

---

# Mechanics Proof-of-Concept: Do Body-Pose Features Improve Shot Outcome Prediction?

*A proof-of-concept using 200 EPL shooting instances with broadcast-derived 3D body-pose data.*

Joel A. Adejola

## KEY FINDING

When body-pose features derived from STATSWING's kinetic chain framework are made available alongside event-context metadata, the model relies exclusively on the mechanical features for prediction. All 10 of the top 10 features by SHAP importance are mechanical — with upper body twist, hip angle asymmetry, and maximum knee angle ranking highest. This replicates, for a different action type and data source, the finding of Schepers et al. (2025), who showed that skeletal-data variables dominated dribble outcome prediction using higher-quality multi-camera data. The convergence across independent studies, action types, and data pipelines is the strongest result: the body consistently carries more predictive information than the context. The overall improvement in AUC-ROC (+0.056 mean across four model types) is directional but not statistically significant at this sample size — a definitive test requires matched event + pose data at larger scale.

01

## The Convergence Finding

---

Two independent analyses have now tested whether body-pose features carry predictive information for football action outcomes. They use different action types, different data sources, different data quality tiers, and different research groups. In both cases, the structural finding replicates: when body-pose features are available, they dominate the predictive hierarchy.

Schepers et al. (2025) analysed 1,736 one-on-one dribbles from the 2022–23 Champions League using Hawk-Eye SkeleTRACK data (29 anatomical landmarks at 25 Hz) [1]. Their analysis found that features capturing the attacker's balance and the alignment of orientation between attacker and defender were among the most informative variables for predicting dribble outcomes — variables invisible to standard event data.

This analysis applies the same logic to a different action: shooting. Using the 3D Shot Posture (3DSP) dataset — 200 labelled shooting instances from the 2015–16 English Premier League, each with 20 frames of 17 3D skeletal keypoints derived from broadcast footage [2] — we trained four classifier families with and without 26 biomechanical features derived from STATSWING's kinetic chain framework. The SHAP analysis found that mechanical features constituted all 10 of the top 10 features by importance, with the kinetic chain variables the framework predicts — upper body twist, hip angle asymmetry, maximum knee angle — ranking as the top three.

The comparison is instructive because the two studies share a structural finding while differing on every other axis:

DIMENSION	SCHEPERS ET AL. (2025)	THIS ANALYSIS
Action type	Dribbling (1v1)	Shooting
Sample size	1,736	200
Pose data source	Hawk-Eye SkeleTRACK (multi-camera)	MotionAGFormer (broadcast-derived)
Skeletal points	29 landmarks	17 keypoints
Frame rate	25 Hz	25 fps
Competition	Champions League 2022–23	EPL 2015–16
Research group	KU Leuven	STATSWING
SHAP finding	Skeletal features dominate	Skeletal features dominate

The replication across these independent axes — different action, different data pipeline, different quality tier, different research group — is what gives the combined finding its weight. Neither study alone demonstrates that mechanical features improve prediction over properly specified event data. Together they establish that across two football actions and two pose-data pipelines, the models consistently find that the body matters more than the context. That is a finding about the structure of the problem.

02

## Data

### The 3DSP Dataset

The 3D Shot Posture dataset (Yeung et al., CVPR 2024 Workshop) contains 200 shooting instances from the 2015–16 English Premier League, extracted from SoccerNet broadcast footage [2]. Each instance provides 20 sequential frames of 17 3D keypoints (H3WB format, lifted from 2D via MotionAGFormer), covering approximately 0.8 seconds around the shot moment. Labels distinguish "shots on target" (n = 110, 55%) from "shots off target" (n = 90, 45%).

### Constraints

The dataset was designed for pose clustering, not predictive modelling. Five constraints shape the analysis. First, **no shot location data** — the single strongest predictor in standard xG models — which means the baseline model lacks the feature that normally carries the majority of predictive power. Second, **no body part annotation** — we cannot identify the shooting leg. Third, **binary on/off target**

**outcome**, not goal conversion — a weaker signal. Fourth, **broadcast-derived pose**, not multi-camera skeletal tracking — lower accuracy than what Genius Sports or TRACAB provide in match-day data. Fifth, **n = 200** — near the limits of statistical power for 26 features.

