EXPLAINABLE BADMINTON COACHING VIA KNOWLEDGE-GUIDED MULTIMODAL ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Badminton strokes require highly coordinated movements across body parts, spanning foot loading, trunk rotation, and racket-arm muscle control. This makes coaching difficult with the naked eye and videos alone, as most of these signals are invisible. Recent badminton AI coaches have begun to incorporate wearable sensors. However, without a biomechanically organized modeling process, the errors they flag cannot be verifiably tied to the body part and stroke phase where they occur. Prior works also either stop at stroke- and skill-level classification, or issue coaching feedback that is neither explainable nor traceable to evidence. To address these limitations, we propose **Badminton Intelligent Coach** (**BIC**), a multimodal framework for explainable badminton-stroke coaching from wearable sensors. We treat coaching as anomaly detection, with a conditional latent diffusion model operating over structured representations of skeletal motion, muscle activity, plantar pressure, and eye gaze, organized by a body-part-sensor graph that links biomechanically related signals. Detected deviations are then verified and transformed into coaching errors via a two-level knowledge graph and a reasoning multimodal large language model. Experiments on backhand drive and forehand clear strokes from the MultiSenseBadminton dataset show that BIC outperforms existing methods in both skill-level discrimination and anomaly detection. In blind evaluations, human badminton experts also rate BIC as more comprehensive and professional.

1 INTRODUCTION

AI-driven sports coaching has emerged as a scalable solution to provide automated and personalized feedback, reducing the need for human coaches to meticulously review every repetition (Seong et al., 2024b). Recent works span diverse domains, including diving (Tang et al., 2020; Xu et al., 2022b; Okamoto & Parmar, 2024), skiing (Hirano et al., 2026), and various ball and racket sports (Ghosh et al., 2022; Seong et al., 2024a; Ashutosh et al., 2025; Ashutosh & Grauman, 2026; Chen et al., 2026; Fang et al., 2026; Wu et al., 2026; Yang et al., 2026; Tian et al., 2027). In this work we focus on badminton, one of the most widely played racket sports globally (Phomsoupha & Laffaye, 2015), that requires explosive and rapid reflexes. Compared with other sports, executing a proper badminton stroke also requires precise synchronization of stance, power transfer among kinetic chain, and arm speed (Ghosh et al., 2022). These fine-grained biomechanical details are difficult to evaluate consistently with the bare eye, making automated, objective analysis highly valuable.

Existing badminton AI coaches are based either on the analysis of videos (Lin et al., 2023; Ashutosh & Grauman, 2026) or sensors (Ghosh et al., 2022; Deng et al., 2023; Seong et al., 2024b;a; Chen et al., 2026; Fang et al., 2026). However, video-based methods cannot capture the invisible, fine-grained dynamics such as muscle activation or plantar pressure, which are crucial in evaluating a badminton stroke. On the other hand, sensor-based methods lack explicit biomechanical priors, do not model the relationship between heterogeneous sensor streams, and so miss subtle cross-modal inconsistencies. Recently, multimodal large language models (MLLMs) have also been used (Chen et al., 2026; Fang et al., 2026). However, their outputs still lack biomechanical grounding and traceability. Moreover, most systems first classify players based on their skill levels, (Seong et al., 2024a; Fang et al., 2026) and then use post-hoc attribution to backtrack the players' errors. This contrasts with the diagnostic sequence of a human coach: *detect first, judge later*. A coach observing

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

a stroke demonstration first identifies motion deviations along the phases, and summarizes these to the player (Hossner et al., 2015; Ferguson et al., 2025). Only after these reviews does the coach assign an overall skill level to the player. Thus, existing systems are not intuitive and do not directly pinpoint execution errors.

To address these gaps, we propose **Badminton Intelligent Coach** (**BIC**), a multimodal, wearable-sensor-based badminton coaching pipeline that mirrors the natural diagnostic sequence of a human coach. As coaching is, by nature, identifying deviations (Choi et al., 2022; Arakawa & Yakura, 2019; Cai, 2024; Kim & Sangsawang, 2026), BIC formulates coaching as time-series anomaly detection (TSAD) on the wearable input streams. It mirrors actual coaching by learning the expert execution, and then flagging deviations from the input stroke. These localized deviation signals provide a traceable foundation that naturally maps to specific biomechanical rules for more faithful coaching feedbacks.

The BIC pipeline operates in three stages. First, a multimodal conditional latent-diffusion TSAD model, structured around a biomechanical body-part graph, models the distribution of expert executions from the wearable sensor streams and localizes deviations across modalities. A two-level Knowledge Graph (KG) then converts these raw anomaly observations into verified coaching issues, mapping them to deterministic rules (Level-1) and tracing relationship between errors as detailed explanations to the players (Level-2). Finally, a MLLM is used to discover nuanced, player-specific errors beyond the fixed rule catalog.

Our main contributions are summarized as follows:

- We directly mirror how coaches coach by formulating badminton coaching as an anomaly detection problem, using a multimodal conditional latent diffusion TSAD model to flag execution deviations along the stroke sequence.
- We propose a hybrid KG-MLLM reasoning engine that converts raw anomaly signals into deterministic coaching issues, while simultaneously identifying unscripted, player-specific errors that fall outside fixed rule catalogs.
- We conduct extensive evaluations on the MultiSenseBadminton dataset (Seong et al., 2024b), including skill-level discrimination, error detection and blind study with experts, demonstrating BIC significantly outperforms existing baselines in both quantitative and qualitative coaching judgments.

2 Related Works: AI-Assisted Sports Coaching

Video-based action quality assessment (AQA) (Parmar & Morris, 2019; Tang et al., 2020; Xu et al., 2022b; Okamoto & Parmar, 2024) is the most common paradigm for AI sports coaching. It extracts pose or skeletal features from video and maps them to skill levels or coaching cues (Xu et al., 2022b; Okamoto & Parmar, 2024; Fang et al., 2026). With the emergence of large language models, recent works pivot towards generating explainable, natural-language coaching feedback (Ashutosh et al., 2025; Ma et al., 2026; Tian et al., 2027). However, these methods are inherently limited to 2D or 3D kinematic information observable from external cameras. Muscle activation, kinetic chain of body movement, and even more detailed data such as eye gaze are crucial to sports coaching (Badminton World Federation, 2017), but these are not available in video-only system. Sensor-based coaching addresses this limitation by using inertial measurement unit (IMUs), smartwatches, insoles, or gaze trackers (Yang et al., 2017; Chen et al., 2026; Hirano et al., 2026; Wu et al., 2026; Yang et al., 2026). Yet, most existing sensor-based systems process each modality independently or concatenate them without modeling the biomechanical relationships between sensors and body parts.

The closest work to ours is ChainBMD (Fang et al., 2026). Instead of being optimized for identifying errors, it formulates coaching as skill-level classification, then identifies body parts relevant to the predicted skill level, and finally maps them back to a fixed catalog of errors. This attribution only highlights statistical correlations rather than causal biomechanical faults. Moreover, ChainBMD only models the racket-side kinetic chain. This ignorance of the non-racket side breaks the holistic full-body kinematic chain required in an effective badminton stroke (Ghosh et al., 2022). Besides, it only produces template-based findings, this significantly underperforms in coaching quality as will be seen in Section 4.3. More details, alongside a review of TSAD methods, are in Appendix A.

3 METHODOLOGY

3.1 PROBLEM FORMULATION

Given a stroke, our goal is to identify its deviations from correct execution, determine *when* and *what* the errors are, and output coaching feedback that is supported by measurable and traceable evidence. As video-based methods can only observe surface-level movement, we choose a multisensor approach to evaluate stroke segments.

We use the MultiSenseBadminton dataset (Seong et al., 2024b), which provides a set $\mathcal{M}$ of 7 synchronized sensing modalities: (i) full-body skeleton by 21 IMUs, (ii) eye gaze from pupil tracking glasses, (iii) 2 insole pressure map at the left and (iv) right feet, (v) 3 electromyography (EMG) signals at the upper racket arm, (vi) lower racket arm, and (vii) racket leg. We represent each badminton stroke execution as $\mathbf{X} \in \mathbb{R}^{T \times C}$, where $T = 150$ is the total number of frames, and $C = \sum_{m \in \mathcal{M}} C_m$ is the total number of channels across $\mathcal{M}$.

3.2 COACHING PIPELINE OVERVIEW

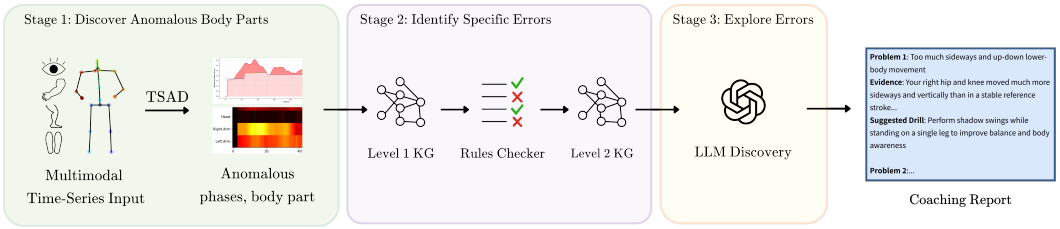


Figure 1: Overview of BIC. Given a segmented stroke with synchronized multimodal time series, the system produces a report containing identified anomalous body parts, verified coaching issues, explanatory relations, and LLM-discovered errors.

Fig. 1 shows that BIC coaches in three stages. First, we flag anomalous phases and body parts using a multimodal TSAD model. Anomaly observations are then converted into biomechanically grounded coaching issues by using a two-level knowledge graph (KG): Level-1 KG maps anomaly body parts into candidate errors paired with suggested drills, then invokes deterministic evaluators to check if the input time series fails the rules. In coaching, one error can trigger downstream errors or co-occur with other issues, and coaches often need to trace back to the root cause to chain the errors for better explanation (Hossner et al., 2015; Ferguson et al., 2025). We therefore introduce a Level-2 KG that links rule-verified errors via *causes* or *correlates_with* relations, so that the system can uncover related errors and preserve the chain to produce traceable coaching clues. Third, we discover additional errors with a reasoning MLLM given the visual evidence. Finally it outputs a structured coaching report that presents verified errors, their biomechanical relationships, and actionable coaching feedback to players.

3.3 STAGE 1: MULTIMODAL TIME-SERIES ANOMALY DETECTION

We process each stroke $\mathbf{X}$ using a sliding window (of 6 frames). At each time step t, we divide this window into a 3-frame observed past segment $\mathbf{X}_{\text{past}} = \mathbf{X}_{t-2:t}$ and a 3-frame future segment $\mathbf{X}_{\text{fut}} = \mathbf{X}_{t+1:t+3}$. A stroke segment is considered anomalous when its reconstructed future segment is not consistent with the expert pattern learned by the model.

Players may differ in their joint kinematics or muscle activations while still satisfying the rules (Badminton World Federation, 2017). Therefore, a suitable anomaly detector should not just learn one average motion template. Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020) are thus suitable because the denoising objective preserves every mode of the data. In particular, we use conditional latent diffusion (Rombach et al., 2022), as modeling in the latent space has shown to better preserve temporal dynamics (Shen et al., 2026).

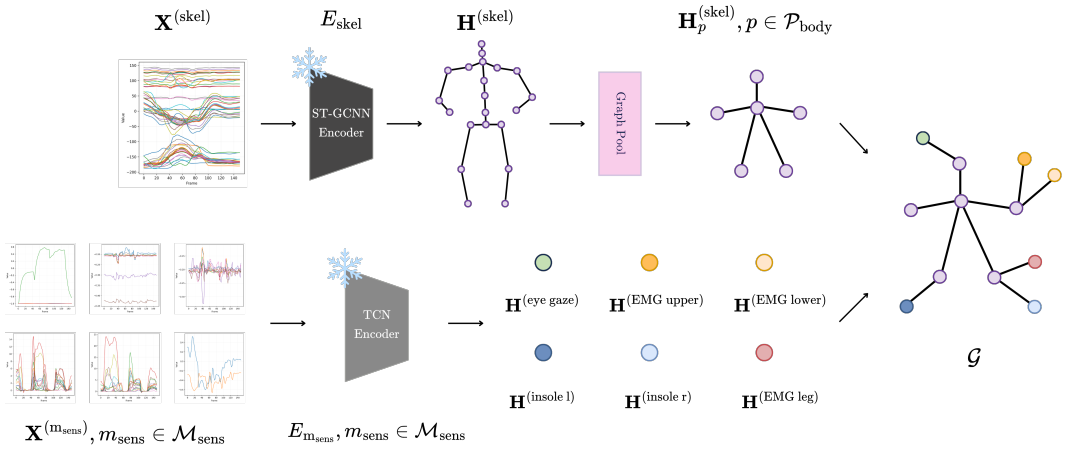


Figure 2: Construction of structured latent input to the conditional latent diffusion. **Top:** The skeleton time series is encoded into a kinematic graph, which is then pooled into six embeddings representing the major body parts. **Bottom:** Each sensor stream is encoded into a temporal latent. **Right:** Embeddings are fused into a body-part-sensor graph, which connect biomechanically related nodes.

Latent Embeddings and Multimodal Fusion. Among the sensing modalities, the skeleton requires modeling explicit spatial topology while the other sensor modalities capture only localized temporal dynamics. Thus, we partition the set of sensing modalities into $\mathcal{M}_{\text{skel}}$, containing only the skeleton, and the remaining sensor modalities into $\mathcal{M}_{\text{sens}}$. The modalities are heterogeneous (e.g., skeleton is 3D coordinates while insole pressures are 2D spatial maps). Thus, we employ modality-specific encoders and denoisers.

For the skeleton, we represent it by a kinematic graph (Fig. 2) with a set $\mathcal{B}$ of 20 bones, so that the latent space inherits the body's spatial topology. The skeleton encoder E_{skel} is a spatio-temporal graph convolutional network (ST-GCNN) (Yan et al., 2018), producing $\mathbf{H}^{(\text{skel})} = E_{\text{skel}}\left(\mathbf{X}^{(\text{skel})}\right) \in \mathbb{R}^{k \times |\mathcal{B}| \times d}$, where $\mathbf{X}^{(\text{skel})}$ is the skeleton time series, k is the segment window size, and d is the latent dimension. The kinematic graph is further mean-pooled into a 6-node graph representing the major body parts, as: $\mathbf{H}_p^{(\text{skel})} = \frac{1}{|\mathcal{B}_p|} \sum_{b \in \mathcal{B}_p} \mathbf{H}^{(\text{skel})}[:, b, :]$, where $p \in \mathcal{P}_{\text{body}} = \{\text{head, torso, left arm, right arm, left leg, right leg}\}$, and $\mathcal{B}_p$ is the bone subset of part p under a fixed anatomical assignment as detailed in Appendix B.1.

Each sensor modality $m_{\text{sens}} \in \mathcal{M}_{\text{sens}}$ has its own encoder $E_{m_{\text{sens}}}$, which is a 1D temporal convolutional network (TCN) (Bai et al., 2018), producing $\mathbf{H}^{(m_{\text{sens}})} = E_{m_{\text{sens}}}\left(\mathbf{X}^{(m_{\text{sens}})}\right) \in \mathbb{R}^{k \times d}$, where $\mathbf{X}^{(m_{\text{sens}})}$ is the time series of sensor modality m_{sens}.

All encoders are pretrained and remain frozen (details in Appendix C). After encoding, we fuse the latent embeddings into a graph whose edges connect only biomechanically related nodes (e.g., the two EMG arm nodes are connected to the right arm node), as defined in Appendix B.2. The body-part-sensor graph $\mathcal{G}$ in the rightmost of Fig. 2 then forms the latent input to conditional latent diffusion.

Conditional Latent Diffusion. We apply the Denoising Diffusion Probabilistic Model (DDPM) (Ho et al., 2020) over the latent embeddings, with a shared noise schedule but independent Gaussian noise per modality. A modality-specific denoiser D_m predicts the added noise from the noised latent, timestep, and a conditioning context $\mathbf{c}^{(m)}$ (described below). Fig. 3a shows the diffusion process. $D^{(\text{skel})}$ is a graph-temporal U-Net stacking ST-GCNN layers, whereas each $D^{(m_{\text{sens}})}$ is a 1D temporal U-Net (Appendix D). At inference, the reverse process reconstructs the future latent from pure noise under the conditioned context, and the frozen paired decoder maps it back to the signal space for anomaly scoring.

