disease to understand the best sensor types, sensor locations and outcomes to assess standing balance and walking stability parkinson's patients. In [40], studied the usability of wearable devices in tracking the spine movement, which could provide substantial insight into the low back pain. SmartCuff [13] introduced a wearable real-time platform (namely, Smart-Cuff) for quantifying peripheral edema (Leg swelling produced by retention of fluid in leg tissues known as peripheral edema) continuously and accurately. In an extension of wearable usage, AAC [46] considered the wearables that perform mutually to identify not only activities but also to evaluate them qualitatively using the data of several sensor nodes attached to the body. It also provide detailed feedback for the improvement of the execution. AAC focuses on the online assessment of periodic human activity within a wireless body area network.

To the best of our knowledge, not too many literary works have concentrated on the players' performance assessment. Similar to [32], SkillNet also uses multiple body-worn IMU sensors focused on assessing sports activity. However, it differs in that the intuitive assessment is accomplished through a novel multi-tasking approach that operates end-to-end.

Just Accepted

## 3 Preliminary state-of-the-art study

In the state-of-the-art studies [17] [18], propose an instance-based template matching scoring algorithm, a distance-based error learning *(DBEL)*. It enables the capture and detection of minute discrepancies and distinctive traits of the lower limbs (both legs) between professional players and other players from different levels of expertise. The objective was to estimate the players' performance using a body sensor network environment. The studies denote upper limbs movements as ***stroke*** and lower limbs as ***stance*** shown in Figure 3. ***Error*** is defined as the minute discrepancies between the stance & stroke performed by the professional w.r.t the other players. In a badminton shot, a professional player's data samples might incur some data variance, which is trivial to understand. Furthermore, to quantify the inconsistency and variations in the movements of the players' limbs w.r.t the professional players, they computed the Euclidean distance of lower limb samples between the professional and other players.

**Table 1.** *Component-wise time complexity* profiling of the *DBEL* module

| **Component** | **Execution Time** |
|---|---|
| B(i) | ≈ 13.78 minutes |
| B(ii) | ≈ 7.50 minutes |
| B(iii) | ≈ 4.20 minutes |
| B(iv) | ≈ 1.58 minutes |
| **B (Total)** | ≈ 27.00 minutes |

However, the limitations of a few SOTA preliminary studies motivated us to develop a multi-task learning framework that can overcome the challenges. Furthermore, we highlighted a few of the challenges of the preliminary studies along with relevant solutions below:

### 3.1 Challenges

- ***Computation resource requirements***: Profiling results (Table 1) indicate that components B(i) and B(ii) constitute the primary performance bottlenecks, together accounting for the majority of the ≈ 27-minute total execution time and consuming 60% of the available 64 GB RAM. Such high resource demands render deployment of the *DBEL* module impractical for resource-constrained devices.
- ***Technical limitations***: The *DBEL* framework exhibits limited generalizability and scalability due to its dependence on professional datasets. It relies on handcrafted error metrics (*MSE*, *RMSE*, *MdAE*, etc.) for player performance estimation, which introduce significant computational overhead. This high cost creates bottlenecks in real-time deployment scenarios. Additionally, the adaptability of the framework across diverse tasks and activities remains unverified.
- ***Limited data availability at inference***: The *DBEL* framework processes only lower-limb sensory data at inference, restricting its ability to fully capture and evaluate overall player performance.

### 3.2 Solution

- ***BSN pipeline for collective data inference***: We postulate that multiple incoming sensory data with collective inference will help to build a robust and generalized error estimation framework.
- ***Adaptability and generalizability characteristics***: To overcome the adaptability and generalizability characteristics, we need to ascertain that the proposed framework can be adopted to different tasks/activities and also different scale of error-estimation granularities (activity, limb and user).
- ***Deep learning (DL) based error Estimation Propagation***: We believe that an end-to-end deep learning architecture might help to mitigate and tackle the technical limitations. DL architectures have a higher convergence rate than any shallow learning algorithms due to their ability to learn the feature representation in the latent space. Therefore, *Multi-task Learning* (MTL) architecture gives an advantage where the distinctive traits in an inflow of multiple sources sensory data.

Just Accepted

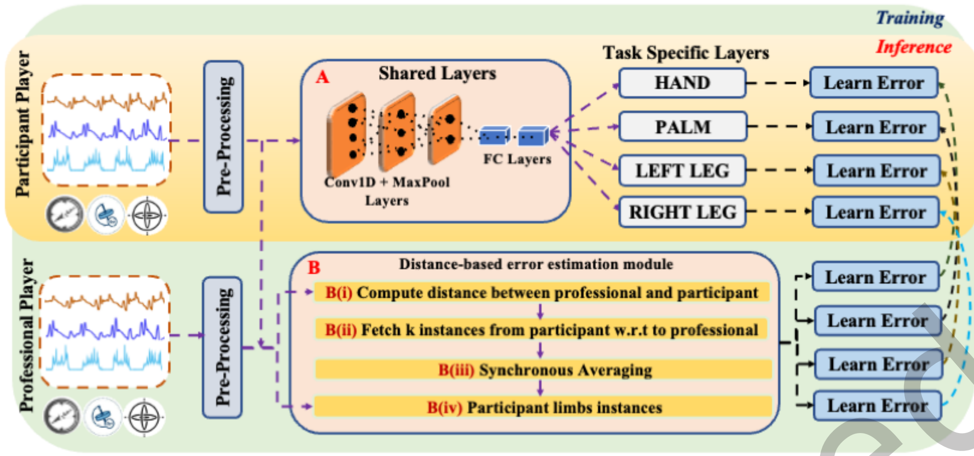



**Fig. 4.** Overall Architecture: **(A)** *Multi-task learning module*: An end-to-end data-driven error estimation pipeline to study and capture the distinctive traits between each player's limb simultaneously. **(B)** *Distance-based error learning (DBEL) module* to determine the closest $k$ number of data instance (instance-based matching learning) approach to learn the ideal body posture from other players w.r.t the professional players

- ***Limb-wise error estimation***: Lastly, to mitigate the modularity challenge, we adopt a limb-wise error estimation module, where we compute the error and capture the distinctive traits for each limb of the players, respectively. Such modularity helps to estimate and predict the most-error prone limb of the players.

## 4 System Design

We discusses the problem formulation and proposed framework adopted to tackle the above-discussed challenges.

### 4.1 Problem Formulation

Multi-task learning is an inductive transfer learning approach designed to learn multiple tasks simultaneously while exploring the feature representations corresponding to the different task similarities and dissimilarities. Given a dataset, $D$ consists of $N$ denote the number of samples, $(x_i^N, y_i^N)_{i=1}^N$ where $x_i$ and $y_i$ represent the data sample and corresponding label, respectively. The dataset, $D$, can be considered a group of $n$ sub-datasets without overlapping them, $D_1, ..., D_n$, where the sub-datasets are leveraged to learn different tasks, $T_1, ...T_n$ respectively. Equation 1, is the conventional mathematical formulation of the MTL algorithm [3, 65] where $z_i^N$ corresponds to the weight vector (regression parameters) for the $N^{th}$ task where it maps the $x^N$ sample to the corresponding label $y^N$ and $Z$ denotes concatenation of all weight vectors$z_1, z_2, ..., z_n$]. The regularizer $Reg(Z)$ denotes the regularization of constraints for $Z$ w.r.t. the prior knowledge of the data and different hypotheses of the relationship among tasks. Additionally, $\beta$ is the regularization parameter that stabilizes and balances between the regularizer $Reg(Z)$ and the overall loss optimized function shown in Equation 1.

$$\min_{z=z_1,z_2..z_n} \sum_{i=1}^{N} \mathcal{L}(x_i^N, y_i^N, z_i^N) \ + \beta Reg(Z) \tag{1}$$



Just Accepted

We aim to measure the errors from each limb of the participants in a collective inference environment. Therefore, we formulate the error measurements at different limbs as the different tasks of the proposed multi-task learning framework. To accomplish our fundamental goal, we first classify the strokes and corresponding stances by adopting the same classification module reported in the papers [17]. Secondly, we leveraged the computed handcrafted error metrics matrix employed in the state-of-the-art; ***DeCoach*** [17] work obtained from the ***DBEL*** module as depicted in Figure 4.