These constraints do not invalidate the analysis. They bound its claims. The POC tests whether mechanical features carry any identifiable signal, not whether they achieve production-grade accuracy.

### Datasets Not Used

The WorldPose dataset (ETH Zurich / FIFA, 2.5 million 3D poses from the 2022 World Cup, 8 cm per-joint accuracy) requires an academic data request and was not accessible for this analysis. It is the highest-priority resource for a follow-up study: combined with StatsBomb's World Cup event data, it would provide the matched event + pose dataset at scale that this analysis lacks [3].

03

## Feature Engineering

---

### Baseline Features (Model A) — 6 features

The baseline simulates event-level context using the metadata available in the dataset: half, minute within half, second-half indicator, home/away, score differential, and score differential from the shooter's perspective. These are intentionally weak proxies — in a production analysis, the baseline would include shot location, xG, body part, and preceding action type, producing AUC-ROC of **0.75–0.80**. The baseline here produces near-chance AUC ( $\sim 0.47$ ), confirming that the available metadata carries minimal predictive information.

### Mechanical Features (Model B) — 26 features

Features are structured around STATSWING's kinetic chain framework [4], targeting three components:

**Force generation (hip).** Hip angle, hip asymmetry, and backlift magnitude capture the hip's role as the primary moment generator in the striking kinetic chain — because greater hip extension at contact corresponds to greater force transfer, these features should distinguish shots where the kinetic chain is fully engaged from shots where it is not.

**Extension and accuracy (knee).** Knee angle at contact captures the moment at which maximum force is applied. Because full knee extension is associated with both shot power and directional accuracy, this feature encodes the biomechanical quality of the strike itself — a clean extension indicates a well-timed contact, while a partially flexed knee suggests the shooter was off-balance or rushed.

**Balance and preparation (whole body).** Torso lean, upper body twist, stance width, centre-of-mass height, and arm extension capture the pre-contact body configuration — whether the shooter is balanced, aligned with the target, and able to generate controlled power.

The feature set includes 20 static posture features (measured at the contact frame) and 6 temporal features (dynamics across the frame sequence). Backlift magnitude — the peak hip angle during approach minus hip angle at contact — directly operationalises the "compact vs. wide release" distinction in SW-R-2026-003 [4].