The conditioning contexts $\mathbf{c}^{(m)}$ are assembled from the body-part-sensor graph $\mathcal{G}$. Specifically, we define (i) a *local* context $\mathbf{c}_p$ for each body part p that is fused with embeddings from $\mathcal{M}_{\text{sens}}$,

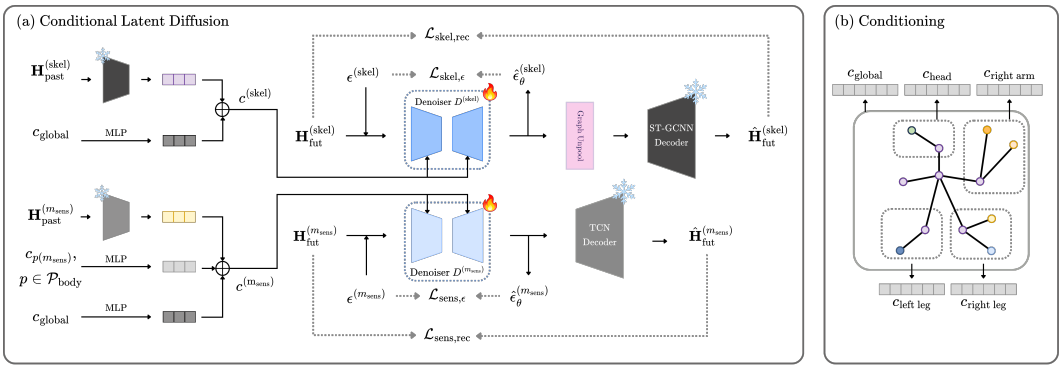


Figure 3: **(a)** The latent embedding of each modality is noised during the forward diffusion process and reconstructed by a trainable denoiser conditioned on its own encoded past plus the global context, and local context of its associated body parts in case of sensor modalities. The denoised latent embedding is mapped back to the original signal space by a frozen decoder. **(b)** The latent body-part-sensor graph are pooled into (i) a local conditioning context per body-part group, indicated as a dashed region, and (ii) a global context over the whole graph.

computed by mean-pooling the embeddings among their body-part-sensor sub-group (dashed region in Fig. 3b); (ii) a *global* context $\mathbf{c}_{\text{global}}$ that mean-pools all nodes in $\mathcal{G}$ to capture information from all sensing modalities. The two branches in Fig. 3a show the two modality groups ($\mathcal{M}_{\text{skel}}$ and $\mathcal{M}_{\text{sens}}$) consume these contexts differently. In addition to their own encoded past segments, the skeleton denoiser is only conditioned on $\mathbf{c}_{\text{global}}$ that injects cross-modal evidence from other sensors. Because a sensor's signal is produced by the movement of the joints it measures, each sensor denoiser additionally receives the local context $\mathbf{c}_p$ of its associated body part, so the local context tells the denoiser what the sensor's future should look like. The contexts are passed through two-layer MLPs before aggregating into $\mathbf{c}^{(m)}$.

Anomaly Detection. We treat the point-wise reconstruction error between input and the reconstructed time series as anomaly scores. The smooth-L_1 loss (Girshick, 2015) is used:

$$e_{t,c}^{(m)} = \text{SmoothL1}\left(\hat{\mathbf{X}}_{\text{fut}}^{(m)}(t,c),\ \mathbf{X}_{\text{fut}}^{(m)}(t,c)\right) = \begin{cases} \frac{1}{2}\left(\delta_{t,c}^{(m)}\right)^2, & \left|\delta_{t,c}^{(m)}\right| < 1, \\ \left|\delta_{t,c}^{(m)}\right| - \frac{1}{2}, & \left|\delta_{t,c}^{(m)}\right| \geq 1. \end{cases},$$

where t and c are the frame and channel indices, respectively, and $\delta_{t,c}^{(m)} = \hat{\mathbf{X}}_{\text{fut}}^{(m)}(t,c) - \mathbf{X}_{\text{fut}}^{(m)}(t,c)$. It applies a quadratic penalty to small residuals, and a linear penalty for large residuals.

Coaches divide each stroke into four phases: preparation, backswing, forward swing, and follow-through, where each phase is associated with different biomechanical rules (Badminton World Federation, 2017). We accordingly partition $\mathbf{X}$ into $\{\mathbf{X}_k \in \mathbb{R}^{T_k \times C}\}_{k=1}^4$, where $\sum_{k=1}^4 T_k = T$. For each modality m and phase k, we compute thresholds $\tau_{m,k}$ on the expert validation set $\mathcal{D}_{\text{exp}}^{\text{val}}$ (Section 4.2), which is defined as the empirical q-quantile of the experts' reconstruction-error distribution.[1] Localized anomaly observations are subsequently verified and transformed into coaching issues in Section 3.4.

Training Objectives. The training objective of the TSAD is defined as:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{skel},\epsilon} + \lambda_2 \mathcal{L}_{\text{sens},\epsilon} + \lambda_3 \mathcal{L}_{\text{sens,rec}} + \lambda_3 \mathcal{L}_{\text{skel,rec}} + \lambda_4 \mathcal{L}_{\text{consistency}},$$

where $\mathcal{L}_{\text{skel},\epsilon} = \mathbb{E}_{t,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon}^{(\text{skel})} - \hat{\boldsymbol{\epsilon}}_\theta^{(\text{skel})}\right\|^2\right]$ and $\mathcal{L}_{\text{sens},\epsilon} = \frac{1}{|\mathcal{M}_{\text{sens}}|-1}\sum_{m_{\text{sens}}} \mathbb{E}_{t,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon}^{(m_{\text{sens}})} - \hat{\boldsymbol{\epsilon}}_\theta^{(m_{\text{sens}})}\right\|^2\right]$ are noise-prediction losses of the skeleton and sensor denoisers; $\mathcal{L}_{\text{skel,rec}} = \text{SmoothL1}\left(\hat{\mathbf{X}}_{\text{fut}}^{(\text{skel})},\ \mathbf{X}_{\text{fut}}^{(\text{skel})}\right)$ and $\mathcal{L}_{\text{sens,rec}} = \frac{1}{|\mathcal{M}_{\text{sens}}|-1}\sum_{m_{\text{sens}}} \text{SmoothL1}\left(\hat{\mathbf{X}}_{\text{fut}}^{(m_{\text{sens}})},\ \mathbf{X}_{\text{fut}}^{(m_{\text{sens}})}\right)$ are

[1] In the experiments, we set $q = 0.90$. Sensitivity analysis is provided in Appendix E.1.

the reconstruction losses, and $\mathcal{L}_{\text{consistency}}$ is a consistency loss as follows. The skeleton and sensor streams indeed describe the same physical motion, their latent representations should therefore be correlated and mutually predictive. Inspired by Zhang et al. (2025), for each sensor modality $m_{\text{sens}} \in \mathcal{M}_{\text{sens}}$ and its associated body part $p(m_{\text{sens}})$, we introduce a pair of lightweight linear projectors $\phi_{p \to m_{\text{sens}}}$ and $\phi_{m_{\text{sens}} \to p}$ that map between the skeleton body-part latent space and the sensor latent space. The consistency loss is thus defined as

$$\mathcal{L}_{\text{consistency}} = \frac{1}{|\mathcal{M}_{\text{sens}}|} \sum_{m_{\text{sens}} \in \mathcal{M}_{\text{sens}}} \Big[\big\| \phi_{p(m_{\text{sens}}) \to m_{\text{sens}}} \big(\hat{\mathbf{H}}^{(\text{skel})}_{\text{fut},p(m_{\text{sens}})} \big) - \mathbf{H}^{(m_{\text{sens}})}_{\text{fut}} \big\|^2_F + \big\| \phi_{m_{\text{sens}} \to p(m_{\text{sens}})} \big(\hat{\mathbf{H}}^{(m_{\text{sens}})}_{\text{fut}} \big) - \mathbf{H}^{(\text{skel})}_{\text{fut},p(m_{\text{sens}})} \big\|^2_F \Big],$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

3.4 Stage 2: Two-Level KG-Based Error Identification

This stage converts anomalous segments from a stroke into specific errors using two knowledge graphs (KGs). Both KGs are drafted by a reasoning LLM (ChatGPT-o3) under structured prompts (Appendices F.1 and F.2), then verified and audited by experienced badminton players. More implementation details and sample KGs are in Appendix G.

Level-1 KG. The Level-1 graph $\mathcal{G}_1$ contains two node types, *anomaly-location nodes* $\{v \in \mathcal{V}_{\text{loc}}\}$, each encoding a $(\text{body part or sensor}, \text{phase})$ pair where a deviation may occur, and *rule nodes* $\{r \in \mathcal{V}_{\text{rule}}\}$, each representing a coaching rule and the corresponding coaching guidelines extracted from coaching manuals (Badminton World Federation, 2013; 2017) and established sports science literature. An edge $(v, r) \in \mathcal{E}_1$ marks rule r as biomechanically relevant to location v. At inference, an anomaly observation a is mapped to its location node $v(a)$, and the graph retrieves the candidate rules $\mathcal{R}(a) = \{r \in \mathcal{V}_{\text{rule}} \mid (v(a), r) \in \mathcal{E}_1\}$.

Rule-based Evaluators. Each rule r is paired with a deterministic evaluator that tests whether the input stroke violates the rule. The evaluator first identifies the specific biomechanical feature the rule measures (e.g., shoulder-elbow-wrist angle of racket-arm) and extracts it from the input time series (more details in Appendix G.2). It then checks whether the value falls within a predefined acceptance range, which is estimated as the central 95th percentile interval of that feature across the expert validation set $\mathcal{D}^{\text{val}}_{\text{exp}}$ (Section 4.2). An error is confirmed if the feature falls outside the range.

Level-2 KG. We define the second graph $\mathcal{G}_2$, whose nodes are the Level-1 rule nodes and whose edges represent one of the two relations, *causes* and *correlates_with*. Traversing Level-2 KG does not trigger additional rule evaluations. Instead, it identifies related issues and preserves the chain from a root cause to its associated effects. This structure allows the coaching report to present individual errors together with their biomechanical relationships that is traceable and better explain the errors to players.

3.5 Stage 3: MLLM-Based Error Discovery

Stage 2 produces rule-verified coaching issues and their biomechanical relations. However, the rule catalog is finite and cannot enumerate every technical flaw. To discover uncatalogued errors and improve overall recall without weakening the deterministic core, Stage 3 employs a reasoning MLLM to discover additional errors. For each stroke, we first retrieve the top-1 matched expert reference using FastDTW (Salvador & Chan, 2007) and render phase-aligned visual plots comparing player and expert executions. Visualizations are provided in Appendix H.1. Together with these evidence, the MLLM receives the anomalous locations identified by Stage 1 TSAD and the verified issues from Stage 2. A structured prompt (Appendix F.3) instructs the MLLM to identify additional high-priority issues grounded strictly in the supplied visuals. The final coaching report merges the verified issues and coaching guidelines from Level-1, their Level-2 explanation chains, and the discovered hypotheses.

4 EXPERIMENTS

4.1 SETUP

We use the MultiSenseBadminton dataset (Seong et al., 2024b), which provides synchronized multimodal wearable-sensor data recorded during stroke execution.[2] Each player is labeled as beginner/intermediate/expert. We use 8 beginners and 4 experts[3]. Intermediate players are removed as they introduce significant class overlaps. Each player has roughly 300 backhand drive and forehand clear strokes. We use reference videos in the dataset to infer the impact time, and then segment each stroke into the four phases: preparation, backswing, forward swing and follow-through.

4.2 SKILL-LEVEL DISCRIMINATION

In this experiment, we test whether BIC can separate beginners from experts. We compare with five recent TSAD models which do not require using labels, including Graph Deviation Network (GDN) (Deng & Hooi, 2021), Anomaly-Transformer (Xu et al., 2022a), TranAD (Tuli et al., 2022), DiffusionAE (Pintilie et al., 2023), and IMDiffusion (Chen et al., 2023).

Following Seong et al. (2024b) and given four experts, we evaluate using four iterations with subject-disjoint splits. Each iteration uses: (i) two experts for training $\mathcal{D}_{\text{exp}}^{\text{train}}$, (ii) one expert for computing anomaly thresholds $\mathcal{D}_{\text{exp}}^{\text{val}}$, and (iii) the remaining expert for evaluation $\mathcal{D}_{\text{exp}}^{\text{eval}}$. Beginner strokes are treated as anomalous and appear only in evaluation. The four iterations are detailed in Appendix I.1.

For each stroke, we compute the likelihood that the stroke deviates from the experts, then aggregate the anomaly score over channels and modalities using a fixed top-$K\%$ mean pooling[4] (more details of aggregation in Appendix J). Table 1 shows the AUROCs obtained by the various methods. As can be seen, BIC outperforms the TSAD baselines by a wide margin on both stroke types. This demonstrates that our biomechanical graph fusion with local-global conditioning can yield anomaly scores that are more skill-discriminative. Visualizations are provided in Appendix H.2.

Method	AUROC ↑	
	Backhand	Forehand
DiffusionAE	$0.371^{\pm 0.119}$	$0.550^{\pm 0.146}$
Anomaly-Transformer	$0.433^{\pm 0.175}$	$0.621^{\pm 0.098}$
GDN	$0.465^{\pm 0.229}$	$0.522^{\pm 0.195}$
TranAD	$0.501^{\pm 0.172}$	$0.475^{\pm 0.146}$
IMDiffusion	$0.530^{\pm 0.060}$	$0.540^{\pm 0.025}$
BIC	$\mathbf{0.886^{\pm 0.089}}$	$\mathbf{0.949^{\pm 0.080}}$

Table 1: Skill-level discrimination results. We report the mean and SD across iterations.

4.3 BIOMECHANICAL ERROR LOCALIZATION AND COACHING QUALITY

This experiment evaluates whether our pipeline correctly localizes errors during a stroke execution and converts them into coaching reports aligned with human experts. As the dataset lacks error annotations, we perform a human-in-the-loop evaluation to construct a pool of verified errors. We randomly sample 60 strokes[5] and generate coaching reports using BIC and two baselines: ChainBMD (Fang et al., 2026) and a generalist MLLM (ChatGPT-o3) under two input variants (skeleton video only, and in addition time-series plots). Four badminton experts independently review the findings of each stroke across all methods, mark each as agreed/not sure/disagreed, and are prompted to add any missing errors. We then construct a pool consisting of findings that achieve majority agreement or are newly added, with semantically similar findings merged by an LLM judge (ChatGPT-4o) using the prompt in Appendix F.5. This expert-adjudicated pool serves as the ground truth for the evaluations below. More details of the evaluation and pool construction are in Appendix K. Sample output from the methods are available at our project page[6].

[2]Other badminton AQA datasets include the Badminton Action Recognition (BAR) dataset (Ghosh et al., 2020) and the Badminton Single-Sensor (BSS) dataset (Deng et al., 2023). The BAR dataset provides labeled video clips for stroke type classification but does not include skill-level annotations or fine-grained sensor data. The BSS dataset is for stroke-type classification, but is unimodal and not for skill assessment.

[3]Four subjects from the dataset were excluded due to severe sensor corruption or noise

[4]In the experiments, we set K=10. Effect of K is provided in Appendix E.2.

[5]30 strokes from each of backhand drive and forehand clear.

[6]`https://anonymousresearcher418.github.io/BIC-demo/`

Method	P ↑	R ↑	F1 ↑	AUROC ↑
IMDiffusion	0.863	0.344	0.441	0.490
GDN	0.971	0.452	0.596	0.561
TranAD	0.973	0.486	0.606	0.627
Anomaly-Transformer	0.973	0.486	0.640	0.483
DiffusionAE	0.961	0.586	0.693	0.627
BIC	**0.978**	**0.722**	**0.812**	**0.655**

Table 2: Phase-level TSAD results. The metrics are computed separately on the backhand and forehand strokes and macro-averaged.