## 4.2 Proposed Architecture

The overall architecture is depicted in Figure 4. The architecture is built upon two majors components: MTL module and error propagation module. We describe both components below in detail:

*4.2.1* ***Multi-task Learning Module.*** The multi-task learning module consists of shared and dedicated layers. Shared layers consists of CNNs layers, max-pooling layers, batch normalization and fully connected layers shared, and responsible for learning the common feature representation among different tasks. We experimented with one-dimensional CNNs layers followed by the max-pooling and batch normalization layers. The extracted features are further processed by the dedicated layers, where the dedicated layers serve as a task-specific regressor unit. We employ four dedicated layers to learn errors from four body sensor positions. Initially, the preprocessed limb data is fed through the share layers and the error measurement module. Then, the final feature representation from the shared layer is forwarded and processed by corresponding dedicated layers, which are further processed with the error measurement module. Only the shared layers are updated simultaneously during the training, whereas each task-specific output layer updates independently.

$$EM = \sum_{i=1,j=1}^{n} [E_{\alpha}^{n}, E_{\beta}^{n}, ..., E_{\gamma}^{n}] \tag{2}$$

Equation 2, EM represents the error matrix obtained for each limb of the participant, where $E_{\alpha}$, $E_{\beta}$,..., $E_{\gamma}$ and $n$ corresponds to the sum of handcrafted error metrics for the task-specific layers (which are user-defined parameters) and denotes to the number of computed error metrics, respectively.

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \log(\cosh(X_i^{actual} - X_i^{predicted})) \tag{3}$$

*4.2.2* ***Error Propagation in MTL Module.*** The phenomenon of multi-task learning is to learn the feature representation of the tasks simultaneously and predict or classify the tasks accordingly. We leverage the computed handcrafted error scores from the *DBEL* module as ***ground truth*** for the *MTL* module show in Equation 2. Equation 3 denotes the optimization of the regressor output function for each limb, where $\mathcal{L}$ symbolizes loss incurred for a single task, $N$ and $T$ represent the number of data instances and the number of tasks, respectively. The ***LogCosh loss*** is calculated independently for each limb. The logCosh loss[2] is similar to the mean squared error (MSE), but the difference is that it is not affected when it is occasionally inaccurate. The objective of the loss function is to minimize the loss for each limb of the players shown in Equation 3 where $X^{actual}$ and $X^{predicted}$ are values obtained from the ***DBEL*** and ***MTL*** modules, respectively.

[2]https://github.com/tuantle/regression-losses-pytorch



Just Accepted

**Algorithm 1:** Distance-based error learning (DBEL) module, where the $P_X$ and $Q_{\overline{X}}$ refers to the professional's and participant's windowed data for each limb respectively whereas $P_y$ and $Q_{\overline{y}}$ refers to the professional's and participant's activity labels of each limb data respectively

**INPUT**: Acquire data from each limb of $P_X$ and $Q_{\overline{X}}$
**OUTPUT**: Handcrafted error matrix for each limb
for each label $\in P_y \cap Q_{\overline{y}}$
  Extract data: $P_X$ *and* $Q_{\overline{X}}$
  Compute euclidean distance $\leftarrow$ Sample $K$ (25) instances from ($P_X$ and $Q_{\overline{X}} \in$ [label = ($P_y \cap Q_{\overline{y}}$)])
  $P_{X-avg} \leftarrow$ Average of $k$ closest sample instances
  $\text{Error}_{limb} \leftarrow append\ [Compute\ Error(Q_{\overline{X}}, P_{X-avg})]$
end for
return $Error_{limb}$

**Algorithm 2:** ***Multi-task learning (MTL)*** module where shared weights = $W$, learning rate = $lr$, number of epochs = *EPOCH*, update weight = $\Delta W$. *Subscript* ($\alpha, \beta, \ldots, \gamma$) and (*H, P, RL, LL*) are user-defined parameters which represents the numbers of activities, users or limbs and corresponds to error obtained from Hand, Palm, Left Leg and Right Leg, respectively

**INPUT**: $Q_{\overline{X}}$ and $\text{Error}_{limb}$ matrix
**OUTPUT**: $E_\alpha$, $E_\beta$,.., $E_\gamma$ represent the errors values
Initialize the MTL pipeline
for epoch = 1 to EPOCH, total epochs do:
 # *Forward Propogation*
  for each *limb*:
    $E_H, E_P, E_{RL}, E_{LL} = \mathcal{L}_{Limb}$
 # *Backward Propogation*
  $\Delta W = lr^* \frac{\partial LE_H}{\partial W}$, $lr^* \frac{\partial LE_P}{\partial W}$, $lr^* \frac{\partial LE_{RL}}{\partial W}$ & $lr^* \frac{\partial LE_{LL}}{\partial W}$
  end for
Until reach the number of epochs $E$
return $E_H, E_P, E_{RL}, E_{LL}$

$$\boxed{\textbf{Performance Indicator (PI)} = \underbrace{W \cdot \left( \frac{\sum_{i=1}^{N,C} P_i^s}{\sum_{i=1}^{N,C} T_i^s} \right)}_{\text{Expert-weighted score}} + \underbrace{(1-W) \cdot \left( 1 - \sum_{i=1}^{N,C} \frac{(X_i^{av} - X_i^{pv})^2}{(X_i^{av} - \text{mean}(X^{av}))^2} \right)}_{\text{AI-weighted } (\mathrm{R}^2) \text{ score}}} \tag{4}$$

Just Accepted

4.2.3 *Expert-Guided Human–AI Feedback Module.* An essential component of the *SkillNet* framework is the ability to evaluate player progression through a unified lens that incorporates both expert insights and AI-generated predictions. To this end, we propose a novel weighted *performance indicator (PI)* that quantitatively reflects personalized improvement by integrating human expert assessments with AI-driven, data-informed estimations. As defined in Equation 4, the PI metric combines two complementary components: (i) the *expert-weighted score* [17], which captures qualitative feedback from domain experts, and (ii) the *AI-weighted score*, which evaluates the model's predictive accuracy. Specifically, $N$ denotes the number of evaluated instances, $C$ reflects the evaluation context (e.g., activity type, sensor position), and $W \in [0, 1]$ represents the relative contribution of expert-based scoring compared to model-based assessment in the weighted performance indicator.

### 4.3 *SkillNet* Training Procedure

Algorithm 1 details the distance-based error learning (DBEL) module, which generates handcrafted limb-wise error metrics used as supervisory signals. Algorithm 2 describes the multi-task error learning (MTEL) module, where each task-specific regression head corresponds to a distinct sensor position in the body sensor network (e.g., hand, palm, right leg, left leg). In MTEL, shared convolutional layers extract high-level, task-agnostic feature representations from the preprocessed multi-sensor input, which are subsequently fed into task-specific layers to predict limb-specific errors. The training objective is the sum of the losses from all task-specific layers:

$$\mathcal{L}_{\text{Total}} = \sum_{i=1}^{T} \mathcal{L}_i \tag{5}$$

where $T$ is the total number of tasks (sensor positions).

For the four-sensor case, this expands to:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{Hand}} + \mathcal{L}_{\text{Palm}} + \mathcal{L}_{\text{Right Leg}} + \mathcal{L}_{\text{Left Leg}} \tag{6}$$

Each $\mathcal{L}_{\text{limb}}$ is computed independently using the Log-Cosh loss between the MTEL-predicted errors and the DBEL-provided ground truth. During backpropagation, shared parameters are updated jointly, while task-specific parameters are updated independently to preserve limb-specific dynamics.

## 5 Experimentation Pipeline

The experiments were conducted on a Linux server. The server housed an Intel i7-6850K CPU, 4x NVIDIA GeForce GTX 1080Ti GPUs and 64 GB RAM. All the codes for data preprocessing and deep learning mechanisms were implemented with python. Especially for deep learning tasks were implemented using PyTorch libraries.

- **RQ1 (Error Quantification and Source Attribution):** How accurately can *SkillNet* quantify prediction errors and trace them to biomechanical, postural, and coordination factors across dynamic wHAR scenarios?
- **RQ2 (Scalability and Robustness Across Granularities):** To what extent does *SkillNet* maintain robustness and scalability across user-, limb-, and activity-level granularities and diverse wHAR datasets?
- **RQ3 (Edge Deployment Profiling and Inference Efficiency):** What are the computational, latency, memory, and power trade-offs of deploying *SkillNet* on embedded edge devices, and how do they impact real-time inference?
- **RQ4 (Effectiveness of Human–AI Collaboration and Performance Assessment):** Does integrating expert feedback with *SkillNet* significantly enhance the accuracy and effectiveness of player performance evaluation?

### 5.1 Datasets

Here, we highlights and enumerates the publicly available datasets used in this study.