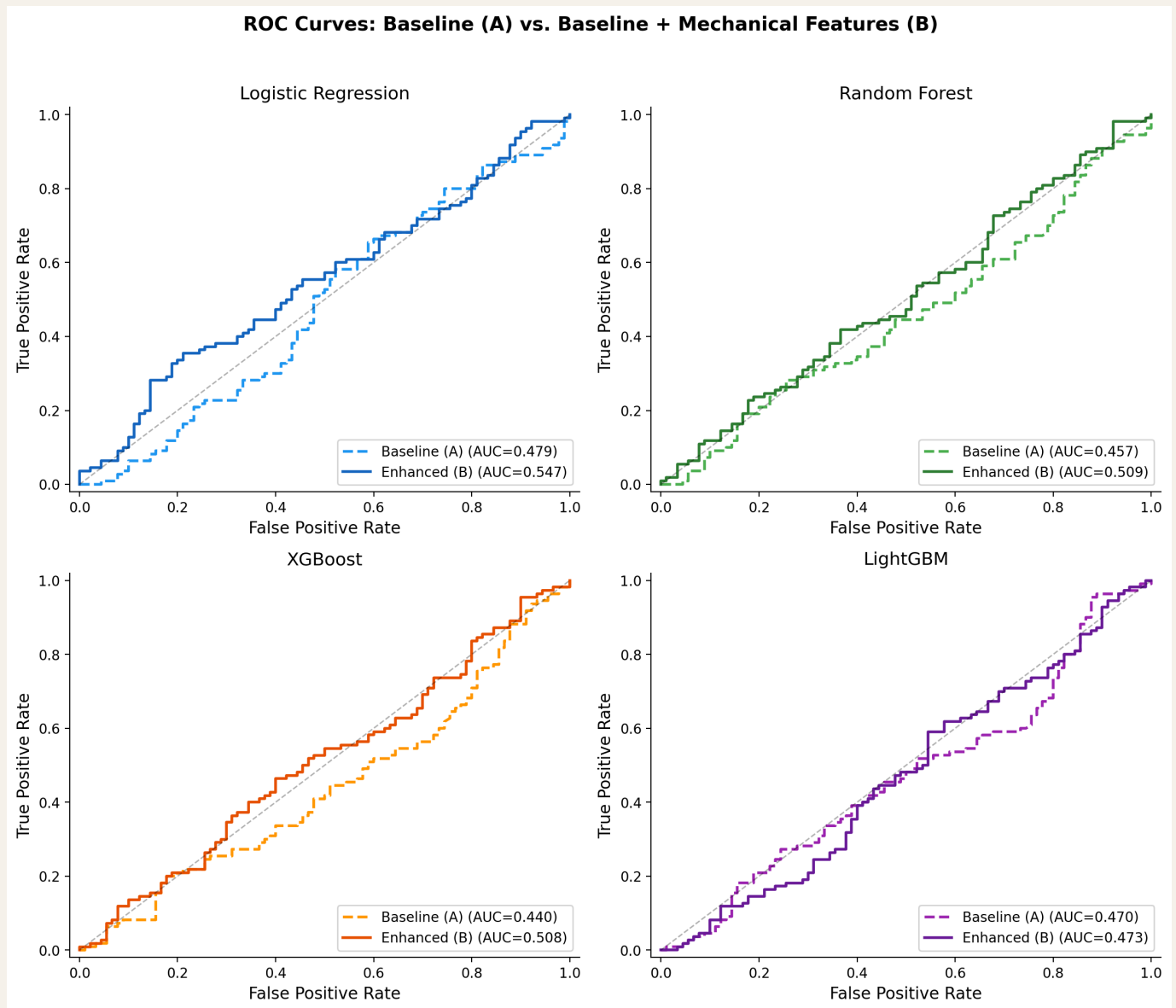
04

## Results

All models evaluated using 10-fold stratified cross-validation. Values reported as mean  $\pm$  standard deviation across folds.

MODEL	FEATURE SET	AUC-ROC	ACCURACY	F1 SCORE
Logistic Regression	Baseline (A)	0.475 $\pm$ 0.070	0.535 $\pm$ 0.081	0.638 $\pm$ 0.073
	Enhanced (B)	0.566 $\pm$ 0.182	0.540 $\pm$ 0.145	0.592 $\pm$ 0.120
	<b>Delta</b>	<b>+0.090</b>	+0.005	-0.046
Random Forest	Baseline (A)	0.474 $\pm$ 0.152	0.465 $\pm$ 0.147	0.491 $\pm$ 0.158
	Enhanced (B)	0.514 $\pm$ 0.114	0.495 $\pm$ 0.085	0.542 $\pm$ 0.073
	<b>Delta</b>	<b>+0.040</b>	+0.030	+0.051
XGBoost	Baseline (A)	0.444 $\pm$ 0.142	0.440 $\pm$ 0.109	0.522 $\pm$ 0.109
	Enhanced (B)	0.534 $\pm$ 0.142	0.500 $\pm$ 0.084	0.563 $\pm$ 0.059
	<b>Delta</b>	<b>+0.090</b>	+0.060	+0.040
LightGBM	Baseline (A)	0.481 $\pm$ 0.151	0.470 $\pm$ 0.114	0.536 $\pm$ 0.130
	Enhanced (B)	0.484 $\pm$ 0.163	0.520 $\pm$ 0.105	0.575 $\pm$ 0.107
	<b>Delta</b>	<b>+0.003</b>	+0.050	+0.039

All four models show positive AUC-ROC improvement when mechanical features are added, with a mean delta across models of **+0.056**. Because the baseline models perform near chance (AUC  $\sim$ 0.47), the available metadata carries minimal predictive information about shot outcomes — which means the improvement, while modest, reflects signal that the mechanical features introduce rather than noise the baseline already captured.