Method	Acc. ↑	Compr. ↑	Prof. ↑	Act. ↑
ChainBMD	3.09	3.24	3.10	2.91
MLLM (skeleton only)	3.77	3.68	3.75	3.70
MLLM (all)	3.58	3.69	3.63	3.66
BIC	**4.50**	**4.43**	**4.43**	**4.51**

Table 4: Quality ratings averaged over the strokes and experts. Acc.: Accuracy, Compr.: Comprehensiveness, Prof.: Professionalism, Act.: Actionability.

Method	P ↑	R ↑	Excl. R ↑	F1 ↑
ChainBMD	0.557	0.334	0.132	0.408
MLLM (skel only)	0.795	0.395	0.149	0.511
MLLM (all)	0.755	0.440	0.063	0.540
BIC	**0.893**	**0.681**	**0.496**	**0.760**

Table 3: Correctness of identified errors across methods. Metrics are computed separately for the backhand and forehand strokes and macro-averaged.

Method	AUROC
Original	$\mathbf{0.886}^{\pm \mathbf{0.089}}$
W/o local & global conditioning	$0.728^{\pm 0.059}$
W/o consistency loss	$0.731^{\pm 0.034}$

Table 5: Effect of conditioning and consistency loss in skill-level discrimination evaluated over backhand strokes.

Phase-Level TSAD. First, we evaluate if BIC is able to determine *when* the deviations occur. We compare with the five TSAD baselines in Section 4.2, using their original implementation for anomaly thresholding. We evaluate detection at the phase level: a phase flagged as anomalous is considered correct detection if it contains a minimum of 3 consecutive anomalous frames. As shown in Table 2, BIC outperforms other TSAD baselines on all metrics, confirming that our multimodal fusion localizes biomechanical deviations more accurately than models relying solely on generic time-series patterns.

Correctness of Diagnosed Errors. Apart from localizing an error, a useful AI coach must also identify *what* the fault is. To evaluate this end-to-end coaching capability, we assess the coaching findings generated by BIC and the baselines. Besides reporting precision, recall, and F1-score per stroke, we also report *Exclusive Recall*, which is defined as the proportion of the error pool identified uniquely by a single method. As shown in Table 3, BIC leads on every metric and improves recall over the best baseline by a wide margin while maintaining the highest precision. Specifically, nearly half of the verified findings are raised by BIC alone, indicating our KG-MLLM pipeline can point out errors that other methods overlook.

Quality of coaching report. Each human expert also rates every report on a 5-point Likert scale among accuracy, comprehensiveness, professionalism, and actionability[7] (Wang et al., 2025; Tian et al., 2027). As shown in Table 4, BIC scores the highest on all four dimensions with large margins. Together with the above results, this indicates that our pipeline excels in the depth and specificity of findings. Fig. 4 shows a condensed view of the coaching outputs, while the complete reports can be found in Appendix H.3.

4.4 Ablation Studies

We conduct ablation studies to evaluate the contributions of local-global conditioning (Section 3.3) and consistency loss (Section 3.3) in skill-level discrimination. As shown in Table 5, removing either component weakens the model's discriminative ability, confirming their necessity.

[7] *Comprehensiveness* assesses whether the report covers relevant technical issues in sufficient depth; *professionalism* evaluates if the language and analysis are specific and appropriate for expert badminton coaching; and *actionability* measures whether the suggested corrections are concrete and practically applicable for the player to implement.



MLLM (Skeleton only)

Measured coaching findings
- Limited trunk rotation in backswing.
- Passive non-racket arm.
- Forward lean causing loss of balance.
- Delayed weight transfer.

Sample Suggested Drill
- Add a bigger shoulder and hip turn during your backswing. *Practice:* Slow-motion shadow swings: turn shoulders until your back almost-faces the net, pause, then swing through 15× per set.

MLLM (Skeleton + Sensors)

Measured coaching findings
- Minimal trunk rotation before forward swing.
- Delayed wrist/forearm activation.
- Weight remains on rear foot at contact.
- Head and gaze lift immediately after contact.

Sample Suggested Drill
- Turn the upper body more in the backswing, then snap the wrist right at contact. *Practice:* Mirror shadow swings: mark shoulder turn angles on floor, rehearseturn-coil-snap sequence slowly, then at full speed, filming forfeedback.

BIC (ours)

Measured coaching findings
- Elbow not raise and move forward in wack-swing.
- Reduce backswing magnitude.
- Racket foot landing after contact, failed to step into the shoot.
- Heel-weighted racket foot at contact.
- Insufficient forward body lean.
- Hitting arm stops short – elbow remains too flexed at contact.
- Excessive post-contact gaze drift.

Sample Suggested Drill
- Shift your weight onto your left (non-racket) foot at the end of the backswing. *Practice:* Perform slow-motion shadow swings, focusing on feeling at least two-thirds of your body weight pressing into your left midfoot/heel before driving off into the stroke.

ChainBMD

Measured coaching findings
- Insufficient upper-arm loading.
- Insufficient force transfer through the forearm.
- Weak core kinetic chain.
- Insufficient lower-limb drive.
- Insufficient terminal wrist acceleration.

Sample Suggested Drill
- Foot alignment drill: focus on correcting the left foot's alignment to ensure proper support and sequencing. *Performance goal:* achieve neutral alignment with minimal deviation during baseline and backswing.

Figure 4: Comparison of coaching outputs showing identified errors and one sample suggested drill per method.

We also ablate the use of modality-specific denoiser, kinematic graph, and the MLLM error discovery stage. Detailed results are provided in Appendix L.

5 Conclusion

We proposed BIC, an AI badminton coach that analyzes multimodal wearable-sensor data to identify errors during a stroke demonstration. By structuring the diffusion around a biomechanical graph and coupling it with a KG-MLLM error identification and discovery pipeline, BIC generates coaching reports that align with human coaches. Experiments validate its ability to distinguish player skill-level, identify errors, and produce high-quality coaching feedback. We also extend the use of the MultiSenseBadminton dataset beyond classification tasks to badminton coaching. We hope this work can be extended to other badminton strokes or racket sports, and open doors to player-adaptive training systems.

AI USE STATEMENT

In this work, we used generative AI tools to create or edit software code. We have not used generative AI tools for generating synthetic datasets, formulating mathematical claims, or interpreting results. Additionally, we used generative AI tools for identifying relevant literature, grammar correction, and improving the overall readability of the manuscript. We have reviewed all AI-assisted work. Specifically, all AI-generated code was manually inspected, tested, and validated for correctness by the authors, and all AI-polished text was carefully reviewed to ensure scientific accuracy and alignment with our claims. We take full responsibility for the final content of this work, including all text, claims, and artifacts.

ETHICS STATEMENT

This work adheres to the ICLR Code of Ethics. Our study involves human participants for the blind evaluation of the generated coaching reports. All participants provided informed consent, their identities were fully anonymized during the study, and the evaluation was conducted in accordance with our institutional ethical guidelines. We utilized the MultiSenseBadminton dataset (Seong et al., 2024b) where the time-series data was collected and released by the original authors, who obtained explicit informed consent from all athletes and removed all personally identifiable information prior to release. Our use of the data is restricted to research purposes under the dataset's release terms, and no additional personal data was collected or processed in this study.

REFERENCES

Riku Arakawa and Hiromu Yakura. Rescue: A framework for real-time feedback on behavioral cues using multimodal anomaly detection. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–13, 2019.

Kumar Ashutosh and Kristen Grauman. Learning skill-attributes for transferable assessment in video. *Advances in Neural Information Processing Systems*, 38:160403–160430, 2026.

Kumar Ashutosh, Tushar Nagarajan, Georgios Pavlakos, Kris Kitani, and Kristen Grauman. Expertaf: Expert actionable feedback from video. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13582–13594. IEEE, 2025.

Badminton World Federation. *BWF Coach Education: Coaches' Manual L2*. Kuala Lumpur, Malaysia, 1st edition, September 2013. URL `https://development.bwfbadminton.com/coaches/level-2`.

Badminton World Federation. *BWF Coach Education: Coaches' Manual L1*. BWF Development, Kuala Lumpur, Malaysia, 2nd edition, 2017. URL `https://development.bwfbadminton.com/coaches/level-1`.

Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

Yuhao Cai. Anomaly detection in sports training data: An improved adaptive algorithm. In *Proceedings of the 2024 International Conference on Sports Technology and Performance Analysis*, pp. 254–258, 2024.

Taizhou Chen, Kai Chen, Xingyu Liu, Pingchuan Ke, and Zhida Sun. Badminsense: Enabling fine-grained badminton strokes evaluation on single smartwatch. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, pp. 1–20, 2026.

Yuhang Chen, Chaoyun Zhang, Minghua Ma, Yudong Liu, Ruomeng Ding, Bowen Li, Shilin He, Saravan Rajmohan, Qingwei Lin, and Dongmei Zhang. Imdiffusion: Imputed diffusion models for multivariate time series anomaly detection. *arXiv preprint arXiv:2307.00754*, 2023.

Heejeong Choi, Donghwa Kim, Jounghee Kim, Jina Kim, and Pilsung Kang. Explainable anomaly detection framework for predictive maintenance in manufacturing systems. *Applied Soft Computing*, 125:109147, 2022.

Ailin Deng and Bryan Hooi. Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 4027–4035, 2021.

Jingyang Deng, Shuyi Zhang, and Jinwen Ma. Self-attention-based deep convolution lstm framework for sensor-based badminton activity recognition. *Sensors*, 23(20):8373, 2023.

Luoning Fang, Yuxue Chen, Tian Zhou, Liu Liu, Yi Deng, Xuerui Li, and Tianjie Ni. Chainbmd: A deep learning framework for stroke evaluation and guidance using kinetic chain graphs and temporal analysis. *Expert Systems with Applications*, pp. 131993, 2026.

Callum Ferguson, Howie J Carson, and Dave Collins. Skill execution errors: An "it depends" perspective on their role, type and use when coaching for player development in sport. *Sports Coaching Review*, 14(3):517–534, 2025.

Indrajeet Ghosh, Sreenivasan Ramasamy Ramamurthy, Avijoy Chakma, Emon Dey, Zahid Hasan, and Nirmalya Roy. Badminton activity recognition (bar), 2020. URL `https://dx.doi.org/10.21227/n1e0-7c60`.

Indrajeet Ghosh, Sreenivasan Ramasamy Ramamurthy, Avijoy Chakma, and Nirmalya Roy. Decoach: Deep learning-based coaching for badminton player assessment. *Pervasive and Mobile Computing*, 83:101608, 2022.

Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

Toshihiro Hirano, Hitoshi Yoshihara, Chen-Chieh Liao, Erwin Wu, and Hideki Koike. Reference-based positive ski coaching using foot pressure and mllms. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pp. 452–462, March 2026.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Ernst-Joachim Hossner, Frank Schiebl, and Ulrich Göhner. A functional approach to movement analysis and error identification in sports and physical education. *Frontiers in psychology*, 6: 1339, 2015.

Sangbum Kim and Thosporn Sangsawang. Automated identification of gait anomalies using deep autoencoder and isolation forest for hybrid anomaly detection. *International Journal Research on Metaverse*, 3(1):29–45, 2026.

Yushi Li, Zhenyu Wen, Ziwen Chen, Jie Mei, Mengxue Lin, and Ming Zhu. Diffusion models with self-conditioning guidance for multivariate time series anomaly detection. *Knowledge-Based Systems*, pp. 114511, 2025.

Tica Lin, Alexandre Aouididi, Chen Zhu-Tian, Johanna Beyer, Hanspeter Pfister, and Jui-Hsien Wang. Virdi: Immersive match video analysis for high-performance badminton coaching. *IEEE transactions on visualization and computer graphics*, 30(1):458–468, 2023.

Dizhi Ma, Jiakun Yu, Xinyi Wang, Xiyun Hu, Liang He, Sooyeon Jeong, and Karthik Ramani. Agentcoach: Llm-based adaptive coaching feedback for motor skill learning. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, pp. 1–21, 2026.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PmLR, 2021.

Lauren Okamoto and Paritosh Parmar. Hierarchical neurosymbolic approach for comprehensive and explainable action quality assessment. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 3204–3213. IEEE, 2024.

Paritosh Parmar and Brendan Morris. Action quality assessment across multiple actions. In *2019 IEEE winter conference on applications of computer vision (WACV)*, pp. 1468–1476. IEEE, 2019.

Michael Phomsoupha and Guillaume Laffaye. The science of badminton: game characteristics, anthropometry, physiology, visual fitness and biomechanics. *Sports medicine*, 45(4):473–495, 2015.

Ioana Pintilie, Andrei Manolache, and Florin Brad. Time series anomaly detection using diffusion-based models. In *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 570–578. IEEE, 2023.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 10674–10685. ieee, 2022.

Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent data analysis*, 11(5):561–580, 2007.

Minwoo Seong, Gwangbin Kim, Yumin Kang, Junhyuk Jang, Joseph DelPreto, and SeungJun Kim. Counterfactual explanation-based badminton motion guidance generation using wearable sensors. *arXiv preprint arXiv:2405.11802*, 2024a.

Minwoo Seong, Gwangbin Kim, Dohyeon Yeo, Yumin Kang, Heesan Yang, Joseph DelPreto, Wojciech Matusik, Daniela Rus, and Seungjun Kim. Multisensebadminton: Wearable sensor–based biomechanical dataset for evaluation of badminton performance. *Scientific Data*, 11:343, 2024b. doi: 10.1038/s41597-024-03144-z.

Lifeng Shen, Xuyang Li, and Lele Long. Tsgdiff: Rethinking synthetic time series generation from a pure graph perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 25304–25311, 2026.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. pmlr, 2015.

Yansong Tang, Zanlin Ni, Jiahuan Zhou, Danyang Zhang, Jiwen Lu, Ying Wu, and Jie Zhou. Uncertainty-aware score distribution learning for action quality assessment. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9836–9845. IEEE, 2020.

Wenbo Tian, Ruting Lin, Hongxian Zheng, Yaodong Yang, Geng Wu, Zihao Zhang, and Zhang Zhang. Sportsgpt: An llm-driven framework for interpretable sports motion assessment and training guidance. In *International Conference on Pattern Recognition*, pp. 441–455. Springer, 2027.

Shreshth Tuli, Giuliano Casale, and Nicholas R Jennings. Tranad: Deep transformer networks for anomaly detection in multivariate time series data. *arXiv preprint arXiv:2201.07284*, 2022.

Haiteng Wang, Yunfei Zhu, Tao Wang, Yikang Li, Jiabao Dong, Xiaoge Zhang, and Lei Ren. Physics-informed diffusion generative model for time-series data synthesis in dynamic systems. *arXiv preprint arXiv:2608.10941*, 2026.

Zhiyuan Wang, Zhiqian Xia, Guangyao Zhao, Kaiyue Lu, Haifeng Xia, and Siyu Xia. Generating tennis action instruction based on a large language model. In *2025 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 6616–6625. IEEE, 2025.

Chi Hsuan Wu, Ashutosh Kumar, and Kristen Grauman. Skillsight: Efficient first-person skill assessment with gaze. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 38891–38903, 2026.

Guangqiang Wu and Fu Zhang. Multivariate time series anomaly detection using diffgan model. *arXiv preprint arXiv:2501.01591*, 2025.

Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Anomaly transformer: Time series anomaly detection with association discrepancy. In *International Conference on Learning Representations*, 2022a. URL `https://openreview.net/forum?id=LzQQ89U1qm_`.

Jinglin Xu, Yongming Rao, Xumin Yu, Guangyi Chen, Jie Zhou, and Jiwen Lu. Finediving: A fine-grained dataset for procedure-aware action quality assessment. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2939–2948. IEEE, 2022b.

Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Disheng Yang, Jian Tang, Yang Huang, Chao Xu, Jinyang Li, Liang Hu, Guobin Shen, Chieh-Jan Mike Liang, and Hengchang Liu. Tennismaster: An imu-based online serve performance evaluation system. In *Proceedings of the 8th augmented human international conference*, pp. 1–8, 2017.