- **Badminton Activity Recognition Dataset [17]** - The BAR dataset is the in-house dataset we collected to study the participants' performance w.r.t the professional player and vice-versa. The dataset acquired from a population of *15* participants (9 males and 6 females, average age: 27 years) collected 30 iterations of each of the *12* strokes. We employ four Shimmer3 IMU wearable units, each equipped with a tri-axial low-noise accelerometer ($\pm 2g$), a wide-range/high-noise accelerometer (up to $\pm 16g$, where $g$ denotes

Just Accepted

**Table 2.** Hyperparameters of Skillnet module

| **Hyper-parameters** | **Values** |
|---|---|
| No. of maximum convolution layers | 3 |
| No. of filters in convolution layers | 256, 196, 128 |
| Convolution filter dimension | 5x1,5x1,5x1 |
| No. of maximum fully connected layers | 3 |
| No. of neurons in fully connected layers | 32, 16, 1 |
| Batch size | 256 |
| Max number of epochs | 64 |
| *W* parameter (**Default**) | 0.5 |

acceleration due to gravity), a gyroscope and a magnetometer. These sensors are strategically placed on the dominant wrist, dominant palm, left leg, and right leg of the participant to capture multi-limb motion dynamics for each stroke. We recorded the data collection session using an action camera to validate activity labeling, which serves as the ground truth for assigning labels to the activities. The dataset consists of two labels: *1) activity label 2) score label* for activity-level granularity. The score varies from [0 to 4] depending on how well the participant played the shot [17], and taking into consideration the scores obtained, we categorize the participants into three expertise groups: *professional, intermediate* and *novice*. Furthermore, in [14], we provide a detailed discussion of the overall experimental setup and annotation procedure, emphasizing that participants' expert performance guided the data collection process in each session. The BAR dataset is categorized into three categories:

- ***Controlled environment subset***: comprises 4 participants specifically chosen based on their expertise and collected within an indoor sports facility, ensuring a standardized badminton sports setting.
- ***Uncontrolled environment subset***: comprises of 7 participants who are randomly selected and collected in an open playground.
- ***Evaluation Longitudinal Subset*:** The study consists of four participants, each involved in two data collection sessions collected at badminton courts in the RAC (Retrievers Activities Center) at the University of Maryland Baltimore County. In the first session, the AI-assisted module analyzes performance by identifying errors and detecting the most error-prone limbs and actions. The second session is conducted on the subsequent day, during which participants receive targeted, video-based feedback clips derived from the BadmintonDB dataset [4], informed by both expert annotations and AI-driven evaluations, as illustrated in Figure 5. An interactive feedback interface, implemented using a flask-based framework, facilitates the user-conditioned rendering of personalized performance feedback. This session also includes a subsequent performance trial to assess the impact of personalized feedback on skill refinement. To mitigate potential biases arising from fatigue, task habituation, and reactive behaviors, the sessions were deliberately distributed across two consecutive days. This design choice was motivated by initial observations that indicated compressing both sessions into a single day led to inconsistent performance, likely due to inadequate physical exertion recovery, thereby compromising the reliability of longitudinal skill tracking.

• **MM-DOS dataset [49]**- The dataset comprised of 50 participants performing *4* workout activities: *squats, lunges, push-ups and shoulder press.* They employed multi-modalities sensing, but for our study, we focus on employing the inertial measurement units (IMU) sensors, which are attached to *9* (right upper arm, right wrist, left upper arm, etc.) different body positions at 50 Hz sampling frequency. The dataset consists of two labels: *1) activity label 2) score label* for activity-level granularity. Lastly, the authors categorize the

Just Accepted

participants into three expertise groups: *expert, intermediate* and *beginner*, in [49], the authors discuss the overall experiment setup in detail, and the *MMDOS* dataset is mostly evenly distributed among the three level of expertise. In total, *18* participants were categorized as intermediates and experts, and *11* were categorized as beginners. Lastly, due to missing data, some subjects' data were not used in our study.

- **Daily and Sports Activity dataset [1]** - The DSADS dataset has a tri-axial accelerometer, magnetometer and gyroscope sensors data from eight users (four males and females). The sensors were placed on five body positions and collected at a *25 Hz sampling frequency. They have performed 19 different activities-jumping, standing, sitting, lying on the back and right, playing basketball, etc.*

## 5.2 Data Preprocessing

This study considers the raw signals acquired from the accelerometer, gyroscope, and magnetometer as input features collected from the body-worn IMU sensor network. The raw data is venerable to noise, such as motion artifacts. So, the raw data acquired were preprocessed using a median filter to eliminate the data's noise. Further, we normalized (min-max scaler) 48 features for the *BAR* dataset. The 48 features comprise three-axis of low noise & wide range accelerometer, gyroscope, and magnetometer sensors data. Due to the low measurement range of the low noise accelerometer sensor, most of the signals were clipped due to high acceleration (more than $\pm 2g$) from jerks and swift shots. Hence, such a phenomenon encouraged us to employ the wide/high range accelerometer raw signals, enabling us to capture those jerk and swift shots. Next, we employed the sliding windowing technique for the raw features. It is widely used in sensor-based human activity recognition problems among signal processing techniques and removes motion or device artifacts from the signal dataset. We employed a sliding windowing with 50% overlap with a window size of 0.125, 0.5 and 0.5 secs at a sampling rate of 512Hz, 50 Hz and 25 Hz for the *BAR*, *MMDOS* and *DSADS* datasets. We employed the majority voting for data labelling for each window segment to select the most activity labels that occurred within each window segment. The overlap-defined windows technique is better for extracting temporal patterns for micro-activities than the activity and event-defined windows [21]. Lastly, most activities requiring high-intensity physical exertion with mostly jerk and swift actions require appropriate limb coordination and tend to require finesse, such as sports-related activities [17].

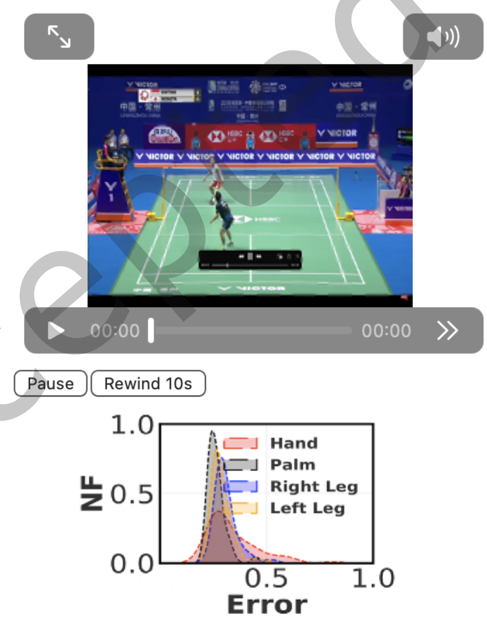



**Fig. 5.** Flask-based interactive player feedback interface tool

## 5.3 Evaluation Strategy and Setup

This section highlights the evaluation strategy and setup utilized to determine the overall performance of the *SkillNet* framework. We computed two evaluation metrics: ***mean squared error (MSE)*** and ***$R^2$ score***. Moreover, the *DBEL* module, we have re-designed each module by utilizing the exact hyperparameters used in the studies [17]. The motivation behind employing the same hyper-parameters is to maintain coupled experiment pipeline throughout the study. We employ a *60-20-20%* dataset split for training, validation and testing sets. The

Just Accepted

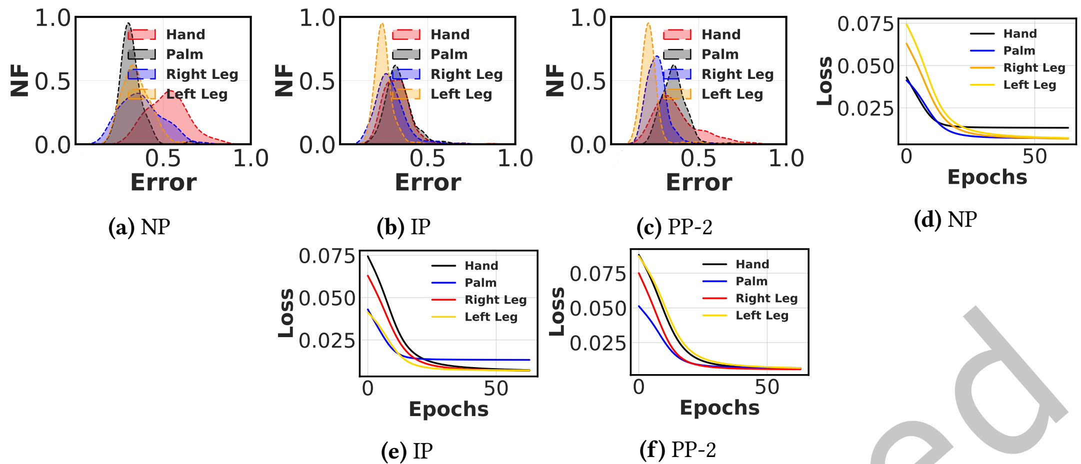



**Fig. 6.** (a-c): Error and loss plots for all the players w.r.t. the PP-1, where NF = Normalized Frequency

**Table 3.** Mean Squared Error and R-Squared Score of the players w.r.t the PP-2

| Players | Sensor Position | MSE | $R^2$ Score |
|---|---|---|---|
| PP-1 | Hand | 0.0057 | 0.5580 |
| | Palm | 0.0064 | 0.6266 |
| | Right Leg | 0.0038 | 0.6952 |
| | Left Leg | 0.0033 | 0.4786 |
| IP | Hand | 0.0011 | 0.7570 |
| | Palm | 0.0021 | 0.7131 |
| | Right Leg | 0.0035 | 0.8189 |
| | Left Leg | 0.0025 | 0.7609 |
| NP | Hand | 0.0030 | 0.4709 |
| | Palm | 0.0042 | 0.4358 |
| | Right Leg | 0.0034 | 0.5265 |
| | Left Leg | 0.0027 | 0.4598 |