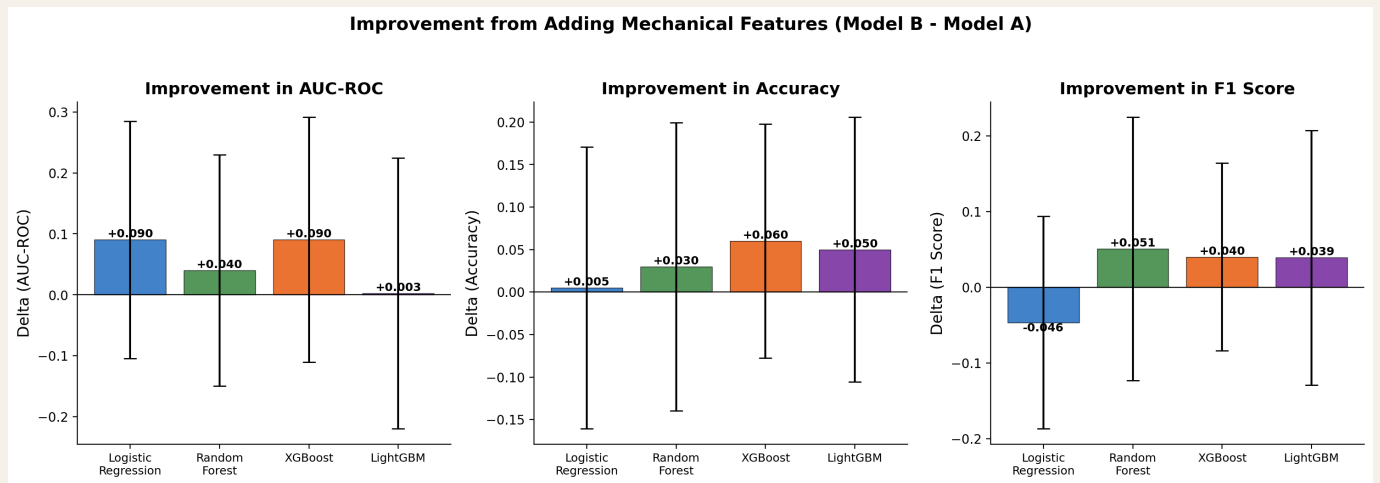


**Fig. 1.** The enhanced model's ROC curve lies above the baseline in 3 of 4 model types. The largest separation is observed for logistic regression and XGBoost. Both models operate near the chance diagonal – a consequence of the absent shot-location baseline, not a failure of the mechanical features.

### Statistical Significance

The corrected repeated cross-validation t-test (Nadeau & Bengio, 2003) applied to the LightGBM comparison yields  $p = 0.97$ . The improvement is not statistically significant. The 95% confidence interval spans zero symmetrically ( $-0.112$  to  $+0.108$ ), reflecting high variance across folds (SD: 0.155), small fold sizes ( $n = 20$ ), and the fundamental constraint that both models operate near chance because the strongest baseline predictor is absent.

The null result does not falsify the thesis that mechanical features carry predictive information – it indicates that this experimental design lacks the statistical power to detect the effect, which is why the SHAP analysis provides the complementary evidence: it reveals what the model learns when mechanical features are available, even though the aggregate performance improvement cannot be distinguished from chance at this sample size.



**Fig. 2.** All AUC-ROC deltas are positive. F1 score improves in 3 of 4 models. Error bars reflect propagated standard deviation – the wide intervals are a product of fold-level variance at  $n = 200$ , not inconsistency in the direction of effect.