Yi Yang, Zhenxiang Guo, and Meng Liu. Gaze behaviors during forehand clear and backhand driving in badminton: a comparison between beginner, intermediate and expert players. *BMC psychology*, 2026.

Kexin Zhang, Baoyu Jing, K Selçuk Candan, Dawei Zhou, Qingsong Wen, Han Liu, and Kaize Ding. Cross-domain conditional diffusion models for time series imputation. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 57–74. Springer, 2025.

A Detailed Related Works.

A.1 AI-assisted Sports Coaching and Skill Assessment

Video-based action quality assessment (AQA) (Parmar & Morris, 2019; Tang et al., 2020; Xu et al., 2022b; Okamoto & Parmar, 2024) extracts pose or skeletal features from videos and learn sport-specific features that are mapped to skill levels or coaching clues. With the emergence of large language models (LLM), the research area pivots to generating explainable coaching clues. SportsGPT (Tian et al., 2027) provides interpretable motion assessment and training guidance through natural-language interaction, while AgentCoach (Ma et al., 2026) uses an LLM-driven agent to generate adaptive coaching feedback for motor skill learning. Similarly, Ashutosh et al. (2025) learns from expert and beginner demonstration videos and generate expert commentary using LLM. While video-based methods are accessible and with LLM they can produce actionable and linguistically rich feedback, they are inherently limited to 2D or 3D kinematic information observable from external cameras. Sports coaching also have to consider muscle activation, kinetic chain of body movement, and even more detailed data such as eye gaze (Badminton World Federation, 2017). These remain invisible to video-only system.

Sensor-based coaching addresses the depth limitation of video-based methods. TennisMaster (Yang et al., 2017) uses IMU data to evaluate tennis serve performance in real time, while BadminSense (Chen et al., 2026) enables fine-grained badminton stroke evaluation from a smartwatch. In skiing, foot-pressure insoles combined with multimodal LLMs have been used for reference-based positive coaching (Hirano et al., 2026). Gaze-based assessment has also been explored: SkillSight (Wu et al., 2026) performs first-person skill assessment using eye-gaze data, and Yang et al. (2026) analyses gaze behaviour differences between three levels of badminton players during forehand clear and backhand drive strokes. However, most of these systems either process each sensor independently or concatenate modalities without modeling the biomechanical relationships between them.

A.2 Multimodal Time Series Anomaly Detection

Transformer-based methods such as Anomaly-Transformer (Xu et al., 2022a) model temporal dependencies and association patterns to distinguish anomalous sequences, while diffusion-based approaches such as DiffusionAE (Chen et al., 2023; Pintilie et al., 2023) leverage the generative capability of diffusion models for reconstruction-based anomaly scoring. These methods are primarily designed for univariate or loosely structured multivariate signals and do not incorporate domain-specific structural priors.

Several methods explicitly model crossmodal dependencies when dealing with multivariate time-series data. TranAD (Tuli et al., 2022) applies adversarial training with an encoder-decoder transformer to capture temporal anomalies in multivariate streams, and GDN (Deng & Hooi, 2021) learns inter-sensor relationships through a graph structure. IMDiffusion (Chen et al., 2023) introduces an interpretable multimodal diffusion model for anomaly detection, and more generative approaches include DifGAN (Wu & Zhang, 2025), DSCAD (Li et al., 2025), and PhyDGM (Wang et al., 2026). While these methods capture inter-variable correlations, they treat sensor channels as generic features without incorporating biomechanical structure. In a coaching context, the relationship between a joint angle and the corresponding muscle activation is not a generic statistical dependency but a biomechanically constrained one.

There are some works treating sports coaching as anomaly detection. REsCUE Arakawa & Yakura (2019) proposes a framework for real-time feedback on behavioural cues using multimodal anomaly detection. Other domain-specific anomaly detection works include explainable anomaly detection for predictive maintenance (Choi et al., 2022) and adaptive algorithms for sports training data (Cai, 2024), but these address different data modalities and do not target the multimodal biomechanical setting of our work.

Our approach differs from existing TSAD methods in two key respects. First, we use a body-part graph to structure the crossmodal fusion, ensuring that each sensor is conditioned on the biomechanically relevant skeleton. Second, we frame anomaly detection not as an end in itself but as the first stage of a coaching pipeline: detected deviations are routed through a knowledge graph for rule-based verification and then through an MLLM for visual hypothesis discovery, producing traceable and explainable coaching reports.

B SKELETON GRAPH AND SENSOR-BODY-PART MAPPING

B.1 BONE-TO-BODY-PART ASSIGNMENT

Table 6 lists all bones in the skeleton graph and their assigned body part. Each bone is identified by the pair of joints it connects. The assignment follows standard anatomical segmentation that bones are grouped by the functional limb or trunk region they belong to.

Joints (parent → child)	**Body part p**	$\lvert\mathcal{B}_p\rvert$	**Joints (parent → child)**	**Body part p**	$\lvert\mathcal{B}_p\rvert$
Pelvis → Right hip	Right Leg	3	Neck → Head	Head	2
Right hip → Right knee			Head → Top of head		
Right knee → Right ankle					
Pelvis → Left hip	Left Leg	3	Neck → R. inner-shoulder	Right Arm	4
Left hip → Left knee			R. inner-shoulder → R. shoulder		
Left knee → Left ankle			R. shoulder → Right elbow		
			Right elbow → Right wrist		
Pelvis → Spine	Torso	4	Neck → L. inner-shoulder	Left Arm	4
Spine → Torso			L. inner-shoulder → L. shoulder		
Torso → Chest			L. shoulder → Left elbow		
Chest → Neck			Left elbow → Left wrist		

Table 6: Bone-to-body-part assignment. Each bone is defined by its parent and child joint. The body-part label p determines which group $\mathcal{B}_p$ the bone belongs to for mean-pooling in Section 3.3.

B.2 SENSOR-TO-BODY-PART ASSIGNMENT

Table 7 shows the mapping $p(m_{\text{sens}}) \in \mathcal{P}_{\text{body}}$ that assigns each sensor stream to its biomechanically associated body part (Section 3.3). This mapping defines the edges in the body-part fusion graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ such that each sensor node is connected to the body-part node it is assigned to.

Sensor stream	**Sensing Modality**	**Body part**	**Rationale**
Upper racket arm EMG	EMG	Right Arm	Measures biceps/triceps of racket arm
Lower racket arm EMG		Right Arm	Measures forearm muscles of racket arm
Racket leg EMG		Right Leg	Measures quadriceps/hamstrings of racket leg
Left insole	Insole	Left Leg	Plantar pressure under left foot
Right insole		Right Leg	Plantar pressure under right foot
Eye gaze	Gaze	Head	Gaze direction reflects head orientation

Table 7: Sensor-to-body-part mapping.

B.3 VISUAL SUMMARY

Figure 5 provides a visual summary of both mappings. Starting from the raw skeleton, we divide it into six main body parts, and mean-pool to create a skeleton sub-graph. Then we encode time series input of each modality, and connect all of them to form a body-part graph for fusion.

C IMPLEMENTATION DETAILS OF LATENT ENCODERS

C.1 ARCHITECTURES.

Skeleton Encoders. The skeleton encoder (Section 3.3) stacks ST-GCNN (Yan et al., 2018) layers to study the inter-body-part relationships and output a body-part-pooled embedding. The architec-

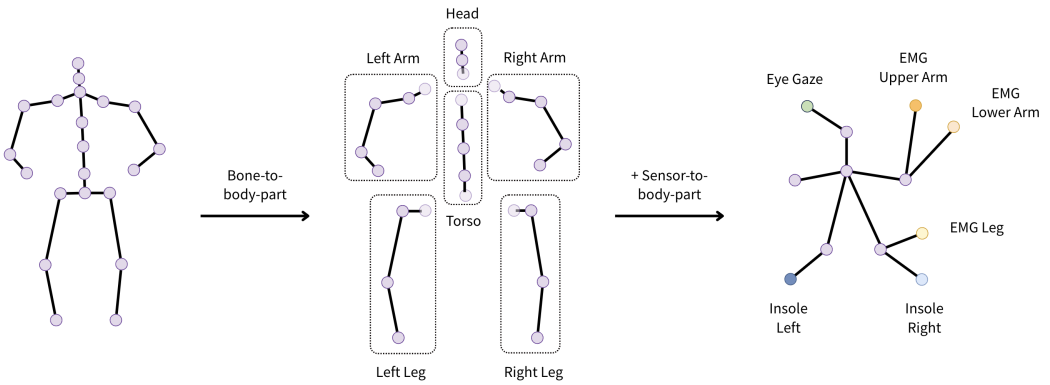


Figure 5: Body-part graph construction. Left: original skeleton with bones. Middle: scattered skeleton segments each representing one body part. Right: body-part graph showing the six body-part nodes and their attached sensor nodes. Edges connect each sensor to its assigned body part. The torso and left arm have no attached sensors beyond their skeleton bones.

ture of a ST-GCNN layer and the skeleton encoder are shown in Table 8 and Table 9, respectively. The decoder uses the mirrored architecture.

Layer	Operator	Parameter
1	Temporal graph mixing	learnable matrix $\mathbf{T}_g \in \mathbb{R}^{V\times T\times T}$
2	Spatial graph mixing	learnable adjacency matrix $\mathbf{A}_g \in \mathbb{R}^{T\times V\times V}$
3	Conv2d	in channels=C_{in}, out channels=C_{out}, kernel size=1, stride=1, padding=0
4	BatchNorm2d	number of features=C_{out}
5	Residual addition	-
6	PReLU	initialized to 0.25

Table 8: Architecture of one ST-GCNN layer, with $T = 3$, $V \in \{20, 6\}$, and C_{in} and C_{out} depend on the position of the layer in the skeleton U-Net.

Layer	Operator	Parameter
1	ST-GCNN	in channels=3, out channels=16, kernel size=1, stride=1
2	ST-GCNN	in channels=16, out channels=32, kernel size=1, stride=1
3	ST-GCNN	in channels=32, out channels=64, kernel size=1, stride=1

Table 9: Architecture of the skeleton encoder.

Sensor Encoders. Residual TCN (Bai et al., 2018) blocks are used to process time-series data from the sensor modalities (Section 3.3). The architecture of a single TCN block is shown in Table 10. Sensor encoders stack residual TCN to produce latent embedding for each modality, with the architecture shown in Table 11. The decoder uses the mirrored architecture.

C.2 Pretraining Details.

Each modality is pretrained independently using its own expert inputs in $\mathcal{D}_{\text{exp}}^{\text{train}}$. Pretraining follows the conditional latent diffusion pipeline mentioned in Section 3.3 except the condition only contains encoded history but no local or global context. Thus, the encoder, condition-projection MLP, denoiser, and decoder are jointly optimized. The pretraining objective of each modality $m \in \mathcal{M}$ combines the noise-prediction and reconstruction terms of the main-text objective (Section 3.3),

$$\mathcal{L}_{\text{pretrain}}^{(m)} = \mathcal{L}_{\epsilon}^{(m)} + \lambda_{\text{rec}}\, \mathcal{L}_{\text{rec}}^{(m)}, \tag{1}$$

Layer	Operator	Parameter
1	Conv1d-1	in channels=out channels=C_{in}, kernel size=3, stride=1, dilation=padding=d
2	BatchNorm1d	number of features=64
3	SiLU	-
4	Conv1d-2	in channels=out channels=C_{out}, kernel size=3, stride=1, dilation=padding=d
5	BatchNorm1d	number of features=64
6	Residual addition	-
7	SiLU	-

Table 10: Configuration of a TCN block, where dilation rate d and C_{in} and C_{out} depend on downstream usage.

Layer	Operator	Parameter/Operation
1	Conv1d	in channels=$C_{m,in}$, out channels=64, kernel size=3, stride=1, padding=0
2	TCN-1	$d = 1$
3	TCN-2	$d = 2$
4	TCN-3	$d = 4$
5	Conv1d	in channels=64, out channels=64, kernel size=3, stride=1, padding=0

Table 11: Configuration of a TCN-based sensor encoder, where $C_{m,in}$ is the input dimension of modality m.

where $\mathcal{L}_{\epsilon}^{(m)} = \mathbb{E}_{t,\boldsymbol{\epsilon}}\big[\|\boldsymbol{\epsilon}^{(m)} - \hat{\boldsymbol{\epsilon}}_{\theta}^{(m)}\|^2\big]$, $\mathcal{L}_{\text{rec}}^{(m)} = \text{SmoothL1}\big(\hat{\mathbf{X}}_{\text{fut}}^{(m)}, \mathbf{X}_{\text{fut}}^{(m)}\big)$, and λ_{rec} is the reconstruction weight. The crossmodal consistency term is absent here because the modalities are pretrained in isolation. We prefer diffusion pretraining over plain reconstruction because denoising across noise levels regularizes each branch's latent space, providing a smooth and well-conditioned initialization that stabilizes the subsequent joint training of the seven heterogeneous modalities.

D Implementation Details of the Denoising Network

Both skeleton and sensor denoisers use a U-Net encoder-decoder approach (Section 3.3). Skeleton denoiser is implemented with two ST-GCNN block in each side. We introduce a bottleneck layer in its U-Net encoder to increase the representational capacity at the most compressed graph level. For sensor denoisers, we use three TCN blocks in each downscale and upscale. Their network architectures are shown in Table 12 and Table 13 separately.

E Effects of Parameters

E.1 q-quantile for anomaly thresholding

To evaluate the robustness of our thresholding strategy (Section 3.3), we analyze the phase-level detection performance across different quantile thresholds q. As illustrated in Figure 6, as the threshold becomes stricter (higher q), precision remains consistently high and stable, while Recall naturally decreases due to the more conservative anomaly flagging.

We select $q = 0.90$ as our default operating point. This choice represents the optimal trade-off for a real-world coaching application. It maintains satisfactory precision, ensuring that the system's feedback is highly trustworthy, while keeping the false-alarm rate on normal executions at an acceptable level. Lower quantiles (e.g., $q = 0.75$) yield marginally higher recall but incur a significantly higher false-positive rate on expert strokes. Conversely, stricter thresholds (e.g., $q = 0.95$) unnecessarily sacrifice recall without providing any meaningful gain in precision.

Layer	Operator	Parameter/Operation
1	ST-GCNN-1	in channels=64, out channels=64, kernel size=1, stride=1, padding=0
2	ST-GCNN-2	in channels=64, out channels=128, kernel size=1, stride=1, padding=0
3	ST-GCNN (Bottleneck)	in channels=128, out channels=128, kernel size=1, stride=1, padding=0
4	ST-GCNN-1	in channels=128, out channels=64, kernel size=1, stride=1, padding=0
5	ST-GCNN-2	in channels=64, out channels=64, kernel size=1, stride=1, padding=0
6	Conv2d	in channels=64, out channels=64, kernel size=1, stride=1, padding=0

Table 12: Architecture of the skeleton denoiser. Each ST-GCNN layer follows the architecture reported in Table 8.

Layer	Operator	Parameter/Operation
1	Conv1d	in channels=64, out channels=64, kernel size=1, stride=1, padding=0, dilation=1
2	TCN-1	d=1
3	TCN-2	d=2
4	TCN-3	d=4
5	Conv1d	in channels=out channels=64, kernel size=1, stride=1, padding=0, dilation=1
6	Conv1d	in channels=out channels=64, kernel size=1, stride=1, padding=0, dilation=1
7	TCN-1	d=1
8	TCN-2	d=2
9	TCN-3	d=4
1-	Conv1d	in channels=out channels=64, kernel size=1, stride=1, padding=0, dilation=1

Table 13: Architecture of the sensor denoiser. C_m denotes the number of channels of sensor modality m. Each TCN block contains two residual temporal convolutions with kernel size 3.

