Scalable Injury-Risk Screening in Baseball Pitching from Broadcast Video

Jerrin Bright[1,2], Justin Mende[2], John Zelek[1,2]
[1] Vision and Image Processing Lab [2] University of Waterloo, Canada

TL;DR

Broadcast video can scale pitching injury screening.

We estimate clinically meaningful pitching biomechanics from a single broadcast camera, validate them against professional tracking data, and convert mechanics, workload, and history into high-recall injury-risk scores.

16/18 metrics validated · **0.811** TJ AUC · **90.3%** sensitivity

Research Question

Can monocular broadcast video replace stadium-scale biomechanics capture for screening?

- Recover 18 clinically relevant pitching metrics from a single camera.
- Validate pose-derived kinematics against professional tracking data.
- Convert per-pitch mechanics into pitcher-level injury-risk signals.
- Prioritize high sensitivity for triage, not clinical diagnosis.

Validation Data

Paired tracking set. 13 professional pitchers, 156 paired pitches, all eight pitch types, and delivery slots from overhand to sidearm.

Large-scale deployment. 119,561 broadcast pitching sequences and 7,348 pitchers for injury-risk modeling.

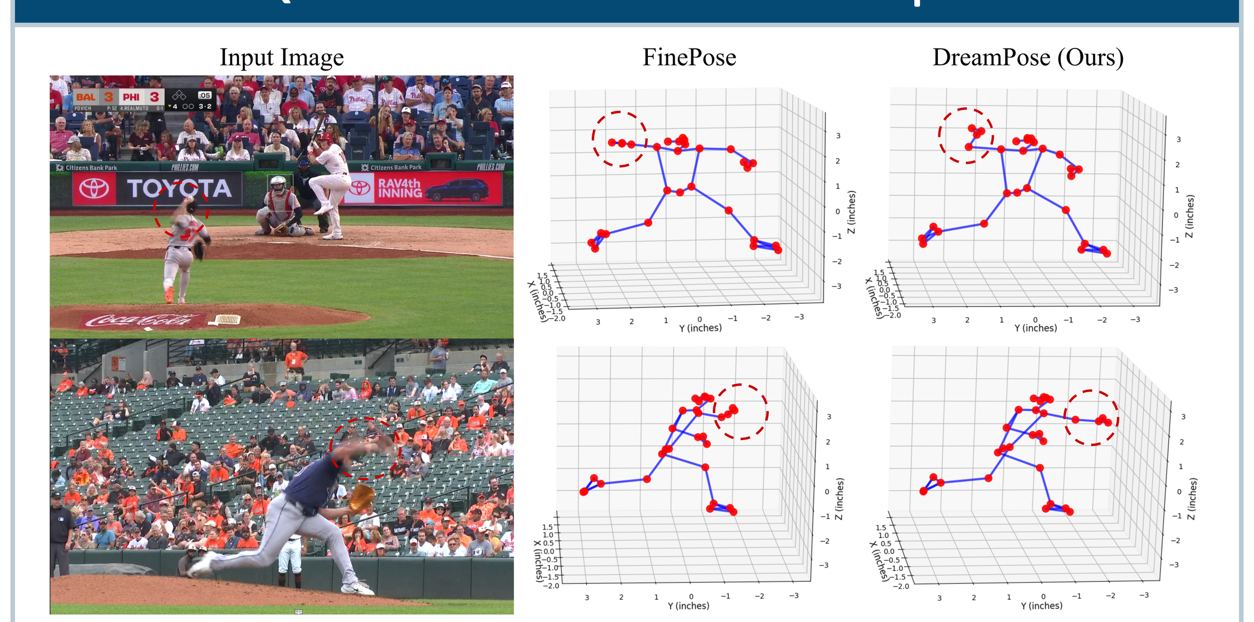
The paired set tests metric fidelity; the large cohort tests whether those metrics scale to screening.

Item	Count
Pitchers	13
Paired pitches	156
Metrics	18
Broadcast clips	119,561
Screening cohort	7,348