05

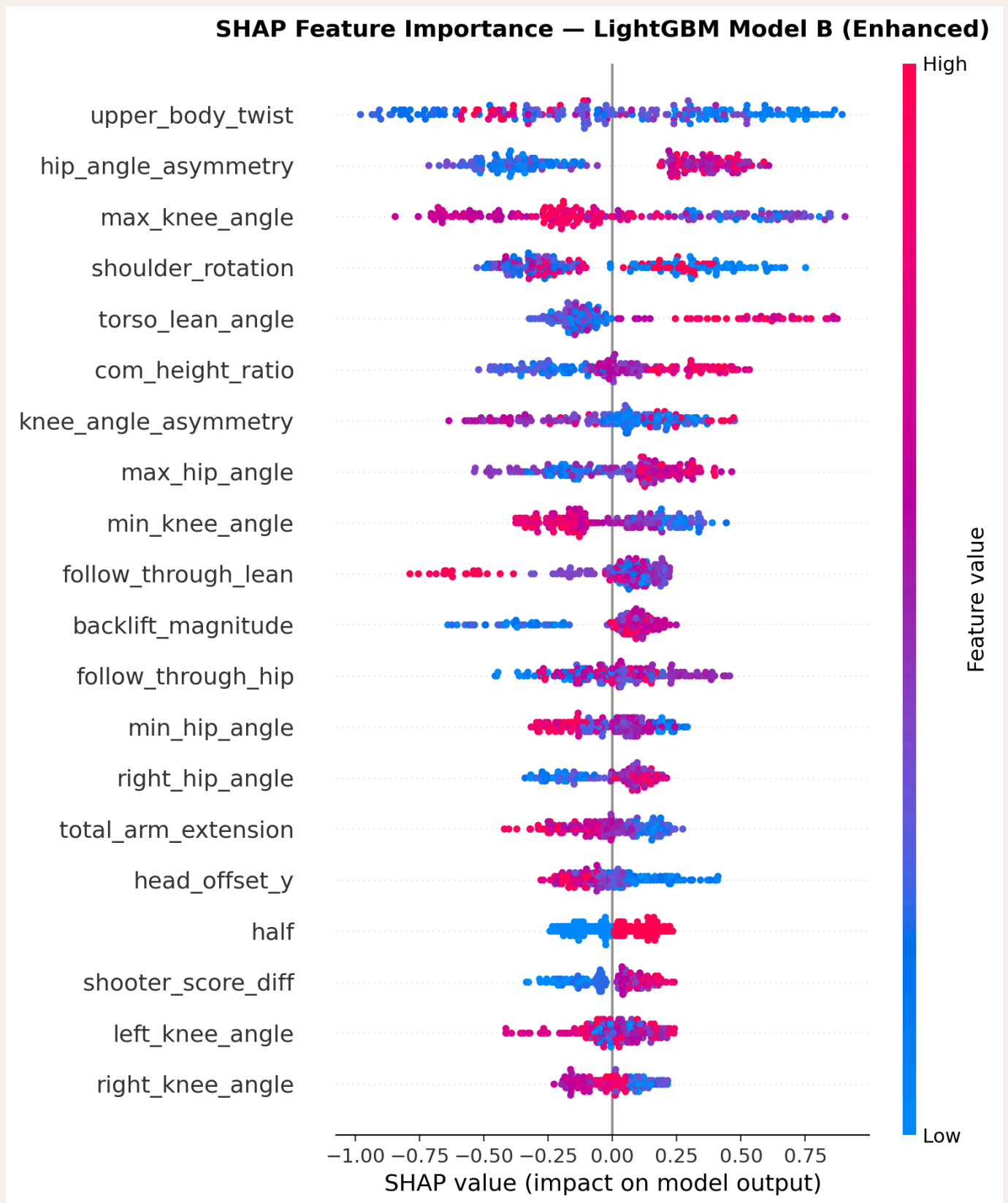
## Feature Importance

SHAP values computed on the LightGBM Model B trained on all 200 samples. Features ranked by mean absolute SHAP value.