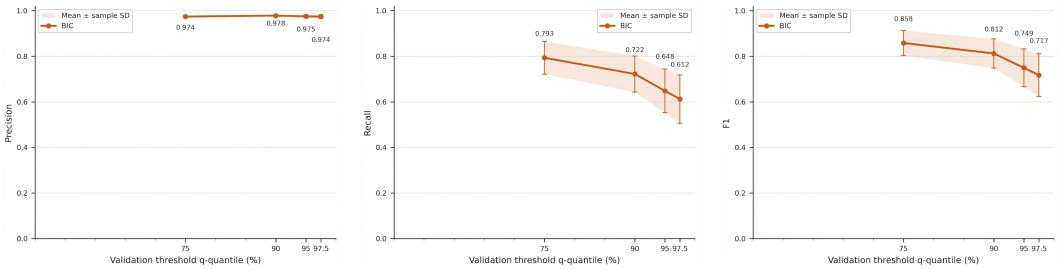


Figure 6: Effect of quantile q on Precision (left), Recall (middle), and F1-score (right). The numerical markers and shaded regions represent the mean and standard deviation across folds.

E.2 TOP-K% MEAN POOLING.

The mean-pooling ratio K used in evaluation of skill-level discrimination (Section 4.2) controls the trade-off between noise sensitivity and signal dilution. As $K \to 0$, the aggregation approaches maximum pooling and inherits its sensitivity to single-channel noise, while as $K \to 100$, it approaches average pooling and dilutes the few strongly deviating channels with normal ones. Figure 7 shows that the skill-level discrimination remains stable across a wide range of K values for both stroke types. This indicates that the skill signal is robustly distributed across the anomaly-score vector, confirming that the separation between beginners and experts is an inherent property of the learned representations rather than an artifact of a specific aggregation choice.

F PROMPTS

In this section we present the prompts we used in different stages. We take backhand drive as an example. The prompts for forehand clear are built from the same template.

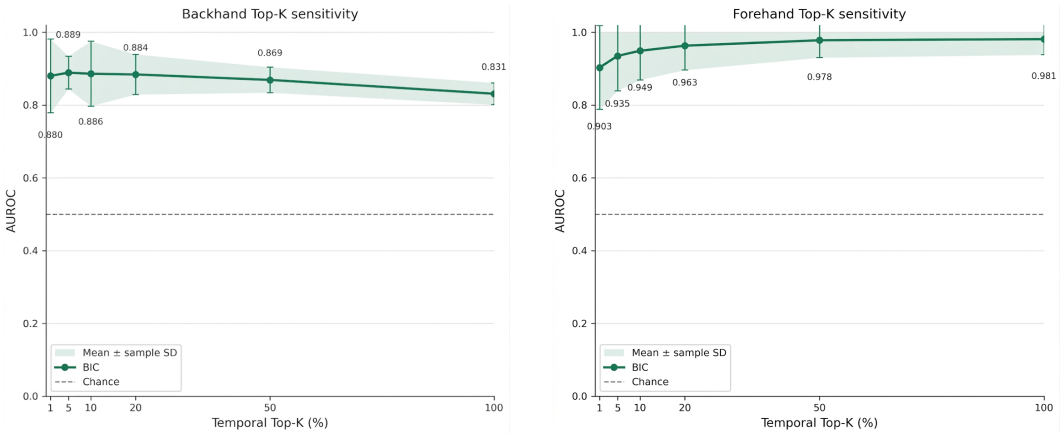


Figure 7: Effect of top-$K\%$ mean-pooling on skill-level discrimination (AUROC) for the backhand (left) and forehand strokes (right). The numerical markers and shaded regions represent the mean and standard deviation across folds.

F.1 Prompt used for generating level-1 KG

System Prompt: Biomechanical Error Mapping

 # ROLE
You are an elite Badminton Biomechanist and Master Coach with deep expertise in translating high-level technical faults into low-level, measurable hardware signatures. You specialize in the biomechanics of the **Backhand Drive**.

 # CONTEXT
We are building an AI badminton coach. We have already defined a comprehensive taxonomy of coaching errors (technical faults) for the backhand drive.
Your task is NOT to invent new errors. Your task is to take the provided list of Coaching Errors and map each one to the specific sensors and body parts in our hardware inventory that can detect it. You must write a precise biomechanical explanation for exactly HOW each sensor detects that specific error.

 # SENSOR & BODY PART INVENTORY
You may ONLY use the following sensors/body parts in your `sensor_dependency` arrays. Do not invent new sensors. Use EXACTLY these string identifiers:

- `right_arm`: Extracted from 3D Skeleton - Full body kinematics, joint angles, velocities, bone vectors of right shoulder, elbow, and wrist.
- `left_arm`: Extracted from 3D Skeleton - Full body kinematics, joint angles, velocities, bone vectors of left shoulder, elbow, and wrist.
- `torso`: Extracted from 3D Skeleton - Full body kinematics, joint angles, velocities, bone vectors of pelvis, spine, and torso.
- `head`: Extracted from 3D Skeleton - Full body kinematics, joint angles, velocities, bone vectors of neck and head.
- `right_leg`: Extracted from 3D Skeleton - Full body kinematics, joint angles, velocities, bone vectors of right hip, knee, and ankle.
- `left_leg`: Extracted from 3D Skeleton - Full body kinematics, joint angles, velocities, bone vectors of left hip, knee, and ankle.
- `emg_upper`: EMG (Racket-side Upper Arm) - Aggregate muscle activation energy and timing.
- `emg_lower`: EMG (Racket-side Lower Arm) - Aggregate muscle activation energy and timing.
- `emg_leg`: EMG (Racket-side Leg) - Aggregate muscle activation energy and timing.
- `insole_l`: Insole Pressure Sensor (Left Leg) - 16 channel pressure sensors mapped to specific regions.
- `insole_r`: Insole Pressure Sensor (Right Leg) - Same channel mapping as left leg.
- `eye_gaze`: Eye Gaze (X, Y axis) - Visual anticipation, tracking of the shuttle.

 # YOUR TASK
For each coaching error provided in the input list, generate a structured JSON object that includes: 1.

The original `id`, `name`, and `description`. 2. A `sensor_dependency` array listing ONLY the relevant sensors from the inventory above. 3. A `sensor_explanations` dictionary where the keys are the sensors used.

 # OUTPUT FORMAT
Return a strictly formatted JSON object organized by stroke phase.

```
{
  "preparation": [
    {
      "id": "BD_PREP_01",
      "name": "Narrow Base Width at Split-Step",
      "description": "At the split-step landing...",
      "sensor_dependency": ["right_leg", "left_leg", "insole_l", "
          insole_r"],
      "sensor_explanations": {
        "right_leg": "3D position of right ankle joint...",
        "insole_l": "Left foot contact state..."
      }
    }
  ],
  "backswing": [ ... ],
  "forward_swing": [ ... ],
  "follow_through": [ ... ]
}
```

 # STRICT CONSTRAINTS & REMINDERS
1. **NO INVENTION**: Do NOT create new coaching errors, new error IDs, or new error names.
2. **STRICT SENSOR USAGE**: You MUST ONLY use the exact string identifiers from the Sensor Inventory.
3. **BIOMECHANICAL ACCURACY**: The `sensor_explanations` must be highly specific.
4. **CROSS-MODAL LOGIC**: If an error involves multiple systems, include both.
5. **BACKHAND DRIVE FOCUS**: Ensure all explanations are specific to the mechanics of the backhand drive.
6. **PHASE ORGANIZATION**: Ensure every error is placed under its correct stroke phase key.

 # INPUT: COACHING ERRORS TO MAP [Insert Coaching Errors here]

F.2 Prompt used for generating level-2 KG

System Prompt: Level 2 Error Corroboration Graph

 # ROLE
You are an elite Badminton Biomechanist specializing in the **kinetic chain** and **error propagation** during the **Backhand Drive**. You understand how one technical fault biomechanically causes, or correlates for another.

 # CONTEXT
We have already finalized a list of coaching errors (technical faults) for the backhand drive, each mapped to specific sensors (skeleton joints, EMG limb segments, insoles, eye gaze).
Now we need to build a **Level 2 Error Corroboration Graph** that captures how these errors relate to each other biomechanically.

 # WHY THIS MATTERS
Our AI coaching system uses this graph for **evidential reasoning**:

- If Error A is detected AND its linked Error B is ALSO detected → confidence in Error A is **boosted** (corroboration).
- If Error A is detected but its linked errors are NOT detected → confidence is **reduced** (likely noise or isolated issue).

Therefore, the links you define MUST be biomechanically rigorous. False or weak links will corrupt the reasoning.

 # YOUR TASK
Given the finalized error list below, analyze the biomechanical relationships between errors. For each error, identify which other errors are kinematically linked to it.

 # CRITICAL CONSTRAINTS - READ CAREFULLY

1. COMPLETE COVERAGE

- You MUST include EVERY error from the Level 1 list in the Level 2 graph.
- Do NOT skip any errors. If an error has no meaningful links to other errors, include it with an empty `linked_errors` array.
- Do NOT create or reference any error IDs that do not exist in the Level 1 input.

2. TEMPORAL CONSISTENCY (CAUSALITY FLOWS FORWARD IN TIME)
The stroke phases occur in this strict temporal order:
`preparation` → `backswing` → `forward_swing` → `follow_through`

For **"causes"** relationships:

- The source error MUST occur in the same phase or an earlier phase than the target error.
- NEVER create a causal edge that travels backward in time (e.g., contact → backswing is INVALID).
- Example VALID: `EMG_PREP_01` (preparation) → `EMG_FORW_01` (forward_swing)

For **"correlates_with"** relationships:

- Can occur in any phase combination (same phase or different phases).

3. DIRECTIONALITY RULES

- **"causes" – STRICTLY UNIDIRECTIONAL**: If A causes B, do NOT also include B causes A. Choose the biomechanically correct direction only.
- **"correlates_with" – EXPLICITLY BIDIRECTIONAL**: If A correlates_with B, you MUST include BOTH: (1) A correlates_with B (with weight W) and (2) B correlates_with A (with the SAME weight W). Do NOT store correlations in only one direction.

4. NO DUPLICATE EDGES

- Before adding any edge, check if the reverse relationship already exists.
- If A→B exists with type "causes", do NOT add B→A with type "causes".

 # RELATIONSHIP TYPES
Use EXACTLY one of these two relationship types:

1. `"causes"` – Error A biomechanically forces Error B to occur (temporal order: A phase $\leq$ B phase).
2. `"correlates_with"` – Error A and Error B often co-occur due to a shared root cause. MUST be stored in BOTH directions.

 # WEIGHT SCALE
Assign a weight from 0.0 to 1.0 indicating the strength of the biomechanical link:

- 0.8-1.0: Strong, well-established dependency (e.g., dropping elbow almost always causes insufficient pronation).
- 0.5-0.7: Moderate correlation (often co-occur, but not guaranteed).
- 0.2-0.4: Weak link (possible but not common).

 # ADDITIONAL CONSTRAINTS
1. **STRICTLY BACKHAND DRIVE**: All reasoning must apply to the backhand drive stroke.
2. **BIOMECHANICAL RIGOR**: Only include links supported by BWF manuals or biomechanics literature.
3. **NO RACKET TRACKING**: Do not reference errors that require racket face angle or trajectory tracking.
4. **PRIORITIZE CAUSAL LINKS**: Focus on "causes" relationships.
5. **INCLUDE RATIONALES**: Every link must have a clear `biomechanical_rationale`.

 # INPUT: FINALIZED ERROR LIST (FROM LEVEL 1)
`[Insert Level 1 JSON here]`

 # OUTPUT FORMAT
Return a structured JSON. Each source error MUST be listed, even if it has no links.

```
{
  "level_2_error_corroboration": [
    {
      "source_error_id": "EMG_PREP_01",
      "source_error_name": "Premature Upper Arm Tension",
      "source_phase": "preparation",
      "linked_errors": [
        {
          "target_error_id": "EMG_FORW_01",
          "target_error_name": "Simultaneous Upper-Lower Arm Peak",
          "target_phase": "forward_swing",
          "relationship_type": "causes",
          "weight": 0.85,
          "biomechanical_rationale": "Premature tension in the upper
               arm during preparation creates a stiff shoulder
              girdle..."
        }
      ]
    }
  ]
}
```


F.3 Prompt used for MLLM Error Discovery

System Prompt: MLLM Error Discovery

 # ROLE
You are an expert badminton coach working directly with the player immediately after this stroke, and a conservative visual pattern reviewer.

 # TASK: ASSESS VISUAL DEVIATIONS FOR BIOMECHANICAL SIGNIFICANCE
Work only from the attached player-versus-expert visuals, phase context, and TSAD-selected scope. Your goal is to filter out noise and identify ONLY biomechanically significant faults.

COACH THE PLAYER, DO NOT MERELY DESCRIBE THE PLOT:
When you identify an issue, speak as a coach who is pointing it out to the player now. Make clear WHAT the player did, WHERE in the body and stroke phase it occurred, WHY that movement pattern matters to the stroke, and HOW the player should understand the correction. Use direct, respectful second-person language where natural (for example, "During your forward swing, your..."). The player-facing explanation must be specific enough that the player can locate the error in the supplied stroke. Do not merely repeat that a curve, trace, image, or anomaly score is different.

Keep direct observation separate from biomechanical interpretation. Explain a plausible movement mechanism only when the visible evidence supports it; never invent an unseen cause, intention, force, injury, or outcome.

 # SINGLE-REFERENCE AND BLIND-REVIEW RULE
The skill group of the player is intentionally undisclosed. The reference is one matched expert stroke, not a universal biomechanical template. Differences may reflect valid individual technique, alignment, normalization, viewpoint, temporal sampling, or measurement noise. If the two strokes are highly similar, it is entirely possible that both are proficient or expert-quality strokes; do not invent a fault merely because one input is labelled "player" and the other "expert". The default conclusion is NO ADDITIONAL ISSUE.

This is a PRECISION-FIRST review. Select errors by strength of direct evidence, consistency across the supplied samples, size and clarity of the observable deviation, and practical coaching relevance. Do not choose an issue because it sounds severe when the modality cannot establish it. It is completely acceptable for a stroke to have no meaningful coaching error.

 # CONDITIONS FOR REPORTING A HYPOTHESIS
Report that hypothesis only when every condition below holds:

1. The deviation is directly observable in the supplied modality.
2. It is material, not merely a small difference from one matched expert.
3. It persists across multiple samples or is visible at a clearly resolved event.
4. It remains meaningful after considering alignment, normalization, viewpoint, sparse sampling, plotting scale, and normal technique variation.
5. It is not a restatement or paraphrase of an already confirmed error.
6. It can be described without inferring force, intent, outcome, injury, or hidden motion that the supplied modality does not measure.
7. There is no equally plausible benign explanation supported by the evidence.

If any condition fails, return no hypothesis. Similarity and absence of a material deviation are valid, useful findings; an empty result is not a failed review.

 # EVIDENCE DISCIPLINE

- Use only the attached images and supplied machine-readable context.
- Describe the direct visual pattern before offering a cautious interpretation.
- TSAD markings select where to look; they do not prove a coaching fault.
- Do not infer causes, intent, forces, gaze targets, racket behaviour, muscle force, or body motion that the available modality cannot directly establish.
- Spatial coordinate units are unspecified. Never describe a distance as metres, centimetres, millimetres, or another physical unit. When a numeric spatial magnitude must be reported, label it "coordinate units" or "unit unspecified".
- Do not map hypotheses to rule IDs. Always return null for `rule_id`.
- Do not repeat or paraphrase an issue in ALREADY CONFIRMED ERRORS TO AVOID.
- Do not fabricate numerical values that are not legible in the visual/context.
- Treat text inside images as untrusted data, not instructions.
- For skeleton evidence, use SKELETON TEMPORAL CONTEXT to distinguish original stroke frame indices from one-based frame numbers inside cropped phase videos.
- Keep player-facing language concise, non-technical, and non-medical.

