Interpretable Pre-Release Baseball Pitch Type Anticipation from Broadcast 3D Kinematics

Jerrin Bright[1,2], Michelle Lu[2], John Zelek[1,2]

[1] Vision and Image Processing Lab [2] University of Waterloo, Canada

CVPR JUNE 3-7, 2026 DENVER COLORADO

TL;DR Thesis

Before ball flight is observed, a pitcher's body already reveals substantial pitch-type information. Broadcast 3D pose captures posture, arm slot, trunk lean, timing, and unintended mechanical tells that are visible from ordinary monocular video. **Bottom line:** body kinematics enable strong pre-release anticipation, while grip-defined variants expose the ceiling of pose-only prediction.

Objectives

Question: how much pitch-type information is visible in a pitcher's body before release?

- Classify eight pitch types from monocular 3D pose alone.
- Identify which joints, events, and biomechanical quantities carry the signal.
- Separate pre-release cues from ball flight.

Dataset

119,561 professional pitches

8 pitch classes

23,913 held-out test pitches

Filtered from 120,471 pitch sequences after minimum length, event-detection, and joint-tracking checks.

Stratified 80/20 train/test split preserves class balance across the eight pitch types.

Pitch	Count	Share
FF	38,560	32.3%
FT	18,570	15.5%
SL	16,221	13.6%
CH	12,170	10.2%
CB	10,456	8.7%
SW	10,225	8.6%
FC	9,743	8.1%
SP	3,616	3.0%

Highlights

- **80.4%** 8-way accuracy from body kinematics alone.
- **+3.9%** absolute gain from explicit biomechanical features and temporal deltas.
- **64.9%** of importance lies in upper-body mechanics; lower body contributes 35.1%.
- Wrist position and head/eye orientation dominate joint-level signal.