**Table 4.** Mean Squared Error and R-Squared Score of the players w.r.t the PP-1

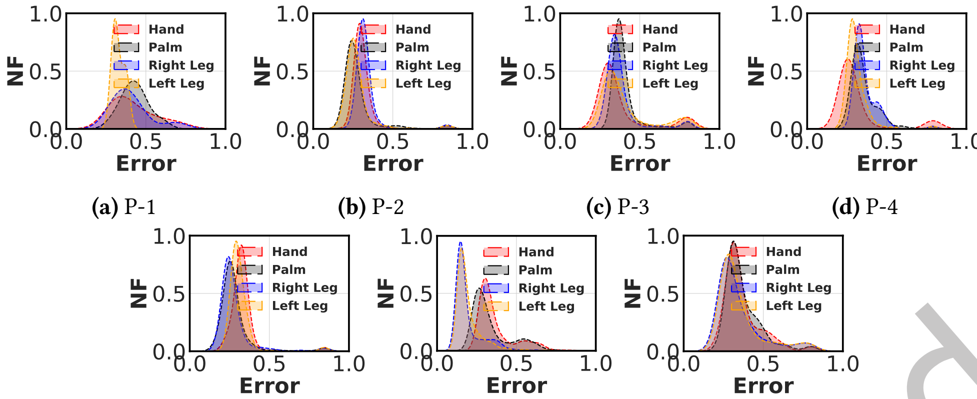
| Players | Sensor Position | MSE | $R^2$ Score |
|---|---|---|---|
| PP-2 | Hand | 0.0038 | 0.5979 |
| | Palm | 0.0020 | 0.5343 |
| | Right Leg | 0.0054 | 0.5775 |
| | Left Leg | 0.0050 | 0.4924 |
| IP | Hand | 0.0049 | 0.5582 |
| | Palm | 0.0032 | 0.4924 |
| | Right Leg | 0.0038 | 0.4378 |
| | Left Leg | 0.0041 | 0.2617 |
| NP | Hand | 0.0069 | 0.3381 |
| | Palm | 0.0058 | 0.2938 |
| | Right Leg | 0.0054 | 0.3689 |
| | Left Leg | 0.0050 | 0.3839 |

validation set was used to fine-tune the hyperparameters of the *MTL* pipeline, which is shown in table 2. The validation and test datasets were not utilized during the training phase.

## 6 Results and Discussion

We present the experimental results and analysis using the *BAR*, *MMDOS*, and *DSADS* datasets. The shared layers of the MTL pipeline employ convolutional neural network (CNN) layers to exploit their shift-invariant property for learning deep, discriminative feature representations from raw limb-wise sensor data. The results are organized according to the four research questions outlined in Section 5: *(RQ1) error estimation and error-prone source analysis, (RQ2) scalability and robustness across granularities, (RQ3) edge deployment profiling and inference efficiency*, and *(RQ4) effectiveness of human–AI collaboration and performance assessment.*

Just Accepted




**(a)** P-1 **(b)** P-2 **(c)** P-3 **(d)** P-4

**(e)** P-5 **(f)** P-6 **(g)** P-7

**Fig. 7.** (a-g): Error plots all the players w.r.t. the PP-2, where NF corresponds to normalized frequency

## 6.1 Error Estimation and Error-Prone Source Analysis (RQ1)

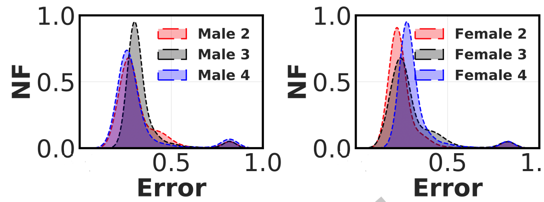
Here, we highlight and enumerate the results where we analyzed three different level-granularities (*users', limbs', and activities'*) performance assessment on the three publicly available datasets evaluated in this work. Firstly, we conduct the initial experiment to quantify the performance estimation based on *limb-level granularity*. Moving forward, Figures 6(a-c) shows the errors and losses obtained from each limb of the three participants (PP-2, IP and NP) collected in a controlled setting, which are examined with respect to the PP-1. The loss graphs show that the network successfully learned each participant's limb discrepancies and unique traits (style, speed, definitive and repetitive patterns, limb movements, etc.). Furthermore, we noticed that the loss values reduce with each iteration of the experiment, which specifies that the network is trying to learn the feature representation of the limb data of the player. Finally, figures 7(a-g) shows the errors obtained from each limb of the seven players (P-(1-7)) collected in an uncontrolled setting, which are examined w.r.t the PP-2. We plotted the probability density function ($pdf$) plots to exhibit the error values obtained. Moreover, the y-axis corresponds to the normalized frequency (probability of error occurrence) on a scale of 0-1. The x-axis corresponds to the error scores; the higher the error scores, the $pdf$ plot will be closer to 1 and vice versa. Moreover, the higher overlap between $pdf$ plots denotes higher synchronous errors occurring between the limbs.

Moreover, we extend our study to *user-level* performance assessment estimation, where we quantify the most error-prone individual among the pool of other individuals and try to understand the dynamics of the individual movements. For the experiment, we assumed the first player from the users list [1] from the female and male categories is referred to as our professional player for this study. Hence, our proposition for the DSADS dataset is based on the result we obtained through our experiments. Moreover, we perform user-level error propagation to showcase how well *SkillNet* performs in evaluating and determining users' performance for activities of daily living (ADLs) and instrumental activities of daily living (iADLs). We conducted the same experiments as those in the BAR dataset and reported the findings below. Interestingly, we observed that both male and female users exhibited similar traits. We obtained $R^2$ Score of ***80.28%*** and ***79.12%*** for female and male users, respectively. Similar trends can be observed in the error figures, where the overlap probability between the error plots is high, as shown in Figures 8 (a-b). Furthermore, we obtain ***96.89%*** and ***97.34%*** F1-score for male and female players respectively by employing the classification module reported in [17]. Lastly, from the error plots and F1-score results, we are assertive that the scale of expertise in performing ADLs and IADLs is mostly proximate and

Just Accepted

identical among the male and female players. Similarly, the error plots show a high overlapping probability among the players.

Lastly, to demonstrate the *SkillNet* framework can be effortlessly scaled to activity-level performance assessment. To achieve our objective, we consider two popular strokes played in badminton sports: *clear lob overhead forehand* and *forehand service* stroke. The results are shown in Figure 9 and Table 5, and our findings are intriguing. We found that, particularly for both strokes, the novice players are most error-prone and have a high probability of errors occurring. Similarly, the phenomenon is observed for the clear lob overhead forehand stroke, where we found that novice players (NPs) fail to play the stroke. But, due to the power exertion required to play the stroke successfully, the stroke is low. Similarly, for the forehand service stroke, because this particular stroke is one of the most effortless strokes to play in badminton sports, so we notice that there is a high probability of overlapping between the players. But then we witnessed that the intermediate player (IP) has a bit of a high error compared to the other players. The reason behind such an occurrence is that (i) the IP is a dominant right-hand player, whereas PP-1 is a left-hand player, and (ii) the power exerted and sequence of the hand movements (angle of swinging the racket) are different compared to the other players.