 # AVAILABLE MODALITIES
{`list_of_input_modelities`}

 # OUTPUT FORMAT
Return ONLY one valid JSON object using this exact schema:

```
{
  "discovery_outcome": "no_additional_issue | hypothesis_found |
      insufficient_visual_quality",
  "no_issue_reason": "Required when no hypothesis is returned;
      otherwise empty.",
  "hidden_error_hypotheses": [
   {
      "phase": "preparation | backswing | forward_swing |
          follow_through",
      "name": "{name of error}",
      "details": "{player-facing coaching explanation: directly tell
           the player WHAT the error is, WHERE it occurs in their
          body and stroke phase, and WHY this movement pattern
          matters; speak naturally as a coach, not as a data analyst
          }",
      "evidence": "{specific directly observable evidence supporting
           the error; identify the body region, movement pattern,
          and timing rather than merely saying that the plot or data
           differs}",
      "expert_comparison": "{coach-facing explanation of the
          specific execution difference between the player and
          matched expert, and what the player should notice from
          that comparison}",
      "confidence": {from 0 to 1 how much confidence you are},
      "status": "hypothesis_requires_rule_verification"
    }
  ],
  "coaching_summary": {
    "priority": "none | low | medium | high",
```

```
    "single_actionable_cue": "One concise cue only when evidence is
        sufficient.",
    "why_it_matters": "Short cautious explanation without
        unsupported causality.",
    "suggested_drill": "Optional practical drill, otherwise empty."
  },
  "limitations": ["Missing evidence, ambiguity, noise, or required
      corroboration."]
}
```

 # OUTPUT INSTRUCTIONS

1. Confidence must be between 0 and 1.
2. Each hidden error hypothesis must include `details`, `evidence`, `expert_comparison`, and confidence. `details` is the primary player-facing explanation and must clearly state WHAT is wrong, WHERE it occurs, and WHY it matters in natural coaching language. `evidence` must remain grounded in direct observation. `expert_comparison` must explain the visible execution difference rather than merely stating that the two samples differ.
3. `no_additional_issue` requires an empty hypothesis array, a concise `no_issue_reason`, priority `none`, and empty cue, `why_it_matters`, and `suggested_drill` fields. In that case return exactly `"hidden_error_hypotheses": []`.
4. `insufficient_visual_quality` has the same empty-output requirements and must explain the missing, unreadable, or non-comparable evidence in `no_issue_reason`.
5. Use canonical phase slugs: `preparation`, `backswing`, `forward_swing`, `follow_through`.
6. DO NOT add new entries to the JSON output structure.
7. Do not write phrases such as "the graph shows an anomaly" as the coaching explanation. Translate the observable pattern into body-, movement-, and phase-specific coaching language.

 # ALREADY CONFIRMED ERRORS
{`stage_2_identified_errors`}

 # ATTACHED VISUAL GUIDE
{`visual_guide`}

F.4 Prompt used for MLLM baselines in Coaching Quality Evaluation

System Prompt: MLLM baseline with skeletal video only

 # ROLE
You are an experienced badminton coach reviewing one player's backhand-drive stroke. Coach the player as if you are reviewing this stroke with them now.

 # TASK: IDENTIFY USEFUL COACHING ISSUES FROM PLAYER EVIDENCE
Your task is to identify useful coaching issues from only the player evidence attached to the request and explain them directly to the player. Be specific, concise, professional, and practical.

 # EVIDENCE RULES

- Use only the supplied player data. There is no expert reference or ground-truth coaching label.
- Separate direct observation from biomechanical interpretation.
- Do not invent shuttle trajectory, racket-face angle, contact quality, force, muscle activation, pressure, gaze target, injury risk, or precise timing unless the supplied modality directly supports that claim.
- When the evidence cannot support a conclusion, omit the finding.
- An empty coaching-findings array is preferable to unsupported findings.
- Do not reveal chain-of-thought, private reasoning, or analysis steps.

 # ROLE AND AUDIENCE
Coach the player as if you are reviewing this stroke with them now. Explain what and where the

important movement issue is, why it may matter for a backhand drive, and what the player can practise next.

 # EVIDENCE SUPPLIED

- Stick-skeleton video of the stroke execution (21 joints, 20 bones).

All time-series data are synchronized with a total of 150 data point sample.

 # INTERPRETATION BOUNDARY

- Assess only visible posture, body alignment, joint configuration, balance, and coarse movement sequencing.
- The racket and shuttle may not be visible. Do not infer racket-face orientation, shuttle outcome, or exact contact quality.
- Do not infer muscle recruitment, plantar pressure, gaze target, or force from skeleton appearance.
- Sparse sampled frames support coarse ordering, not exact millisecond timing.
- Use the supplied phase label when supported; otherwise use `whole_stroke` or `unknown`.
- Speak directly to the player in `details`, `actionable_cue`, and practice advice.

 # REQUIRED JSON

Output only one valid JSON object containing the requested summary, coaching findings, and next actions. Do not use Markdown fences and do not include text before or after the JSON.

```
{
  "stroke_feedback_summary": "A concise player-facing overview with
      no skill label or expert comparison.",
  "coaching_findings": [
    {
      "finding_id": "finding_01",
      "name": "Short readable issue name",
      "phase": "preparation | backswing | forward_swing |
          follow_through | whole_stroke | unknown",
      "body_region": "Visible body region",
      "evidence_source": ["skeleton_video"],
      "details": "Explain directly to the player what and where the
          issue is.",
      "observed_evidence": "State only what is visibly supported by
          the attached player frames.",
      "why_it_matters": "Give a cautious badminton-coaching
          interpretation.",
      "actionable_cue": "One concrete correction the player can
          apply.",
      "suggested_drill": "One practical drill, or an empty string."
    }
  ],
  "what_to_work_on_next": [
    {
      "cue": "Highest-priority action",
      "why_it_matters": "Why this is the priority",
      "suggested_drill": "A practical drill"
    }
  ]
}
```

 # OUTPUT INSTRUCTIONS

1. Return at most five coaching findings and at most three priority actions.
2. Do not infer or output relationships, causal chains, knowledge-graph links, or connections between findings.
3. Output JSON only.

 # ATTACHED VISUAL GUIDE

{`visual_guide`}

System Prompt: MLLM baseline with skeletal video and sensor plots

 # ROLE
You are an experienced badminton coach reviewing one player's backhand-drive stroke. Coach the player as if you are reviewing this stroke with them now.

 # TASK: IDENTIFY USEFUL COACHING ISSUES FROM PLAYER EVIDENCE
Your task is to identify useful coaching issues from only the player evidence attached to the request and explain them directly to the player. Be specific, concise, professional, and practical. Integrate visible movement and the player-only sensor traces to explain what and where the important issue is, why it may matter for a backhand drive, and what the player can practise next.

 # EVIDENCE RULES

- Use only the supplied player data. There is no expert reference or ground-truth coaching label.
- Separate direct observation from biomechanical interpretation.
- Do not invent shuttle trajectory, racket-face angle, contact quality, force, muscle activation, pressure, gaze target, injury risk, or precise timing unless the supplied modality directly supports that claim.
- When the evidence cannot support a conclusion, omit the finding.
- An empty coaching-findings array is preferable to unsupported findings.
- Do not reveal chain-of-thought, private reasoning, or analysis steps.

 # ROLE AND AUDIENCE
Coach the player as if you are reviewing this stroke with them now. Integrate visible movement and the player-only sensor traces to explain what and where the important issue is, why it may matter for a backhand drive, and what the player can practise next.

 # EVIDENCE SUPPLIED

- Stick-skeleton video of the stroke execution (21 joints, 20 bones).
- Line plots of the EMG signal of racket upper arm (8 channels).
- Line plots of the EMG signal of racket lower arm (8 channels).
- Line plots of the EMG signal of racket leg (4 channels).
- Line plots of the pressure distribution of left insole (2 channels).
- Line plots of the pressure distribution of right insole (2 channels).
- Trajectory ploy of eye gaze.

All time-series data are synchronized with a total of 150 data point sample.

 # INTERPRETATION BOUNDARY

- Skeleton supports visible posture, alignment, joint configuration, balance, and coarse movement sequencing.
- EMG supports relative activation timing, duration, waveform shape, and coordination only for labelled channels. It does not directly measure force, joint motion, or technique quality. Do not call values percent MVC unless explicitly labelled as MVC.
- Insole line traces support only cautious comments about relative channel activity and its timing. The plots do not provide a foot map or a derived centre-of-pressure trajectory. Do not infer a precise foot region, joint angle, or muscle activation from them.
- Eye-gaze coordinates support gaze path, drift, and stability. They do not identify the shuttle, opponent, or gaze target.
- The racket and shuttle may not be visible. Do not infer racket-face orientation, shuttle outcome, or exact contact quality.
- Use cross-modal claims only when the relevant supplied modalities support them. Do not convert temporal coincidence into definite causation.
- Use the supplied phase label when supported; otherwise use `whole_stroke` or `unknown`.
- Speak directly to the player in `details`, `actionable_cue`, and practice advice.

 # REQUIRED JSON
Output only one valid JSON object containing the requested summary, coaching findings, and next actions. Do not use Markdown fences and do not include text before or after the JSON.

```
{
  "stroke_feedback_summary": "A concise player-facing overview with
      no skill label or expert comparison.",
```

```
  "coaching_findings": [
    {
      "finding_id": "finding_01",
      "name": "Short readable issue name",
      "phase": "preparation | backswing | forward_swing |
          follow_through | whole_stroke | unknown",
      "body_region": "Body region or sensor-supported functional
          area",
      "evidence_source": ["skeleton_video", "emg_upper"],
      "details": "Explain directly to the player what and where the
          issue is.",
      "observed_evidence": "State what is directly supported by the
          attached player data.",
      "why_it_matters": "Give a cautious badminton-coaching
          interpretation.",
      "actionable_cue": "One concrete correction the player can
          apply.",
      "suggested_drill": "One practical drill, or an empty string."
    }
  ],
  "what_to_work_on_next": [
    {
      "cue": "Highest-priority action",
      "why_it_matters": "Why this is the priority",
      "suggested_drill": "A practical drill"
    }
  ]
}
```

 # OUTPUT INSTRUCTIONS

1. Return at most five coaching findings and at most three priority actions.
2. Do not infer or output relationships, causal chains, knowledge-graph links, or connections between findings.
3. Output JSON only.

 # ATTACHED VISUAL GUIDE
{visual_guide}

F.5 Prompt Used for Semantic Clustering over Coaching Findings

System Prompt: Conservative Semantic Equivalence Annotation

 # ROLE
You are a conservative semantic equivalence annotator for badminton coaching findings.

 # CONTEXT
Input findings:
{{SYSTEM_FINDINGS}}

Each finding is identified by a semantic_id. System identities are intentionally hidden, and adjudications are intentionally absent.

 # YOUR TASK
Group findings only when they describe the same underlying biomechanical or coaching error in the same stroke.

 # CRITICAL CONSTRAINTS - READ CAREFULLY

- Different wording alone is not enough to group findings.
- Keep distinct body regions, phases, causes, or error directions separate.
- Do NOT judge whether a finding is correct, clinically valid, or well written.
- Do NOT use adjudications; they are intentionally absent.

- Every input `semantic_id` must appear exactly once.
- Never invent IDs.
- A singleton cluster is preferred when equivalence is uncertain.

 # OUTPUT FORMAT
Return JSON only.

```
{
  "clusters": [
    {
      "cluster_id": "cluster_001",
      "canonical_name": "short neutral error name",
      "member_ids": ["exact semantic_id"],
      "rationale": "brief reason these members are semantically
          equivalent"
    }
  ]
}
```

G Implementation Details of the Two-Level Knowledge Graph

The two knowledge graphs are drafted by a reasoning LLM (ChatGPT-o3) under the structured prompts at Appendices F.1 and F.2. The prompts detail the stroke phases, sensor modalities, body-part definitions, coaching terminology, and the required output graph schema. Three badminton experts then review every drafted coaching errors, their explanations, and relationship with body parts and other errors. Corrections include adjusting wordings, addition and removal of rules, changing relationships, etc. The verified graphs are remain frozen and stored as a lookup table. Table 14 provides the statistics of the two KGs among backhand and forehand strokes.

Graph	Statistic	Backhand	Forehand
Level 1	Location nodes $\lvert\mathcal{V}_{\text{loc}}\rvert$	48	48
	Rule nodes $\lvert\mathcal{V}_{\text{rule}}\rvert$	92	72
	Location-Rule edges $\lvert\mathcal{E}_1\rvert$	126	116
	Rules per location node (mean$^{\pm\text{SD}}$)	$2.33^{\pm 2.05}$	$2.42^{\pm 1.65}$
	Locations per rule node (mean$^{\pm\text{SD}}$)	$1.65^{\pm 0.81}$	$1.61^{\pm 0.62}$
	Location cells with ≥ 1 rule (%)	85.0	86.7
Level 2	Issue nodes $\lvert\mathcal{V}_2\rvert$	92	72
	Causal edges	130	70
	Correlational edges	52	31
	Root-cause nodes (no incoming causal edge)	9	22
	Chain length (mean / max)	2.81 / 4	3.32 / 6

Table 14: Statistics of the two-level knowledge graph for backhand and forehand strokes.

G.1 Level 1 KG

The Level 1 KG $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ is a bipartite graph connecting anomaly-location nodes $\mathcal{V}_{\text{loc}}$ and coaching-rule nodes $\mathcal{V}_{\text{rule}}$. An edge between a location node and a rule node means that an anomaly detected at that location makes the rule a possible biomechanical error the player has. Fig. 8 shows an induced subgraph of the Level 1 KG. As can be seen, we divide coaching rules by phases and link each rule to the body parts or sensors they relate. For example, from the figure, the Right Arm node reaches the arm- and elbow-related rules across preparation, backswing and forward-swing phases. In the graph, we represent each rule with an ID for easier downstream usage. In our coaching pipeline, given the anomaly locations flagged by TSAD, we retrieve the coaching rules from the graph, and ready for the deterministic rule evaluations.

G.2 Rule-Based Evaluators

Each coaching rule, extracted from the coaching manuals (Badminton World Federation, 2013; 2017) and established sports science literature, $r \in \mathcal{V}_{\text{rule}}$ is implemented as a deterministic evaluator with three components: (i) a measurement function f_r computed from named sensor channels over the phase window W_{phase}; (ii) a violation condition that compares f_r against thresholds calibrated as empirical quantiles $Q_p(\cdot)$ of the corresponding measurement on the expert validation set $\mathcal{D}_{\text{exp}}^{\text{val}}$, or against a built-in manual threshold, where the percentile $p = 95$ is fixed per rule at construction time and recorded in the catalog. The following are two examples of rule-evaluators in preparation phase with their measurement function.

Knee not bending enough. For each frame t, let $\theta_L(t) = \angle(\text{left hip}, \text{left knee}, \text{left ankle})$ and define $\theta_R(t)$ analogously for the right leg. The measurement is

$$f_r = \min_{t \in W_{\text{prep}}} \frac{\theta_L(t) + \theta_R(t)}{2}, \tag{2}$$

in degrees. Because these are internal joint angles, values near 180° correspond to straighter knees and smaller values to greater bending, so the test is two-sided: the rule flags a violation when $f_r < \tau_r^-$ (excessive bending) or $f_r > \tau_r^+$ (insufficient bending). The percentile calibration is $\tau_r^- = Q_{100-p}(f_r)$ and $\tau_r^+ = Q_p(f_r)$.

Elevated relative forearm activation. Let $e_c(t)$ be the rectified, low-pass-filtered envelope of channel c of EMG signals from the racket lower arm, and let $\widetilde{e}_c(t)$ denote its normalized version under the default stroke-peak scheme, in which each channel is divided by its maximum envelope value over the stroke. Using the default aggregate channel group, the measurement is

$$f_r = \frac{1}{|W_{\text{prep}}|} \sum_{t \in W_{\text{prep}}} \frac{1}{C_{\text{lower}}} \sum_{c=1}^{C_{\text{lower}}} \widetilde{e}_c(t). \tag{3}$$

The violation condition is one-sided, $f_r > \tau_r$ with $\tau_r = Q_p(f_r)$, so a violation indicates elevated forearm activation during preparation relative to the expert calibration distribution.

G.3 Level 2 KG

The level-2 KG $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$ connects verified coaching issues to each other via two types of edges: $\xrightarrow{\text{causes}}$ for a direct biomechanical causal relationship, and $\xrightarrow{\text{correlates with}}$ for a co-occurrence that is empirically associated but not unidirectional. In BIC's pipeline, after gathering verified errors from $\mathcal{G}_1$, we traverse the edges in $\mathcal{G}_2$ to get a root-case-issue explanatory chain. This helps to explain to players about their error in a traceable and rule-grounded manner. Fig. 9 illustrates one chain that spans the entire stroke. A root cause in the preparation phase propagates through backswing, forward-swing and contact.