Key Findings

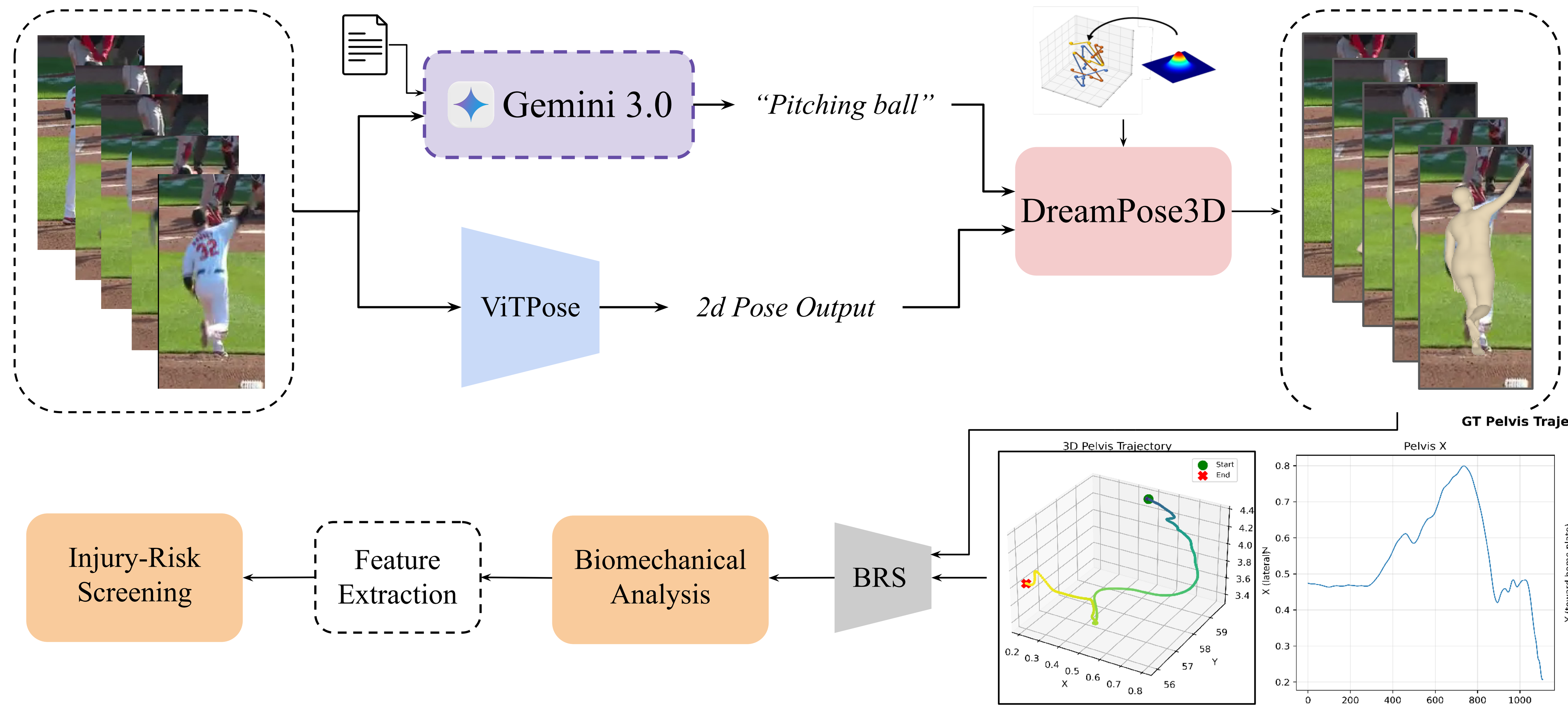
- **16/18** metrics meet the sub-degree or sub-0.1 ft validation target.
- The final ensemble reaches **0.811** AUC for Tommy John surgery and **0.825** AUC for significant arm injury.
- At threshold 0.25, sensitivity reaches **90.3%**, suitable for early screening workflows.
- Range, coefficient-of-variation, and P90 mechanics are stronger injury signals than simple averages.

Qualitative MLB Pose Examples



Qualitative MLB examples show that the pose-estimation stack remains stable across broadcast viewpoints, pitcher scale changes, occlusion, and in-game camera compression.

Method



- DreamPose3D/PitcherNet-style pose recovers temporally consistent 17-joint pitching sequences from broadcast footage [1,2].
- PGLM lifts pelvis-rooted poses into global kinematics using velocity-based trajectory prediction and sliding-window commits.
- BRS enforces constant bone lengths, joint-limited IK, smoothing, and bilateral symmetry before metric computation.
- Per-pitch metrics are summarized into mean, range, P90, coefficient-of-variation, workload, age, and history features.