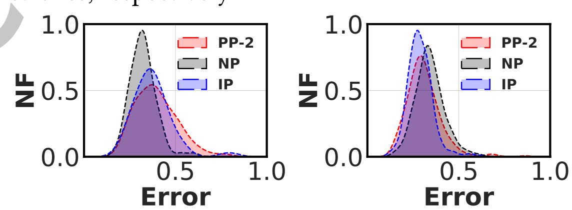

(a) Male Participants (b) Female Participants

**Fig. 8.** (a–b): Error plots for male and female participants for the DSADS dataset.

**Table 5.** Mean Squared Error and R-Squared Score for each player w.r.t the PP-1 for *clear lob overhead forehand* and *forehand service* strokes

| Strokes | Players | MSE | $R^2$ Score |
|---|---|---|---|
| Clear Lob Overhead Forehand | PP-2 | 0.0183 | 0.8128 |
| | IP | 0.0102 | 0.7804 |
| | NP | 0.0113 | 0.7920 |
| Forehand Service | PP-2 | 0.0181 | 0.7120 |
| | IP | 0.0099 | 0.7286 |
| | NP | 0.0198 | 0.6896 |

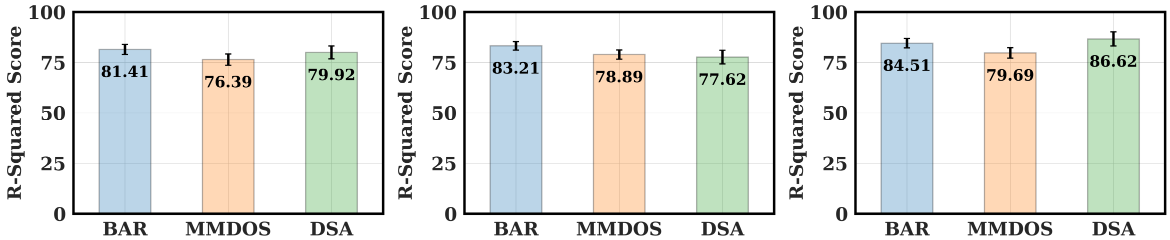
**Fig. 9.** Error plots of the players w.r.t. the PP-1 for *clear lob overhead forehand* and *forehand service* strokes, respectively




## 6.2 Scalability and Robustness Evaluation Across Granularities (RQ2)

This subsection highlights the results where we exhibit the scalability and robustness characteristics of the *SkillNet* framework. First, to accomplish our goal, we ought to learn the unique patterns and discrepancies from user-, limb- and activity-level granularities and then assess and estimate the error/performance of the individuals. Moving forward, to showcase the proposed framework's scalability characteristics that can learn and quantify the discrepancy across multiple diverse tasks, we experiment with the proposed framework across three datasets evaluated in this work shown in Figures 10. Additionally, from this investigation, we determine that the *SkillNet* framework can learn the micro-complex activities across different levels of mobility dynamics and unique activities of the individuals' actions. To support our findings and introspections, we notice that the *SkillNet* framework obtains a high R-squared score of ≈ 85% to estimate the participants' performance across sports, activities of daily living (ADLs) and workout exercises activities. Moreover, we establish the *SkillNet* framework performs

Just Accepted




**(a)** Activity-level **(b)** Limb-level **(c)** User-level

**Fig. 10.** (a-c): Scaling the SkillNet from user-, limb- and activity-level to quantify the error and assess the performance w.r.t the professional players across three datasets.

competently and comparably to the BAR and MMDOS datasets. The BAR, DSADS, and MMDOS datasets achieve a high R-squared score of $\approx 86.5\%$, which determines that the *SkillNet* works well and not only in sports analytics domain-related activities but also across diverse activity recognition tasks.

Next, we examined an intrigue phenomena, Figures 6 (c) and 11 (c), where each experiment examined w.r.t the PP-1 and PP-2, respectively. The study shows that PP-1 and PP-2 have similar trends in both experiments. To strengthen our proposition, Figs. 6, PP-2 has the highest vulnerability probability of receiving a low error score for the left leg, whereas the palm has the low vulnerability probability of receiving a high error score. Compared to in Figure 11 (c), PP-1 has the highest vulnerability probability of receiving a low error score for the left leg, whereas the palm has the low vulnerability probability of receiving a high error score. To bolster our discussion, we computed *MSE* and $R^2$ *scores* shown in Tables 3 and 4, and undoubtedly we examined similar trends on the *MSE* and $R^2$ *scores*. We believe that such phenomena are because the PP-1 & PP-2 are right and left-handed players. The style, swing, posture and muscle exertion are changed w.r.t the players' dominant limbs. We are assertive that *SkillNet* framework can successfully learn the minute and distinctive traits of the players irrespective of the dominant limb and strengthens the generalizability aspect of the *SkillNet* framework.

Moving forward, we employ a ***cross-person validation*** strategy, similar to [14, 17], where we consider another professional player, PP-2, as the reference player, as shown in the state-of-the-art (SOTA) work [17]. We performed the experiments with the players collected in an uncontrolled environment setting with respect to the PP-2 shown in Figs. 11(a-c). The motivation is to showcase that the *SkillNet* successfully learned the distinctive traits irrespective of the dominant limbs, where we experimented with both left-handed and right-handed dominant players. We determined that P-3 & P-6 and P-2 & P-4 are the best players among other male and female players, respectively. Furthermore, we noticed that the probability of an error occurring, as indicated by the *pdf plots*, is low compared to the other players. To support our propositions, the error plots shown in Figs. 7 and *MSE* and $R^2$ scores shown in Table 6, endorse and contemplate similar trends, moreover, we can also determine that for the P-2, P-3, P-4 and P-6, the error $pdf$ plot has higher overlapping (probability of error occurrence), which determines that the movements of the limbs were similar and synchronous; additionally, insights are listed in [14]. Furthermore, the reason behind this phenomenon is that, for good players, the movements of their limbs are coherent and highly coordinated when playing badminton shots. Moreover, we used the recorded videos of the data collection sessions as the ground truth to validate our conclusions and findings.

To rigorously assess the robustness of the *SkillNet* framework, we introduced controlled jitter noise [53] into one of the task-specific layers during the training process. This perturbation emulates real-world signal degradations, such as those arising from suboptimal sensor placement, inadvertent device handling, or motion-induced artifacts that may manifest during deployment. We postulate that under such perturbations, *SkillNet* should exhibit

Just Accepted

**Fig. 11.** (a–c): Error plots of the players w.r.t. the PP-2

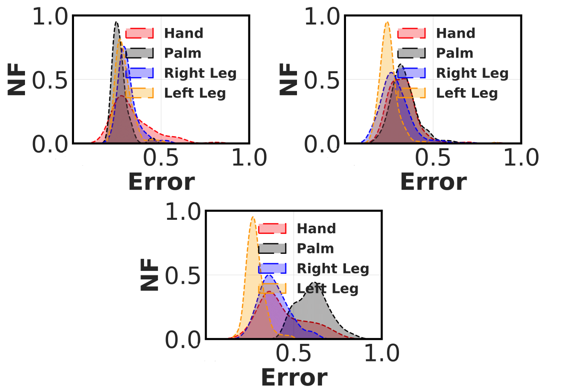


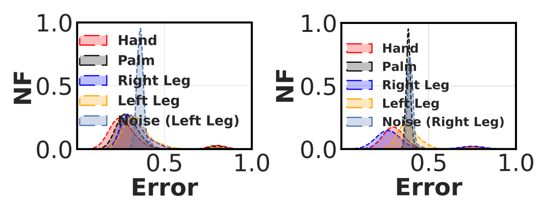

**Table 6.** Mean Squared Error and R-Squared Score for each player w.r.t the PP-2

| Players | Sensor Position | MSE | $R^2$ Score |
|---|---|---|---|
| P-2 | Hand | 0.0181 | 0.7120 |
| | Palm | 0.0099 | 0.7286 |
| | Right Leg | 0.0128 | 0.7341 |
| | Left Leg | 0.0198 | 0.6896 |
| P-3 | Hand | 0.0242 | 0.7545 |
| | Palm | 0.0368 | 0.7722 |
| | Right Leg | 0.0612 | 0.8056 |
| | Left Leg | 0.0199 | 0.7557 |
| P-4 | Hand | 0.0271 | 0.6792 |
| | Palm | 0.0106 | 0.6533 |
| | Right Leg | 0.0335 | 0.6890 |
| | Left Leg | 0.0183 | 0.7153 |
| P-6 | Hand | 0.0081 | 0.7120 |
| | Palm | 0.0095 | 0.7286 |
| | Right Leg | 0.0028 | 0.7341 |
| | Left Leg | 0.0198 | 0.6976 |

elevated error magnitudes and reduced coefficient of determination ($R^2$) values for the perturbed limb-specific models. To empirically validate this hypothesis, jitter noise (magnitude = 0.05) was injected into the task-specific layers corresponding to the left and right legs, followed by limb-wise error estimation (Figure 12). The resulting $pdf$ error plots show a pronounced increase in error frequency, accompanied by a distinct shift toward higher error scales for the perturbed limbs compared to their unperturbed counterparts. Consistently, the $R^2$ values for the right and left legs decreased to *0.085* and *0.093*, respectively. These observations substantiate that *SkillNet* can reliably discriminate between noise-contaminated samples and clean data, thereby demonstrating its resilience to unexpected sensor perturbations in practical deployment scenarios.