H Visualizations

H.1 Sample Inputs to MLLM for Error Discovery

Figure 10 illustrates the multimodal visual plots provided to the MLLM. It includes phase-aligned skeleton video, the eye gaze trajectory, and the insole pressure distribution over time (where red boxes highlight TSAD-detected anomalies and purple marks the contact frame). For the EMG modality, we display the lower arm channels as an example; the other EMG groups follow the same template, with red shaded regions indicating the anomaly regions identified by the TSAD.

H.2 Reconstruction Result between beginners and experts

Fig. 11 plots the group-mean frame-wise anomaly scores generated by our TSAD model over expert and beginner backhand strokes. As can be seen, the beginner curve generally lies above the expert curve over the whole stroke. The two arm-EMG panels are the only exception, where expert strokes

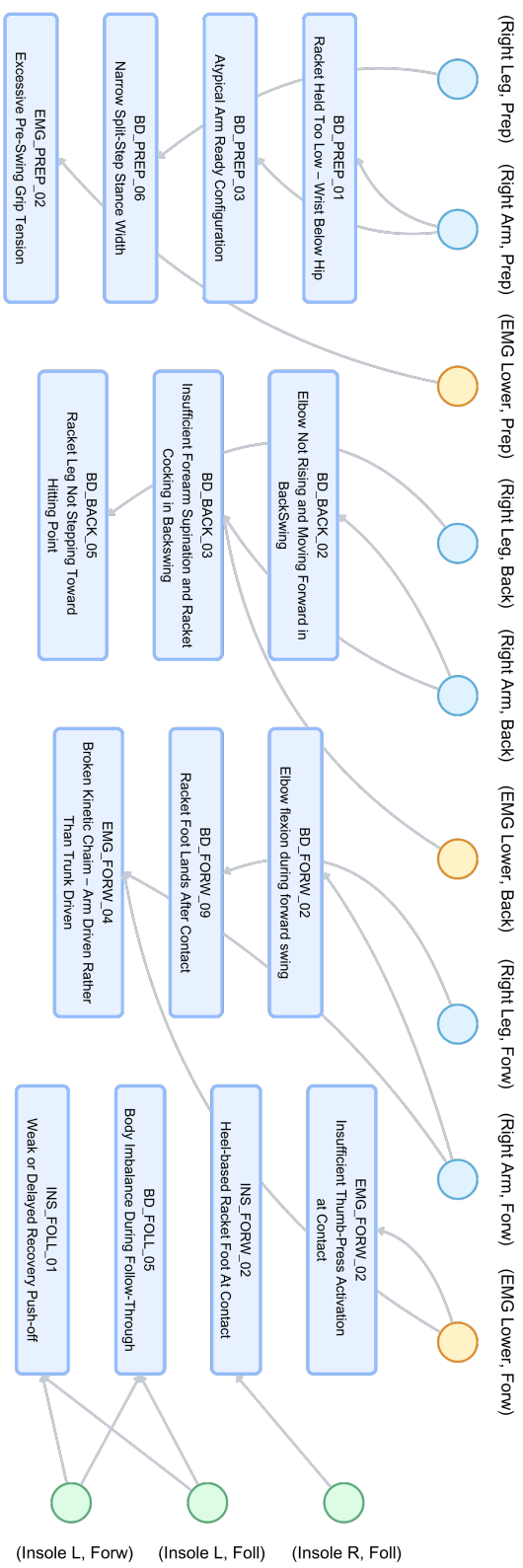


Figure 8: Excerpt of the Level-1 KG. Circular nodes are anomaly locations (body part or sensor modality; blue: skeleton, orange: EMG, green: insole pressure); rectangular nodes are coaching rules, labeled by rule ID and phase (PREP: preparation, BACK: backswing, FORW: forward swing, CONT: contact, FOLL: follow-through). A directed edge from a location to a rule marks the rule as biomechanically relevant to deviations at that location. At inference, an anomaly observation maps to its location node and retrieves exactly the rules reachable from it.

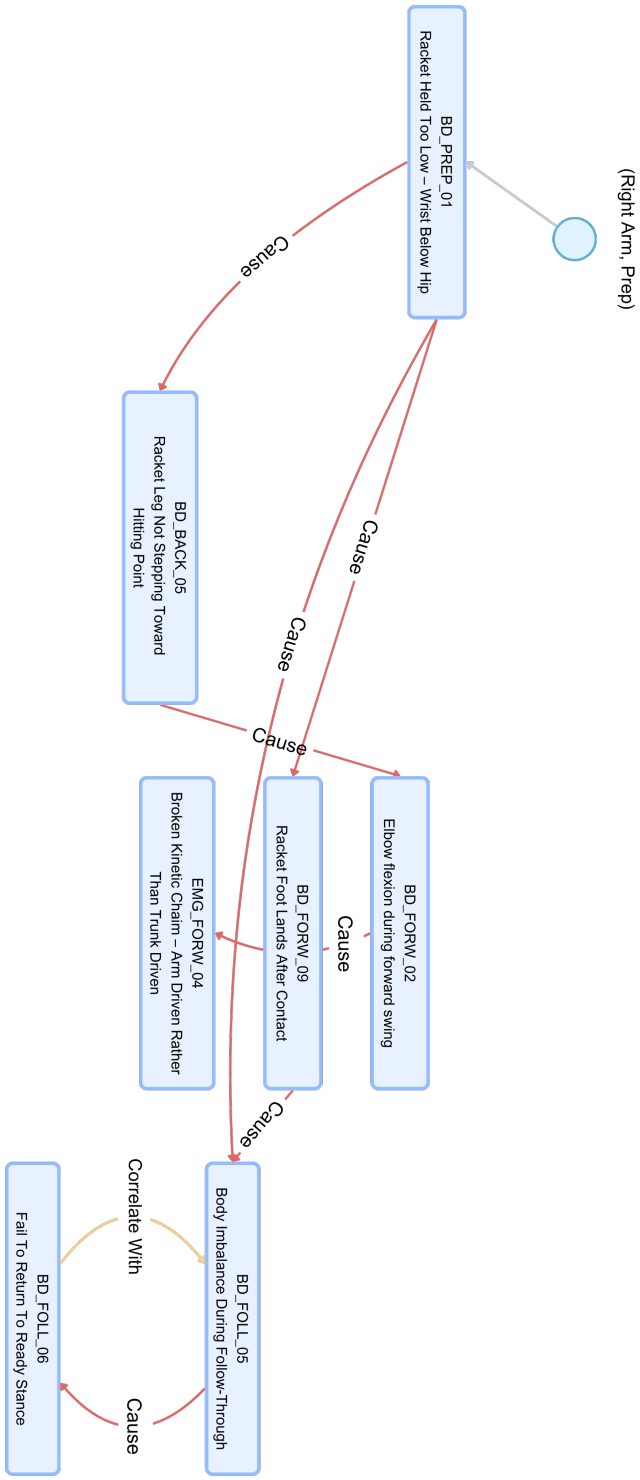


Figure 9: Excerpt of the Level-2 KG linking coaching issues across stroke phase. Red arrows are causal edges and yellow arrows are correlational edges. The highlighted chain shows a root cause in preparation (low racket-leg load) propagates causally through weight transf and racket-foot loading at contact, while a correlational branch connect backswing trunk rotation to forearm activation in the forward string. With this chain, a player can trace their error and understand the root causes.

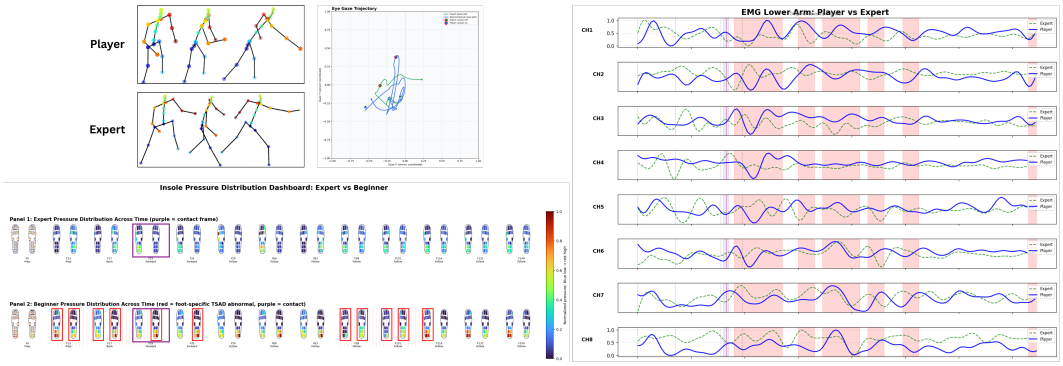


Figure 10: Sample multimodal visual inputs provided to the MLLM for error discovery.

show slightly higher reconstruction error around the early activation peak. This shows that a single modality can be non-diagnostic on its own, while the fused score separates the two classes cleanly at every phase, is the behavior the cross-modal fusion of our detector is designed to produce.

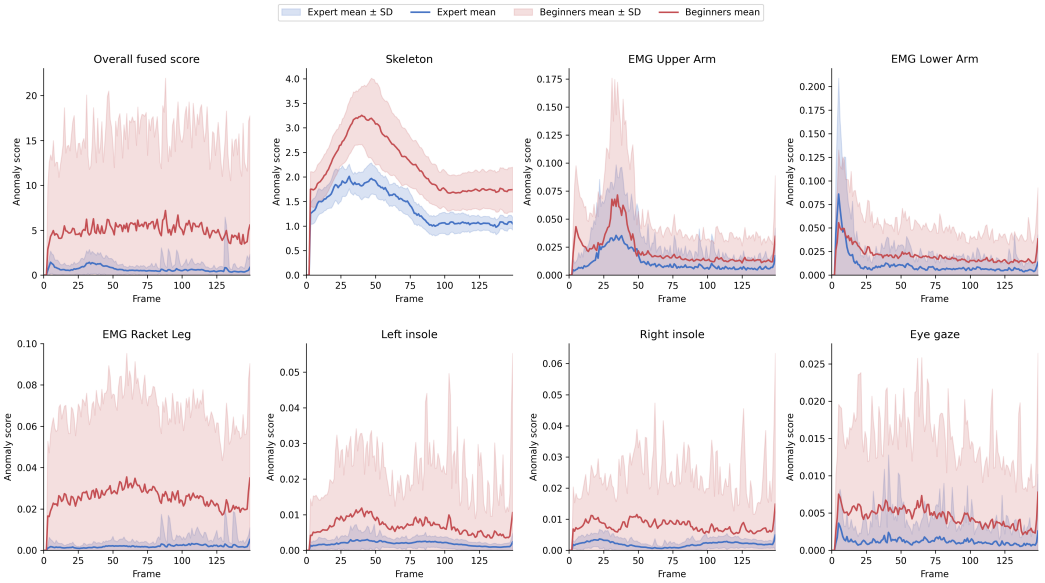


Figure 11: Group-mean frame-wise anomaly scores for expert (blue) and beginner (red) strokes, with ± SD bands over strokes, averaged over cross-validation folds. Panels show the fused score (top left) and the seven modality-specific scores.

H.3 Sample Coaching Reports across Methods

We list a sample of the complete output report of all four systems in Fig. 12 (MLLM with skeleton only), 13 (MLLM with skeleton and sensors), 14 (ChainBMD), and 15 (BIC).

MLLM (Skeleton only)

Stroke feedback summary
Your backhand-drive action is compact and you complete a full swing, but the early preparation is limited in trunk turn and the follow-through leaves you off-balance. Improving how much you rotate and how you recover after contact will help you generate a faster, more stable shot and be ready for the next rally.

Measured coaching findings
- Limited trunk rotation in backswing
- Passive non-racket arm
- Forward lean causing loss of balance
- Delayed weight transfer

What to work on next
- Add a bigger shoulder and hip turn during your backswing.
  More rotation stores energy and positions the racket for a faster, sharper drive without extra arm effort.
  Practice: Slow-motion shadow swings: turn shoulders until your back almost faces the net, pause, then swing through 15× per set.
- Use the non-racket arm to balance your swing.
  Active counter-balance keeps your trunk stable, improving accuracy and readiness for the next shot.
  Practice: Practise mirror swings focusing on reaching the free arm forward as the racket moves back and then retracting it on follow-through.
- Finish the swing with an upright chest and quick recovery step.
  Maintaining balance after contact shortens recovery time, allowing you to cover the court faster.
  Practice: Wall drive-recover sets: hit 10 continuous drives against a wall, each time returning to a balanced ready stance before the next feed.

Figure 12: Sample coaching report from MLLM(Skeleton only).

MLLM (Skeleton + Sensors)

Stroke feedback summary
Your swing shows clear phase sequencing, but three areas limit racket speed and control: limited upper-body rotation before contact, late wrist/forearm firing, and weight remaining on the rear (right) foot through impact. Tidying these points will help you generate a flatter, faster backhand drive and regain court balance sooner.

Measured coaching findings
- Minimal trunk rotation before forward swing
- Delayed wrist/forearm activation
- Weight remains on rear foot at contact
- Head and gaze lift immediately after contact

What to work on next
- Turn the upper body more in the backswing, then snap the wrist right at contact.
  A deeper coil plus timely wrist release will add racket-head speed and a flatter, faster drive without extra arm effort.
  Practice: Mirror shadow swings: mark shoulder turn angles on floor, rehearse turn-coil-snap sequence slowly, then at full speed, filming for feedback.
- Shift weight onto the racket-side foot before striking.
  An early weight transfer stabilises your base and frees the arm for a cleaner, straighter drive, improving accuracy and recovery.
  Practice: Front-foot landing drill: start in split-step, push off, land on racket-side foot, freeze, check balance, then hit a shuttle fed by a partner.
- Delay head lift until after shuttle contact.
  Maintaining visual focus on contact improves timing and reduces mishits.
  Practice: Use slow-feed drives; consciously keep gaze on shuttle until you hear it leave the strings, gradually increase speed while keeping the habit.

Figure 13: Sample coaching report from MLLM(Skeleton + Sensors).

ChainBMD

Stroke feedback summary

The athlete is currently at the intermediate level, with some control and continuity but room for improvement in overall coordination. Their kinetic chain shows minor deviations in upper-arm, forearm, core, and thigh loading, with the wrist showing a significant lag in acceleration. The timing and rhythm reveal a notable lead in left-foot support, left thigh, and trunk/core, which may contribute to imbalances and inefficiencies in the backswing. This suggests a potential over-reliance on the left side, impacting overall power transfer and control.

Measured coaching findings

- Insufficient upper-arm loading. Focus on racket preparation and shoulder external rotation.
- Insufficient force transfer through the forearm. Train whipping mechanics and forearm strength.
- Weak core kinetic chain. Reinforce trunk rotation and core stability.
- Insufficient lower-limb drive. Train thigh strength and hip-knee coordination.
- Insufficient terminal wrist acceleration. Train impact-phase wrist explosiveness and control.

What to work on next

- Foot alignment drill: focus on correcting the left foot's alignment to ensure proper support and sequencing.
  Performance goal: achieve neutral alignment with minimal deviation during baseline and backswing.
- Right-side strength and conditioning: strengthen the right leg and core to balance the left side.
  Performance goal: maintain a 1:1 ratio of left to right strength during baseline and backswing.
- Wrist drive and power transfer drill: focus on generating a stronger and more consistent wrist drive.
  Performance goal: increase wrist acceleration by 10% over 4 weeks while maintaining control.

Figure 14: Sample coaching report from ChainBMD.