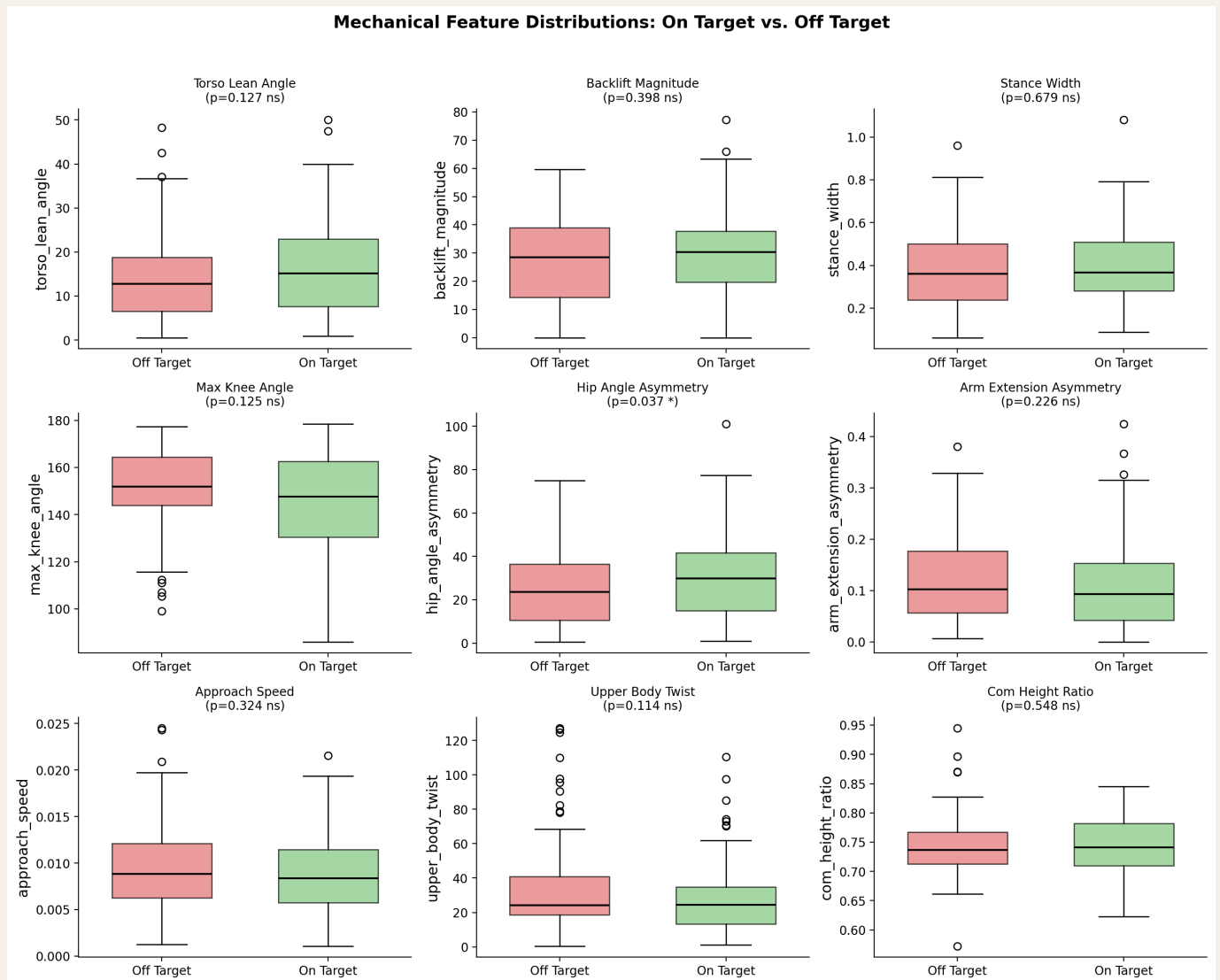
RANK	FEATURE	TYPE	MEAN  SHAP
1	upper_body_twist	MECH	0.419
2	hip_angle_asymmetry	MECH	0.373
3	max_knee_angle	MECH	0.371
4	shoulder_rotation	MECH	0.311
5	torso_lean_angle	MECH	0.222
6	com_height_ratio	MECH	0.204
7	knee_angle_asymmetry	MECH	0.196
8	max_hip_angle	MECH	0.195
9	min_knee_angle	MECH	0.194
10	follow_through_lean	MECH	0.163

Mechanical features constitute **10 of 10** top features and **15 of 15** top features — no baseline feature appears in the top 15. When both feature types are available, the model relies exclusively on mechanical features for prediction. This is a stronger result than incremental improvement: the model does not blend the two feature types, it ignores the baseline entirely, which means the mechanical features carry signal that the metadata does not even partially replicate.

The top three features map directly to the kinetic chain framework described in SW-R-2026-003: **upper body twist** (trunk rotation, the rotational component of force generation), **hip angle asymmetry** (the difference between shooting and standing leg – a committed, well-structured striking posture), and **maximum knee angle** (full knee extension at contact, the biomechanical indicator of a well-timed strike). **Backlift magnitude** (rank 11, SHAP: 0.153) is the feature most directly connected to the "compact vs. wide release" distinction – a smaller value indicates a compact release suited to finishing under defensive pressure.



**Fig. 3.** SHAP summary for the enhanced LightGBM model. Each dot represents one sample; horizontal position indicates the feature's contribution to predicting on-target (positive) or off-target (negative). Mechanical features occupy the entire top of the hierarchy. Baseline features (half, shooter\_score\_diff) appear only at the bottom.