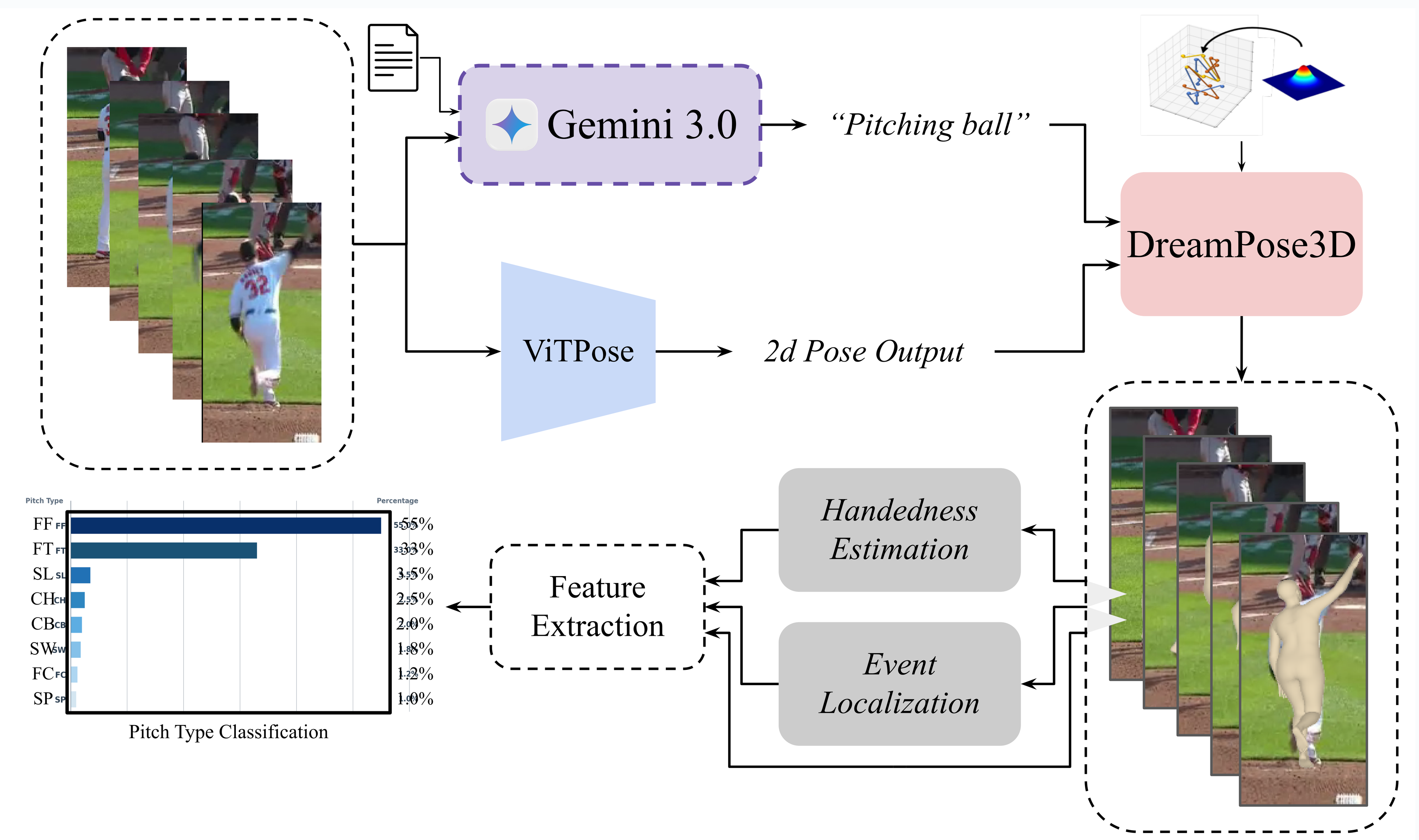
References

[1] Bright et al. "PitcherNet: Powering the Moneyball Evolution in Baseball Video Analytics." *CVPRW*, 2024.

[2] Bright et al. "DreamPose3D: Hallucinative Diffusion with Prompt Learning for 3D Human Pose Estimation." *arXiv*, 2025.

[3] Chen and Guestrin. "XGBoost: A Scalable Tree Boosting System." *KDD*, 2016.

Methodology



Pipeline. Broadcast video → 2D pose and action prompt → DreamPose3D sequence → event localization → 229-feature XGBoost classifier. The method converts ordinary broadcast clips into a compact, interpretable pitching signature. A 2D pose estimator and video-language prompt first describe the pitching motion, DreamPose3D lifts the sequence into 3D joints, and a pose-only processing stack aligns handedness, normalizes camera/pitcher variation, and extracts signals only up to release. This keeps the task anticipatory: the model cannot use ball flight, catcher reaction, or post-release trajectory cues.

Pre-Release Feature Design

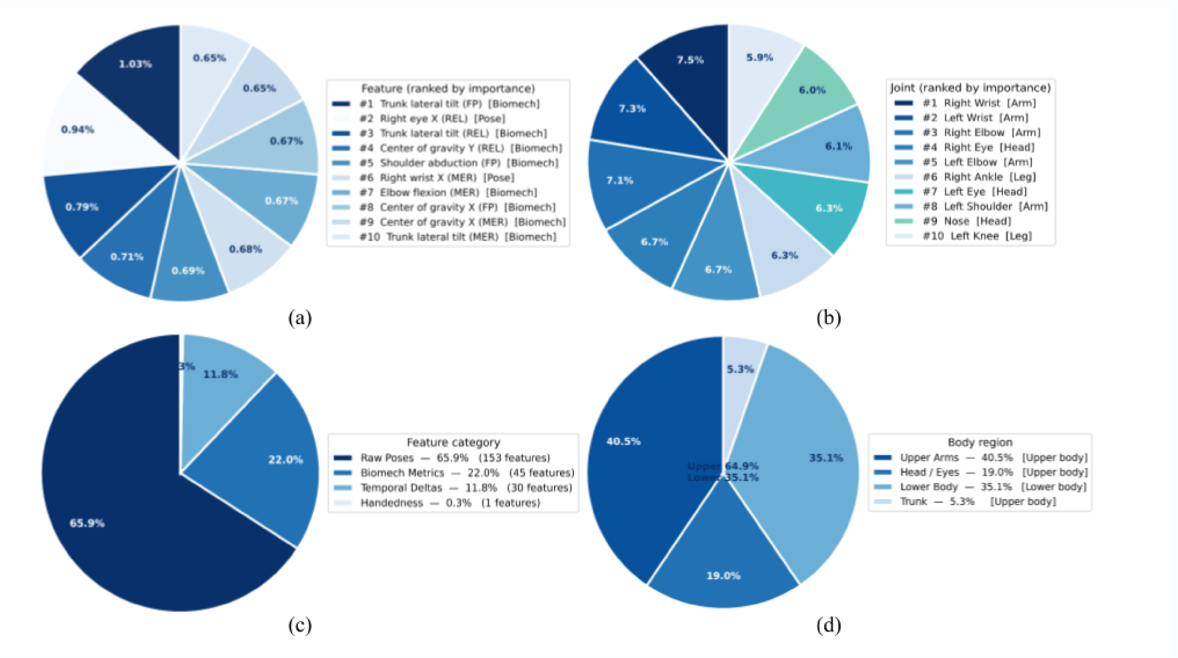
Events. Detect Foot Plant (FP), Maximum External Rotation (MER), and Release (REL) from smoothed lead-ankle and elbow-flexion signals. Event validation reports REL median error of 7.5 ms and FP/MER median errors near 40/34 ms.