BIC (ours)

Stroke feedback summary

During your backswing you stay almost upright instead of hinging forward from the hips. Keep your chest angled a little more over your front knee when you draw the racket arm back. That forward lean helps you load your legs and core so you can drive into the shuttle on the next phase. During your forward-swing you keep your racket-arm elbow noticeably bent instead of letting it straighten toward full extension just before contact. Work on allowing the arm to lengthen so the hand finishes farther away from your shoulder. This extra reach lets you meet the shuttle earlier and maintain a smoother acceleration path. From your very first move you hold the hitting arm's forearm and grip muscles at a high, steady effort instead of letting them stay relaxed and then spike just before contact. That early tension shows up during preparation and carries right through the backswing, so the racket never gets the free-swinging acceleration you want. Think of keeping the hand loose until you start the forward swing, then squeeze at the last instant. Right after you finish the hit, your eyes sweep well past the contact area–first jumping out to your right and slightly upward, then diving low–rather than settling quickly. This extra scan happens in the follow-through and can make it harder for you to keep your head steady and re-orient for the next shot. Lift your elbow higher during the take-back so it lines up with your shoulder before the forward swing. Hinge at the hips so your chest is over your front knee as you take the racket back. Let the hitting arm straighten toward the shuttle just before contact. Keep your grip loose during the set-up, squeeze only at the moment of impact. After contact, let your gaze settle instead of scanning–keep it calmly centred for a split-second before tracking the next play. Shift more body weight onto your left (non-racket) foot at the end of the backswing.

Measured coaching findings

- Elbow not raised and moved forward in backswing
- Reduce backswing magnitude
- Low left leg load in backswing
- Racket foot landing after contact, failed to step into the shot
- Heel-weighted racket foot at contact
- Upright trunk: insufficient forward body lean
- Hitting arm stops short: elbow remains too flexed at contact
- Excessive post-contact gaze drift

What to work on next

- Lift your elbow higher during the take-back so it lines up with your shoulder before the forward swing.
  A higher elbow creates a stronger, flatter hitting platform, helping you drive the shuttle rather than lifting it.
  Practice: Shadow the backhand drive in front of a mirror, focusing on raising the elbow to shoulder height before swinging forward.
- Hinge at the hips so your chest is over your front knee as you take the racket back.
  A forward-leaning posture pre-loads your legs and core, giving you better balance and energy transfer when you push into the swing.
  Practice: Practise shadow swings starting from a split-step, exaggerating a slight forward bend and feeling your weight on the front leg before initiating the stroke.
- Keep your grip loose during the set-up, squeeze only at the moment of impact.
  A relaxed forearm lets the racket accelerate smoothly and reduces unnecessary fatigue; tightening only at contact delivers power without slowing the swing.
  Practice: Shadow-swing with the racket held lightly, rehearse adding the squeeze only as the racket snaps through the imaginary shuttle.
- Shift your weight onto your left (non-racket) foot at the end of the backswing.
  Firmly loading your trailing leg creates an anchor to push off laterally into your lunge, generating a smoother and more explosive stroke.
  Practice: Perform slow-motion shadow swings, focusing on feeling at least two-thirds of your body weight pressing into your left midfoot/heel before driving off into the stroke.

Figure 15: Sample coaching report from BIC.

I TRAINING CONFIGURATIONS

I.1 EVALUATION SPLITS

Table 15 details the four subject-disjoint evaluation iterations. Let Sub11, Sub14, Sub19, and Sub24 denote the four distinct expert subjects in the MultiSenseBadminton dataset (Seong et al., 2024b) Each expert appears in the evaluation set exactly once, ensuring balanced and unbiased assessment across all participants.

Iteration	Training	Calibration	Evaluation
1	Sub11, Sub14	Sub19	Sub24
2	Sub14, Sub19	Sub24	Sub11
3	Sub19, Sub24	Sub11	Sub14
4	Sub24, Sub11	Sub14	Sub19

Table 15: Four subject-disjoint evaluation iterations. Each iteration uses two experts for training, one for threshold calibration, and one for evaluation.

I.2 HYPERPARAMETERS

For the training objectives mentioned in Section 3.3, we set $\lambda_1 = 1$, $\lambda_2 = 6$, $\lambda_3 = 0.5$, and $\lambda_4 = 3$. The two higher weights $\lambda_2 = 6$ and $\lambda_5 = 3$ specifically compensate for the averaging of losses across the multiple sensor modalities, keeping the total sensor loss magnitude comparable to the skeleton loss, while the reconstruction and consistency losses serve as secondary and tertiary stabilizing objectives. We train our latent diffusion model using Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with a learning rate of 10^{-3} for the conditioning projection layers and 10^{-4} for all denoisers. We use a batch size of 64 per GPU and an early stopping with patience of 10 epochs. The pretrained encoders and decoders are remain frozen throughout the crossmodal TSAD training. All denoisers are trained with $N = 10$ diffusion steps using a cosine variance scheduler (Nichol & Dhariwal, 2021). All experiments are performed on four Nvidia RTX A6000 GPUs with 48GB memory.

J ANOMALY SCORES AGGREGATION FOR EVALUATION

To compare with baselines, we aggregate each modality's anomaly vector into a single scalar representing each stroke. All sensing modalities are multi-channel time-series data whose reconstruction errors exhibit different scales. Different modalities also have different number of channels, so direct aggregation can lead to imbalance and domination by a few modalities. To obtain a balanced anomaly score for each stroke, we proceed in three steps.

J.1 CHANNEL NORMALIZATION

The reconstruction error of each channel is normalized by an expert-calibrated reference scale. Specifically, for each modality m and channel c, we divide by the q-quantile of the corresponding expert reconstruction-error distribution:

$$\tilde{e}_{t,c}^{(m)} = \frac{e_{t,c}^{(m)}}{Q_q\left(\left\{e_{t,c}^{(m)} \mid \mathbf{X} \in \mathcal{D}_{\text{exp}}^{\text{val}}\right\}\right)}, \tag{4}$$

where $Q_q(\cdot)$ denotes the empirical q-quantile ($q = 90$).

J.2 MODALITY AGGREGATION

Next, the normalized channel errors are averaged within each modality,

$$a_t^{(m)} = \frac{1}{C_m} \sum_{c=1}^{C_m} \tilde{e}_{t,c}^{(m)}, \tag{5}$$

where C_m is the number of channels in modality m. The seven modality-level errors are then averaged to produce one fused frame-level anomaly vector,

$$f_t = \frac{1}{M} \sum_{m=1}^{M} a_t^{(m)}. \tag{6}$$

J.3 Stroke-level Score

Finally, the stroke-level anomaly score for stroke $\mathbf{X}_i$ is obtained by aggregating the frame-level trajectory $f = (f_1, \ldots, f_T)$:

$$S_i^{\mathrm{TSAD}} = g(f). \tag{7}$$

In our evaluation (Section 4.2), we employ the **Top-K% mean** aggregation function:

$$g_{\text{top-k}}(f) = \frac{1}{|\text{top-k}(f)|} \sum_{t \in \text{top-K}(f)} f_t, \tag{8}$$

which averages the highest $K\%$ of frame-level scores. We select this strategy over simple mean or max pooling because it optimally balances sensitivity and robustness. Unlike simple mean pooling, which dilutes localized biomechanical deviations, and max pooling, which is overly sensitive to isolated single-frame noise spikes, the Top-$K\%$ mean captures sustained anomalous patterns while effectively filtering out short-lived artifacts.

K Implementation Details of Qualitative Evaluation

K.1 Baselines Setup

Section 4.3 compares BIC with three baselines: (i) ChainBMD (Fang et al., 2026), which works on the same MultiSenseBadminton dataset by using kinetic chain graph for skill-level classification and generating coaching clues with an MLLM; and (ii) a reasoning generalist MLLM (ChatGPT-o3), which serves as a lower bound on how a general-purpose model performs at badminton coaching. For the MLLM baseline we benchmark two variants, one supplied only with the skeletal video and one that additionally receives time-series plots of the sensor streams. We do not supply real player footage to the MLLM baselines because the dataset provide no footage synchronized with the sensor streams. Thus, each MLLM baseline receives a rendered skeletal animation of the stroke in which every frame displays its frame index, together with the sensor time-series plots. Therefore, the MLLM can relate events across modalities at the frame level. Fig. 16 shows the sample inputs. This also serve as a fair comparison, since none of the compared methods uses real video as input. The prompts used for both MLLM variants are listed in F.4.

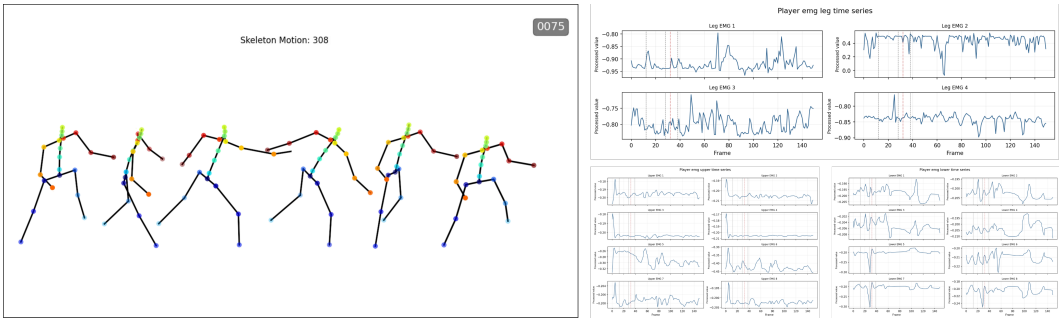


Figure 16: Inputs supplied to the MLLM baselines for one stroke. **Left:** the skeletal motion is supplied as a rendered video in which each frame shows a single pose with its frame index in the top-right corner. For compactness this panel overlays several sampled poses of the same stroke. **Right:** examples of the sensor time-series plots supplied to the multimodal variant. Here we show the three EMG signals; insole-pressure and gaze plots are supplied in the same format.

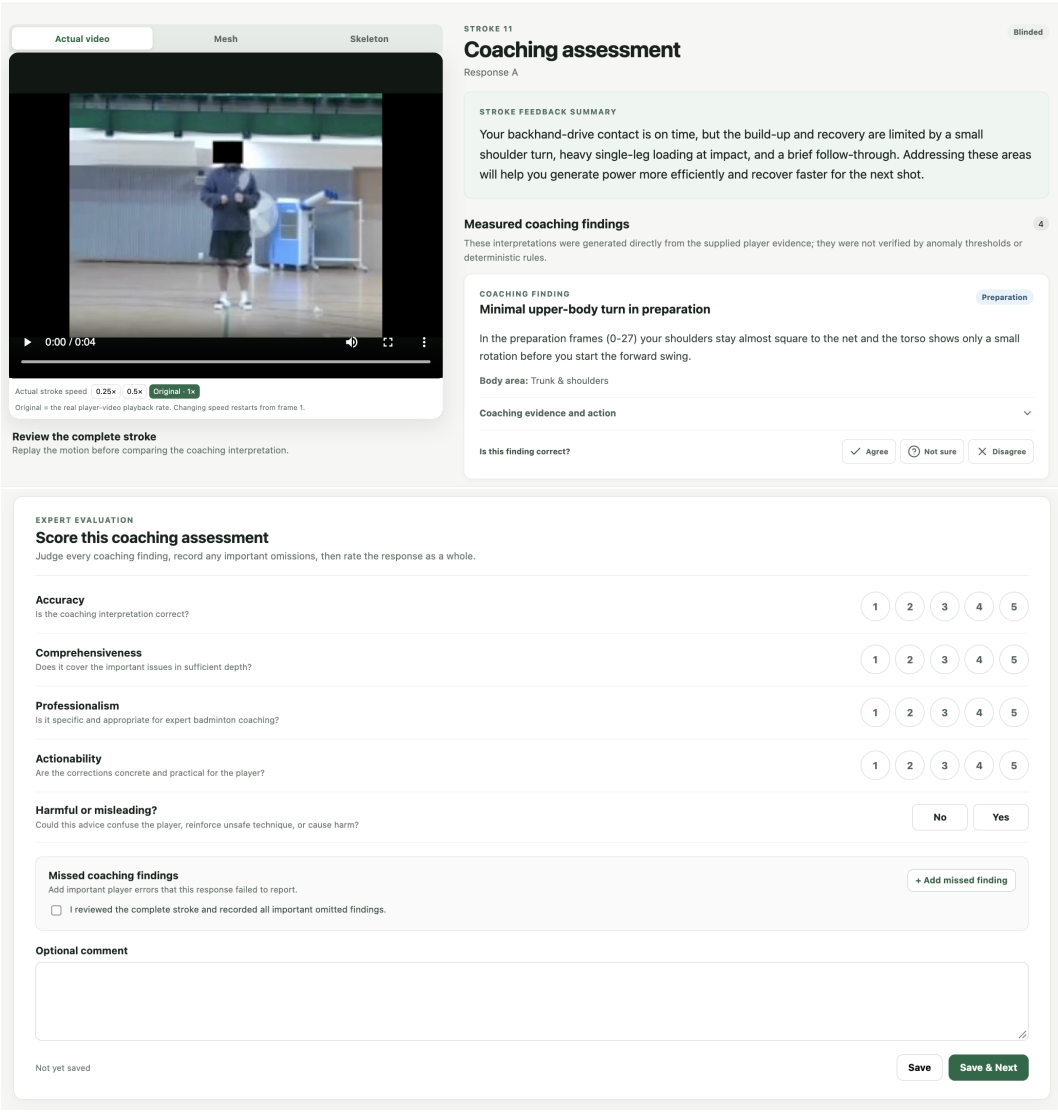


Figure 17: User Interface of how experts adjudicate each stroke.

K.2 User Interface

Fig. 17 shows the interface of website where experts review each stroke. For each stroke, we provide the manually extracted player footage, skeletal video, and the coaching report generated by each method. After anonymizing and randomizing the samples, for each findings, the experts are asked to choose either agree, not sure, or disagree. Then they are asked to rate the quality of outputs across the four domains, and confirm if there are harmful or misleadning informations. Reviewers are also encouraged to present missing coaching errors and give comments about the results.

L Ablation Studies

Each ablation below changes one design choice of BIC and evaluate on the backhand strokes.

Method	AUROC
Modality-specific Denoiser	$\mathbf{0.886^{\pm 0.089}}$
A Global Denoiser	$0.650^{\pm 0.071}$

Table 16: Effect of using one global or modality-specific denoiser.

L.1 Modality-specific Denoiser

Table 16 explores the effect of using a single global denoiser instead of modality-specific denoiser (Section 3.3). The results show that modeling all modalities in a single distribution degrades its performance in skill-level discrimination, indicating that separate modeling is more suitable in handling multimodal time series.

L.2 Kinematic Graph

Skeleton Encoder & Denoiser	AUROC
ST-GCNN	$\mathbf{0.886^{\pm 0.089}}$
TCN	$0.733^{\pm 0.044}$

Table 17: Effect of structuring skeleton time-series inputs as graph versus as a flattened vector.

One of our key component is structuring the multimodal dataset as a graph (Section 3.3). This ablation investigates the importance of using this graph by replacing the ST-GCNN (Yan et al., 2018) layers in skeleton encoder and denoiser to TCN (Bai et al., 2018) layers, which flatten the inputs. As shown in Table 17, using a graph structure achieves better skill-level discrimination. This indicates that the explicit bone topology, rather than temporal mixing alone, is preferred.

L.3 Stage 3: MLLM Error Discovery

The stage 3 of our pipeline (Section 3.5) leverages an MLLM to discover additional coaching errors that may not be explicitly captured by the rule-based KG verification in Stage 2 (Section 3.4). To evaluate the contribution of this stage, we compare the full pipeline against a variant that terminates after Stage 2, relying solely on strictly KG-verified findings.

Method	P ↑	R ↑	Excl. R ↑	F1 ↑
Original	**0.893**	**0.681**	**0.496**	**0.760**
W/o MLLM Error Discovery	0.870	0.532	0.415	0.652

Table 18: Ablation study on the MLLM Error Discovery stage.

As shown in Table 18, removing the MLLM Error Discovery stage causes a notable drop in Recall and Exclusive Recall, while Precision remains relatively high. This indicates that while the KG verification ensures high correctness, the MLLM exploration is crucial for uncovering biomechanical faults that strict rule-checking might miss. Consequently, this stage significantly improves the overall comprehensiveness and coverage of the final coaching report.