**Fig. 4.** Feature distributions for on-target (green) vs. off-target (red) shots. Hip angle asymmetry is the only feature reaching conventional significance ( $p = 0.037$ ). The absence of strong univariate separation – combined with the strong multivariate SHAP signal – indicates that the predictive information resides in feature interactions, not in individual distributions. This is consistent with the kinetic chain model: it is the coordinated configuration of the body, not any single joint angle, that determines execution quality.

06

## What This Means

The analysis produces three findings, each carrying a different level of evidential strength – and the distinction matters, because the strongest result is not the one the headline metrics might suggest.

**Finding 1 (moderate): Mechanical features dominate the feature importance hierarchy.** When both baseline and mechanical features are available, the model assigns zero importance to baseline features and relies exclusively on mechanical features. SHAP values measure marginal predictive contribution – the model has determined that body-pose features carry signal that game-state metadata does not.

This finding carries a caveat: the baseline is six weak metadata proxies, not a properly specified event-data model. The dominance would need to replicate against shot location + xG + body part to support the stronger claim that mechanical features carry signal that event data does not.

**Finding 2 (moderate): The direction of the effect is consistently positive.** All four model types show improvement in AUC-ROC when mechanical features are added. The mean improvement (+0.056) is modest but directionally consistent across all classifier families.

**Finding 3 (weak): The overall predictive improvement is not statistically significant.** The corrected repeated CV t-test yields  $p = 0.97$ . This is a clean null — a product of experimental constraints (a deliberately minimal baseline, a binary outcome variable,  $n = 200$ ), not necessarily a true absence of effect.

The convergence with Schepers et al. is the finding that carries the most weight. Two independent analyses — different actions, different data, different research groups — produce the same structural result: when skeletal features are available, the model prefers them. The replication across these independent axes is how structural claims accumulate evidential weight.

07

## Next Steps

---

### Immediate: WorldPose + StatsBomb

The WorldPose dataset (2.5 million 3D poses from the 2022 World Cup, 8 cm per-joint accuracy) [3] combined with StatsBomb's open event data for the same tournament would resolve every constraint that limits this analysis. It provides the sample size (10–100x increase), the data quality (multi-camera 3D pose at production accuracy), the matched event features this study lacks (shot location, xG, body part), and a stronger outcome variable (goal conversion rather than on/off target). This combination is the most accessible path to a definitive test — because both datasets are available through academic or open-access channels, the barrier is data linkage, not data access.

### Medium-term: Dribbling Extension

Repeat the comparison for dribble outcomes using WorldPose or equivalent skeletal match data. Structure mechanical features around STATSWING's three-phase dribbling framework (preparation, reception, attack). Compare results with Schepers et al. as a direct validation check.

### Long-term: Transfer Prediction

The dataset that would enable this test — pre-transfer mechanical profiles linked to post-transfer performance outcomes — does not yet exist in the public domain. The Premier League and Bundesliga both now generate skeletal match data at every game. The historical archive — if accessible — would

enable retrospective construction of this dataset for transfers from 2022–23 onward.

#### LIMITATIONS

**No real event-level baseline features.** The baseline model is artificially weak. A proper test requires shot location, xG, body part, shot type, and preceding action data matched to the same shooting instances. The SHAP finding that mechanical features dominate over baseline features is therefore a weaker claim than it would be if the baseline included production-grade event data.

**Shooting leg ambiguity.** Without knowing which foot the player used, left/right mechanical features may introduce noise. The asymmetry features partially address this by measuring magnitude regardless of direction.

**Contact frame identification.** Frame 15 of 20 is used as the contact frame. Actual foot-ball contact may vary. Automatic contact detection would improve feature accuracy.

**Pose estimation accuracy.** The 3D keypoints are lifted from monocular broadcast footage, which introduces estimation error that multi-camera skeletal tracking would not.

**Sample size.** With  $n = 200$  and 26 mechanical features, the analysis is underpowered for detecting moderate effect sizes. A power analysis assuming AUC improvement from 0.50 to 0.55 at 80% power would require approximately 800–1,000 samples.

**SHAP on training data.** SHAP values in Section 5 were computed on a model trained on all 200 samples — the same data used for training. Because SHAP measures feature contributions within the model's learned decision surface, training-data SHAP values may reflect overfitting patterns rather than genuine feature relationships. The feature importance hierarchy should be treated as indicative, not definitive, until replicated on held-out data at larger scale.