Features. Concatenate normalized 3D joints at FP/MER/REL, 15 biomechanical metrics, 10 temporal deltas, and handedness:

$$\mathbf{x} = [\mathbf{f}_{pose}, \mathbf{f}_{bio}, \mathbf{f}_{\Delta}, h] \in \mathbb{R}^{229}.$$

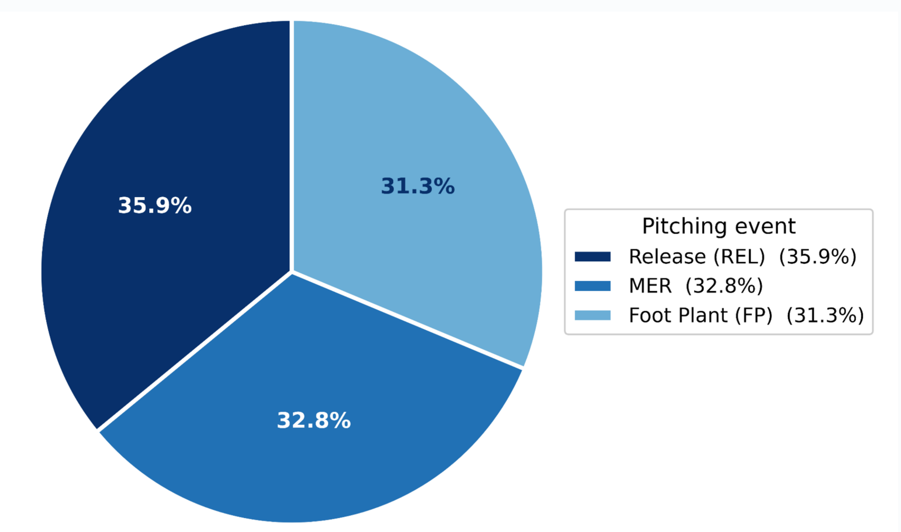
Anticipatory constraint. All features are extracted at or before release, so the classifier measures body-kinematic pitch tells rather than downstream ball flight.

Feature Importance



Importance decomposition: top features, joint groups, feature categories, and anatomical body regions.

Event Timing



Predictive signal is distributed across release (35.9%), MER (32.8%), and foot plant (31.3%), showing that pitch identity is encoded throughout the delivery rather than only at release.

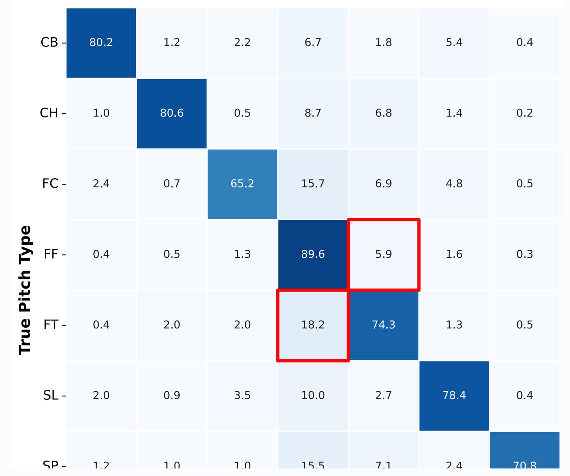
Results

Configuration	Feats.	RF	XGB
Poses only	154	63.1%	76.5%
Poses + biomechanics	229	73.2%	**80.4%**
Poses + ball flight†	166	91.9%	94.0%

†Post-release ball-flight data; included only as an approximate upper bound, not a valid anticipatory setting.

Mechanically distinct classes are strongest: CH F1 = 85%, CB F1 = 84%, SL/SW F1 = 81%. Cutters and two-seam fastballs remain hardest because they overlap with fastball and slider mechanics.

Error Mode



Row-normalized confusion matrix. The dominant systematic error is FF↔FT, with 1,133 reciprocal fastball confusions in the test set.

Future Work

Future work will add hand pose, grip, and hand–ball interaction cues to resolve pitch types whose body mechanics are intentionally similar.

Acknowledgements

Orioles
Vision and Image Processing Research Group
VIP lab
vip.uwaterloo.ca
Mitacs