## 6.3 Edge Deployment Profiling and Inference Efficiency (RQ3)

This subsection highlights the results of our analysis on the feasibility of deploying the *SkillNet* on resource-constrained devices. First, we benchmark the layer-wise AI-assisted module and highlight the hardware (time) resource consumption shown in Table 7, and the codes are released [3]. The motivation is to identify and showcase any performance bottlenecks due to PyTorch operations. Another motivation is to mitigate the high computational resources challenge faced in the *DBEL* module. Table 1, *B(total)*, shows the total time consumed for an experiment. Comparatively, in Table 7, *A (total)*, shows the total time consumed from ***CPU and CUDA (GPU) processors*** took ***99.204 ms*** and ***88.825 ms***, respectively.




**(a)** Left Leg **(b)** Right Leg

**Fig. 12.** Robustness evaluation of the *SkillNet* framework with jitter noise injected into task-specific layers for left and right legs

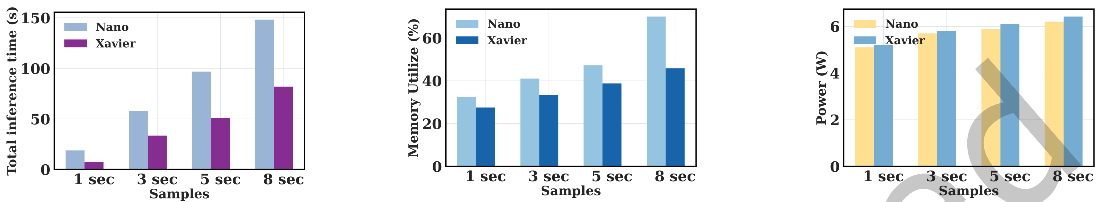
Furthermore, the execution time of the overall *MTL* pipeline is ***3.24 mins*** with total trainable parameters of

[3]https://github.com/indrajeetghosh/SkillNet-Framework

Just Accepted

**Table 7.** *Layer-wise time complexity* profiling

| Layers | CPU Time | Cuda Time |
|---|---|---|
| CNN1D + Maxpool1D + + BatchNormal1D - (i) | $\approx 1.2842\ ms$ | $\approx 13.982\ ms$ |
| CNN1D + Maxpool1D - (ii) | $\approx 1.1022\ ms$ | $\approx 10.987\ ms$ |
| CNN1D + Maxpool1D - (iii) | $\approx 1.112\ ms$ | $\approx 10.980\ ms$ |
| A (Total) | $\approx 99.404\ ms$ | $\approx 88.825\ ms$ |




**(a)** Inference Time **(b)** Memory Utilization **(c)** Power Consumption

**Fig. 13.** (a-c): Benchmark *SkillNet* framework on Jetson Nano and Jetson NX Xavier w.r.t inference time, utilized memory and power consumed, respectively

***511720***, which is comparatively very less than the *DBEL* module. Interestingly, we notice that CPU processors took less time than the CUDA processors because GPU processors work better at performing highly parallel computations, while CPU processors perform more serial computations [29].

Additionally, to determine that the *SkillNet* framework can be readily deployed to the resource-constrained embedded edge devices. We conduct and deploy the *SkillNet* framework on *Jetson Nano* and *Jetson NX Xavier* devices without performing any post-training or training-aware model compression techniques. The devices are equipped with a quad-core ARM Cortex-A57 MPCore processor, 4GB of LPDDR4 memory, a 128-core NVIDIA Maxwell GPU, and an octa-core NVIDIA Carmel ARM v8.2 CPU, as well as 8GB LPDDR5 and a 512-core Volta GPU, for the Jetson Nano and Jetson Xavier NX, respectively. Furthermore, we benchmark the pre-trained *SkillNet* upon total inference time, total memory utilized and power consumed to perform inference on **1, 3, 5,** and **8** seconds of tasks, respectively shown in Figure 13. We examine that Jetson Nano took nearly *148* seconds to run 8 seconds ($\approx$ 4000 data samples of the task); contrastingly, the Jetson NX Xavier took nearly *81.80* seconds to run the same number of samples. The Jetson NX Xavier consumed high power compared to the Jetson Nano. The results also illustrate some key findings and challenges: (i) *Jetson NX Xavier* took comparatively less time and memory but took high power consumption. However, we achieved satisfactory results; however, one of the key challenges remains in reducing computational complexity. Furthermore, this can be achieved by employing post-training or training-aware model compression techniques, as shown in [41], which will help reduce the time and memory utilized. Lastly, we demonstrate a vast range of analyses of the proposed framework *StillNet* from state-of-the-art datasets, robustness in deploying the pre-trained multi-task model to resource-constrained devices for inference without additional effort and training strategies and hypothesis testing to measure the robustness of the relationship between the deep multi-task learning pipeline and DBEL pipeline [17].

Just Accepted

## 6.4 Effectiveness of Human–AI Collaboration and Performance Assessment (RQ4)

Here, we discuss the impact of AI-assisted augmentation on player performance by integrating expert-based video clips to refine action execution while reducing errors. The performance assessment, as presented in Figure 14, comparing handcrafted and AI-based methods across two sessions reveals notable improvements in player evaluation scores. For instance, in player 1, the AI-based score improved significantly from 0.742 in session 1 to 0.881 in session 2, resulting in a weighted PI increase from 0.7010 to 0.8505, with an overall improvement margin of **0.1495**. This substantial enhancement underscores the efficacy of AI in providing more accurate and consistent performance assessments. Similarly, player 2 exhibited a marginal improvement, with AI-based scores increasing slightly from 0.538 in session 1 to 0.566 in session 2, and the weighted PI moving from 0.5515 to 0.5705. The improvement margin for player 2 is **0.0190**, which suggests gradual improvement. Similarly, our ground-truth videos corroborate these findings, showing a similar trend where Player 1 dramatically improved their playing style. The AI-assisted system provided fine-grained error analysis for specific actions and limbs, with expert video clips demonstrating the correct execution of particular actions. This detailed feedback enabled player 1 to refine their technique significantly. In contrast, player 2 did not exhibit substantial changes, indicating a more stable performance with less pronounced improvements.

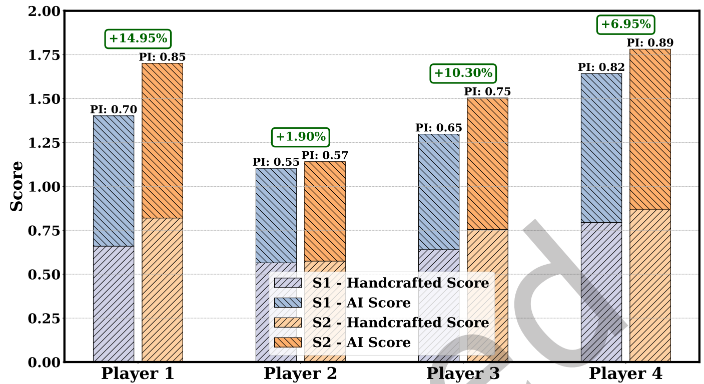



**Fig. 14.** Performance assessment on evaluation subset BAR dataset

Lastly, these results highlight and demonstrate that AI-based scoring mechanisms, augmented with expert video analysis, provide a more refined and precise assessment of player performance. This approach significantly enhances the accuracy of performance metrics by leveraging advanced algorithms to detect and correct subtle errors in technique. Such improvement can be seen in the overall handcrafted score [17] across all the players. This highlights that the integrated modules facilitate a comprehensive evaluation framework that improves performance assessments and contributes to the continuous development and refinement of the overall players' skillsets.

## 7 Limitation and Future Work

In this section, we highlight the key limitations of our work and outline directions for future improvements. First, the proposed framework is trained in a fully supervised manner, which introduces a bottleneck for large-scale and real-time deployment due to the cost and effort required for manual annotation. Second, the model is evaluated using data from only one professional player, limiting the diversity of skill representation. This may restrict the generalization capability of the learned features when encountering highly variable or fine-grained motion patterns from players with different styles. Moreover, while SkillNet achieves a favorable balance between accuracy and energy efficiency for edge deployment, we observed a trade-off wherein further reduction of network parameters or convolutional depth results in non-trivial performance degradation. Developing adaptive architectures that dynamically scale computation based on resource availability remains an important open direction. As future work, we aim to design an invariant and universal template-matching framework robust to

Just Accepted

sensor placement shifts and limb dominance variability. We will further explore label-efficient learning strategies (e.g., semi-/self-supervised learning [37, 59], knowledge distillation[39], etc.) to enhance scalability and edge deployability. Additionally, incorporating individual characteristics (e.g., height, weight, and movement strength) can improve personalization and performance interpretation.