#### REPRODUCIBILITY

**Code:** Analysis code is self-contained: loads 3DSP data, engineers features, trains models, computes SHAP values, generates figures, exports results. Available on request via [research@statswing.com](mailto:research@statswing.com).

**Data:** 3DSP dataset is open access. GitHub: [calvinyeungck/3D-Shot-Posture-Dataset](https://github.com/calvinyeungck/3D-Shot-Posture-Dataset). The engineered feature set (200 rows, 34 columns) is reproducible from the raw 3DSP data using the code above.

**Random seeds:** All models use `random_state=42`. Cross-validation uses `StratifiedKFold(shuffle=True, random_state=42)`. Results are deterministically reproducible.

[PROOF-OF-CONCEPT](#)[BIOMECHANICS](#)[BODY-POSE DATA](#)[SHAP ANALYSIS](#)[SHOOTING MECHANICS](#)[KINETIC CHAIN](#)

---

## IMPLICATIONS

### For the parent publication (SW-R-2026-003)

The structural argument — that mechanics is the missing variable in transfer prediction — does not depend on this proof-of-concept. It rests on the logical structure of the compounding stack, the cross-domain precedent from the NBA, and the Schepers et al. dribbling study. This analysis adds a second independent data point: STATSWING's own kinetic chain framework, when operationalised as body-pose features, produces the feature importance hierarchy the framework predicts. The convergence strengthens the structural claim without constituting the definitive empirical test, which requires matched event + pose data at scale.

### For researchers with access to WorldPose or equivalent data

The feature engineering pipeline and experimental design described here can be applied directly to larger, higher-quality datasets. The 26 mechanical features — structured around force generation (hip), extension and accuracy (knee), and balance and preparation (whole body) — provide a starting framework for a properly powered study. The WorldPose + StatsBomb combination is the most accessible path to the definitive test.

### For data providers

The 3DSP dataset was designed for pose clustering and contains no event-level data. The absence of matched event + pose datasets in the public domain is itself a finding: the analytical community cannot test whether mechanical features improve upon event data because no publicly available dataset pairs the two. A provider that releases matched skeletal + event data — even for a small sample of competitions — would enable the research that this proof-of-concept scopes but cannot yet execute.

---

## REFERENCES

- [1] M. Schepers, P. Robberechts, J. Van Haaren, and J. Davis, "What Makes a Dribble Successful? Insights From 3D Pose Tracking Data," arXiv:2506.22503, June 2025. [arxiv.org](https://arxiv.org/abs/2506.22503)
- [2] C. Yeung, K. Ide, and K. Fujii, "AutoSoccerPose: Automated 3D Posture Analysis of Soccer Shot Movements," CVPR 2024 Workshop on Computer Vision in Sports, arXiv:2405.12070. Dataset: [github.com/calvinyeungck/3D-Shot-Posture-Dataset](https://github.com/calvinyeungck/3D-Shot-Posture-Dataset)
- [3] WorldPose dataset, ETH Zurich / FIFA. 2.5 million 3D poses from the 2022 World Cup. Available via academic data request.
- [4] J. Adejola, "The Execution Layer: Mechanics as the Missing Variable in Transfer Prediction," STATSWING Research SW-R-2026-003, March 2026. [statswing.com/research/mechanics/](https://statswing.com/research/mechanics/)
- [5] C. Nadeau and Y. Bengio, "Inference for the Generalization Error," *Machine Learning*, vol. 52, no. 3, pp. 239–281, 2003.

---

## CITE THIS TECHNICAL NOTE

J. Adejola, "Mechanics Proof-of-Concept: Do Body-Pose Features Improve Shot Outcome Prediction?," STATSWING Technical Note SW-TN-2026-001, March 2026. [statswing.com/research/mechanics-poc/](https://statswing.com/research/mechanics-poc/)

## STATSWING

STATSWING Research publications share non-proprietary findings from our analytical work. They represent a portion of the frameworks and methodology that underpin our intelligence products.

For the full analytical infrastructure – including proprietary ratings, player and team assessments, and decision-support briefs – visit [statswing.com](https://statswing.com).

[statswing.com](https://statswing.com) · [research@statswing.com](mailto:research@statswing.com)