## 8 Conclusion

This paper proposes an end-to-end BSN-driven framework, ***SkillNet*** that enables us to capture the distinctive traits between players' limbs for collective data inference and achieve $R^2$ Score of ***86%*** in predicting the players' performance. We comprehensively analyzed our multi-task learning paradigm and showcased that *MTL* can generate better error representation for the limbs. We also showcase that the proposed framework can be scaled from user-, limb- and activity-level of performance and error assessment. Moreover, we experiment on three publicly available datasets from diverse activities such as workout exercises, sports-related actions, etc. Furthermore, we demonstrate that the proposed framework precisely learns the shared feature representation for each player's limb and will be befitting to estimate the error between the player's limbs w.r.t the professional players. We believe that the model's error can be used as feedback for the player to improve the performance in the game as it will help the users determine the most error-prone limb of the player while playing the game. Moreover, we explored and showcased the adaptability, robustness and scalability characteristics of the *SkillNet* framework. Additionally, we introduced a novel weighted performance indicator (PI) metric to determine the overall improvement on the player's performance and empirical results shows an improvement of 14.95%, demonstrating AI-based scoring mechanisms are valuable in providing a robust benchmark assessment framework. We benchmark the *SkillNet* framework on resource-constrained embedded edge devices (*Jetson Nano and Jetson NX Xavier*) to demonstrate that the pipeline is readily deployable without any additional effort and introduce a layer-wise time complexity profiling to estimate the inference time for real-time deployment.

## Acknowledgments

This work has been partially supported by NSF CNS EAGER Grant #2233879, U.S. Army Grants #W911NF2120076 and #W911NF2410367, NSF IIS Grant #2509680, ONR Grant #N00014-23-1-2119, NSF CAREER Award #1750936, and NSF REU Site Grant #2050999.

## References

[1] Kerem Altun and Billur Barshan. 2010. Human activity recognition using inertial/magnetic sensor units. In *International workshop on human behavior understanding*. Springer, 38–51.

[2] Md Ariful Islam Anik, Mehedi Hassan, Hasan Mahmud, and Md Kamrul Hasan. 2016. Activity recognition of a badminton game through accelerometer and gyroscope. In *2016 19th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 213–217.

[3] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. 2006. Multi-task feature learning. *Advances in neural information processing systems* 19 (2006).

[4] Kar-Weng Ban, John See, Junaidi Abdullah, and Yuen Peng Loh. 2022. Badmintondb: A badminton dataset for player-specific match analysis and prediction. In *Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports*. 47–54.

[5] Onur Barut, Li Zhou, and Yan Luo. 2020. Multi-task LSTM Model For Human Activity Recognition and Intensity Estimation using Wearable Sensor Data. *IEEE Internet of Things Journal* (2020).

[6] Heike Brock, Yuji Ohgi, and James Lee. 2017. Learning to judge like a human: convolutional networks for classification of ski jumping errors. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 106–113.

[7] Gino Brunner, Darya Melnyk, Birkir Sigfússon, and Roger Wattenhofer. 2019. Swimming style recognition and lap counting using a smartwatch and deep learning. In *Proceedings of the 23rd International Symposium on Wearable Computers*. 23–31.

[8] Jie Cao, Youquan Wang, Haicheng Tao, and Xiang Guo. 2022. Sensor-based human activity recognition using graph LSTM and multi-task classification model. *ACM Transactions on Multimedia Computing, Communications and Applications* 18, 3s (2022), 1–19.

[9] Rich Caruana. 1997. Multitask learning. *Machine learning* 28, 1 (1997), 41–75.



Just Accepted

[10] Ling Chen, Yi Zhang, and Liangying Peng. 2020. METIER: A Deep Multi-Task Learning Based Activity and User Recognition Model Using Wearable Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–18.

[11] Zijian Chen, Wei Sun, Yuan Tian, Jun Jia, Zicheng Zhang, Wang Jiarui, Ru Huang, Xiongkuo Min, Guangtao Zhai, and Wenjun Zhang. 2024. Gaia: Rethinking action quality assessment for ai-generated videos. *Advances in Neural Information Processing Systems* 37 (2024), 40111–40144.

[12] Joseph DelPreto, Chao Liu, Yiyue Luo, Michael Foshey, Yunzhu Li, Antonio Torralba, Wojciech Matusik, and Daniela Rus. 2022. ActionSense: A Multimodal Dataset and Recording Framework for Human Activities Using Wearable Sensors in a Kitchen Environment. *Advances in Neural Information Processing Systems* 35 (2022), 13800–13813.

[13] Ramin Fallahzadeh, Mahdi Pedram, Ramyar Saeedi, Bahman Sadeghi, Michael Ong, and Hassan Ghasemzadeh. 2015. Smart-cuff: A wearable bio-sensing platform with activity-sensitive information quality assessment for monitoring ankle edema. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, 57–62.

[14] Indrajeet Ghosh, Avijoy Chakma, Sreenivasan Ramasamy Ramamurthy, Nirmalya Roy, and Nicholas Waytowich. 2022. PerMTL: A Multi-Task Learning Framework for Skilled Human Performance Assessment. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 37–44.

[15] Indrajeet Ghosh, Garvit Chugh, Abu Zaher Md Faridee, and Nirmalya Roy. 2024. Unsupervised Domain Adaptation for Action Recognition via Self-Ensembling and Conditional Embedding Alignment. In *2024 IEEE International Conference on Data Mining (ICDM)*. IEEE, 699–704.

[16] Indrajeet Ghosh, Garvit Chugh, Abu Zaher Md Faridee, and Nirmalya Roy. 2025. High-Order Moments Conditional Domain Adaptation Networks for Wearable Human Activity Recognition. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. 718–728.

[17] Indrajeet Ghosh, Sreenivasan Ramasamy Ramamurthy, Avijoy Chakma, and Nirmalya Roy. 2022. DeCoach: Deep learning-based coaching for badminton player assessment. *Pervasive and Mobile Computing* (2022), 101608.

[18] Indrajeet Ghosh, Sreenivasan Ramasamy Ramamurthy, and Nirmalya Roy. 2020. StanceScorer: A Data Driven Approach to Score Badminton Player. In *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 1–6.

[19] Mahanth Gowda, Ashutosh Dhekne, Sheng Shen, Romit Roy Choudhury, Lei Yang, Suresh Golwalkar, and Alexander Essanian. 2017. Bringing IoT to sports analytics. In *14th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 17)*. 499–513.

[20] Mahanth Gowda, Ashutosh Dhekne, Sheng Shen, Romit Roy Choudhury, Sharon Xue Yang, Lei Yang, Suresh Golwalkar, and Alexander Essanian. 2018. Iot platform for sports analytics. *GetMobile: Mobile Computing and Communications* 21, 4 (2018), 8–14.

[21] Yu Guan and Thomas Plötz. 2017. Ensembles of deep lstm learners for activity recognition using wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 1–28.

[22] Xiaonan Guo, Jian Liu, and Yingying Chen. 2017. FitCoach: Virtual fitness coach empowered by wearable mobile devices. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 1–9.

[23] Nils Y Hammerla, Shane Halloran, and Thomas Plötz. 2016. Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv preprint arXiv:1604.08880* (2016).

[24] Chunhua He, Xiaoman Wang, Yangxing Wen, Shuibin Liu, Zewen Fang, Heng Wu, Maojin Liang, and Songqing Deng. 2025. Sleep Monitoring and Sleep-Aid Intervention Methods: A Review. *IEEE Internet of Things Journal* (2025).

[25] Matthew Holbrook, Jennifer Hobbs, and Patrick Lucey. 2019. Rugby-Bot: Utilizing Multi-Task Learning & Fine-Grained Features for Rugby League Analysis. *arXiv preprint arXiv:1910.07410* (2019).

[26] HM Sajjad Hossain, Sreenivasan R Ramamurthy, Md Abdullah Al Hafiz Khan, and Nirmalya Roy. 2018. An active sleep monitoring framework using wearables. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 3 (2018), 1–30.

[27] Ryan P Hubble, Geraldine A Naughton, Peter A Silburn, and Michael H Cole. 2015. Wearable sensor use for assessing standing balance and walking stability in people with Parkinson's disease: a systematic review. *PloS one* 10, 4 (2015), e0123705.

[28] Thomas Jaitner and Wolf Gawin. 2007. Analysis of badminton smash with a mobile measure device based on accelerometry. In *ISBS-Conference Proceedings Archive*.

[29] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*. 675–678.

[30] Dusan Johnson. [n. d.]. Training sensors for soccer (aka football) players. https://gadgetsandwearables.com/2019/02/19/soccer-training-sensor/, visited 2021-01-05.

[31] Jan Kern, Thomas Lober, Joachim Hermsdörfer, and Satoshi Endo. 2022. A neural network for the detection of soccer headers from wearable sensor data. *Scientific Reports* 12, 1 (2022), 18128.

[32] Aftab Khan, James Nicholson, and Thomas Plötz. 2017. Activity recognition for quality assessment of batting shots in cricket using a hierarchical representation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–31.

Just Accepted

[33] Hao Kong, Cheng Huang, Jiadi Yu, and Xuemin Shen. 2024. A survey of mmwave radar-based sensing in autonomous vehicles, smart homes and industry. *IEEE Communications Surveys & Tutorials* 27, 1 (2024), 463–508.
[34] Chih-En Kuo, Yi-Che Liu, Da-Wei Chang, Chung-Ping Young, Fu-Zen Shaw, and Sheng-Fu Liang. 2016. Development and evaluation of a wearable device for sleep quality assessment. *IEEE Transactions on Biomedical Engineering* 64, 7 (2016), 1547–1557.
[35] Jie Li, Jiaxin Wang, Sen Qiu, Xiaofeng Liu, Jianqing Li, Wentao Xiang, Bin Liu, Songsheng Zhu, Loo Chu Kiong, Angelo Cangelosi, et al. 2024. Smart Swimming Training: Wearable Body Sensor Networks Empower Technical Evaluation of Competitive Swimming. *IEEE Internet of Things Journal* (2024).
[36] Hui Liu, Yale Hartmann, and Tanja Schultz. 2021. CSL-SHARE: A multimodal wearable sensor-based human activity dataset.
[37] Aleksej Logacjov. 2024. Self-supervised learning for Accelerometer-based human activity recognition: A survey. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 4 (2024), 1–42.
[38] Chrsitine F Martindale, Vincent Christlein, Philipp Klumpp, and Bjoern M Eskofier. 2021. Wearables-based multi-task gait and activity segmentation using recurrent neural networks. *Neurocomputing* 432 (2021), 250–261.
[39] Amir Moslemi, Anna Briskina, Zubeka Dang, and Jason Li. 2024. A survey on knowledge distillation: Recent advancements. *Machine Learning with Applications* 18 (2024), 100605.
[40] Enrica Papi, Woon Senn Koh, and Alison H McGregor. 2017. Wearable technology for spine movement assessment: A systematic review. *Journal of biomechanics* 64 (2017), 186–197.
[41] Sreenivasan Ramasamy Ramamurthy, Indrajeet Ghosh, Aryya Gangopadhyay, Elizabeth Galik, and Nirmalya Roy. 2022. STAR-Lite: A light-weight scalable self-taught learning framework for older adults' activity recognition. *Pervasive and Mobile Computing* 87 (2022), 101698.
[42] J Ishan Randika, Kanishka Rajamanthri, Avishka Kothalawala, Niroshan Gunawardana, Ashan Induranga, Pathum Weerakkody, Kaveendra Maduwantha, BTGS Kumara, and Kaveenga Koswattage. 2025. Optimized Motion Capture for Cricket Shot Classification Using Minimal Hardware and Machine Learning. *IEEE Access* (2025).
[43] Chantelle Jean Rigozzi, Gareth A Vio, and Philip Poronnik. 2023. Application of wearable technologies for player motion analysis in racket sports: A systematic review. *International Journal of Sports Science & Coaching* 18, 6 (2023), 2321–2346.
[44] Cristiano Russo, Cristian Tommasino, and Antonio Maria Rinaldi. 2025. Leveraging Graphs for Advanced Analytics in Major Team Sports: A Systematic Mapping Study. *Comput. Surveys* (2025).
[45] MS Salim, HN Lim, MSM Salim, and MY Baharuddin. 2010. Motion analysis of arm movement during badminton smash. In *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. IEEE, 111–114.
[46] Martin Seiffert, Flavio Holstein, Rainer Schlosser, and Jochen Schiller. 2017. Next generation cooperative wearables: Generalized activity assessment computed fully distributed within a wireless body area network. *IEEE Access* 5 (2017), 16793–16807.
[47] Minwoo Seong, Gwangbin Kim, Dohyeon Yeo, Yumin Kang, Heesan Yang, Joseph DelPreto, Wojciech Matusik, Daniela Rus, and SeungJun Kim. 2024. Multisensebadminton: Wearable sensor–based biomechanical dataset for evaluation of badminton performance. *Scientific Data* 11, 1 (2024), 343.
[48] Fatemeh Serpush, Mohammad Bagher Menhaj, Behrooz Masoumi, and Babak Karasfi. 2022. Wearable sensor-based human activity recognition in the smart healthcare system. *Computational intelligence and neuroscience* 2022 (2022).
[49] Ahmed Sharshar, Ahmed Fayez, Ahmed Abo Eitta, and Walid Gomaa. 2022. MM-DOS: A Novel Dataset Of Workout Activities. In *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.
[50] Spektacompowerbat. [n. d.]. https://spektacom.com/product/.
[51] Tim Steels, Ben Van Herbruggen, Jaron Fontaine, Toon De Pessemier, David Plets, and Eli De Poorter. 2020. Badminton Activity Recognition Using Accelerometer Data. *Sensors* 20, 17 (2020), 4685.
[52] Carlos Torres, Jeffrey C Fried, Kenneth Rose, and Bangalore S Manjunath. 2018. A multiview multimodal system for monitoring patient sleep. *IEEE Transactions on Multimedia* 20, 11 (2018), 3057–3068.
[53] Terry T Um, Franz MJ Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. 2017. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In *Proceedings of the 19th ACM international conference on multimodal interaction*. 216–220.
[54] Monireh Vahdati, Fedwa Laamarti, and Abdulmotaleb El Saddik. 2025. Meta-review of wearable devices for healthcare in the Metaverse. *ACM Transactions on Multimedia Computing, Communications and Applications* 21, 7 (2025), 1–36.
[55] Ben Van Herbruggen, Jaron Fontaine, Jonas Simoen, Lennert De Mey, Daniel Peralta, Adnan Shahid, and Eli De Poorter. 2024. Strategy analysis of badminton players using deep learning from IMU and UWB wearables. *Internet of Things* 27 (2024), 101260.
[56] Fei Wu, Qingzhong Wang, Jiang Bian, Ning Ding, Feixiang Lu, Jun Cheng, Dejing Dou, and Haoyi Xiong. 2022. A survey on video action recognition in sports: Datasets, methods and applications. *IEEE Transactions on Multimedia* 25 (2022), 7943–7966.
[57] Jinglin Xu, Sibo Yin, and Yuxin Peng. 2025. Human-centric Fine-grained Action Quality Assessment. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025).
[58] Youssef Yamout, Tashaffi Samin Yeasar, Shahrear Iqbal, and Mohammad Zulkernine. 2023. Beyond smart homes: An in-depth analysis of smart aging care system security. *Comput. Surveys* 56, 2 (2023), 1–35.

Just Accepted

[59] Xiangli Yang, Zixing Song, Irwin King, and Zenglin Xu. 2022. A survey on deep semi-supervised learning. *IEEE transactions on knowledge and data engineering* 35, 9 (2022), 8934–8954.
[60] Jun Ye, Hao Hu, Guo-Jun Qi, and Kien A Hua. 2017. A temporal order modeling approach to human action recognition from multimodal sensor data. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 13, 2 (2017), 1–22.
[61] Minxiang Ye, Cheng Yang, Vladimir Stankovic, Lina Stankovic, and Samuel Cheng. 2019. Distinct feature extraction for video-based gait phase classification. *IEEE Transactions on Multimedia* 22, 5 (2019), 1113–1125.
[62] Hao Yin, Paritosh Parmar, Daoliang Xu, Yang Zhang, Tianyou Zheng, and Weiwei Fu. 2025. A Decade of Action Quality Assessment: Largest Systematic Survey of Trends, Challenges, and Future Directions. *arXiv preprint arXiv:2502.02817* (2025).
[63] Cemil Zalluhoglu and Nazli Ikizler-Cinbis. 2020. Collective Sports: A multi-task dataset for collective activity recognition. *Image and Vision Computing* 94 (2020), 103870.
[64] Ling-An Zeng and Wei-Shi Zheng. 2024. Multimodal action quality assessment. *IEEE Transactions on Image Processing* 33 (2024), 1600–1613.
[65] Yu Zhang and Qiang Yang. 2017. A survey on multi-task learning. *arXiv preprint arXiv:1707.08114* (2017).
[66] Kanglei Zhou, Ruizhi Cai, Liyuan Wang, Hubert PH Shum, and Xiaohui Liang. 2024. A comprehensive survey of action quality assessment: Method and benchmark. *arXiv preprint arXiv:2412.11149* (2024).

Just Accepted