Tracking Agreement

Metric	MAE	r
Lead knee flexion	**0.3°**	.998
Trail knee flexion	**0.4°**	.997
Lead shin angle X/Y	**0.2°**	.999
Trail shin angle X/Y	**0.3°**	.998
Throw / glove elbow flexion	**0.3–0.5°**	.996+
Pelvis rotation	**0.5°**	.999
Torso rotation	0.9°	.997
Hip–shoulder separation	**0.4°**	.998
Trunk forward / lateral tilt	0.2–0.9°	.996+
COG$_x$ / COG$_z$	**0.02–0.03 ft**	.998+
Shoulder abd. throw / glove	6.2–21.4°	.743+

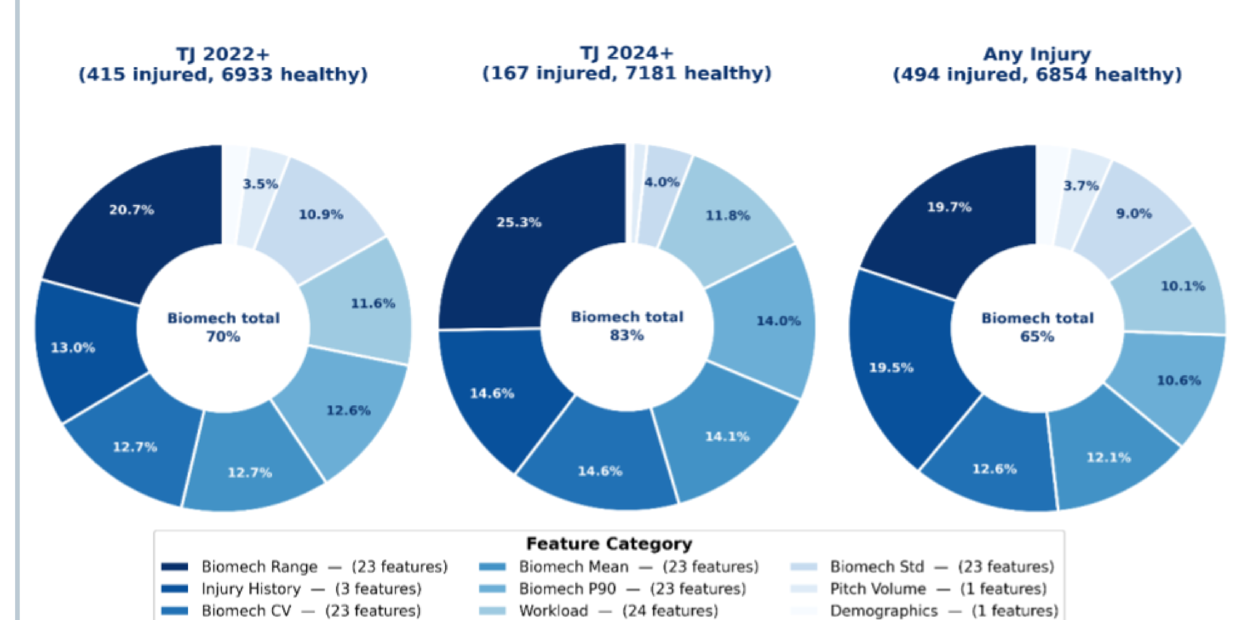
Expanded validation against professional tracking: 16/18 metrics meet the screening threshold of MAE $< 1°$ or < 0.1 ft with $r > .95$.

Strongest	knee flexion, shin angles, elbow flexion, trunk tilt, hip–shoulder separation, and COG$_x$/COG$_z$
Weakest	shoulder abduction, especially glove-side shoulder center localization and monocular depth
Meaning	validated metrics can be used directly; shoulder-dependent metrics should be flagged for cautious review

11 metrics below 0.5° · **5** metrics below 1.0° · **2** metrics needing refinement

This pattern supports scalable screening: most clinically useful lower-body, trunk, elbow, and center-of-gravity quantities transfer reliably from monocular broadcast video.

Risk Signal



Biomechanical features dominate GBM gain across Tommy John and broader arm-injury targets.

Variability	range and coefficient-of-variation features contribute 33–39% of GBM gain
Extremes	P90 hip–shoulder separation captures high-stress mechanics
History	prior injury and workload remain essential context for screening

Feature family	Why it matters	Action
Range / CV	delivery instability	monitor
P90 angles	repeated high stress	review
Workload	exposure volume	adjust

Main Results

Outcome	Score
Validated metrics	16/18
Tommy John AUC	**0.811**
Any arm injury AUC	**0.825**
Sensitivity at 0.25	**90.3%**

16/18 validated · **0.811** TJ AUC · **90.3%** sensitivity

The model is designed as a high-recall triage tool: it flags athletes for review rather than declaring injury causality.

Model stage	TJ AUC
Static thresholds	0.503
Anomalies + workload	0.633
Feature selection	0.734
Enriched ensemble	**0.811**

Interpretation

Why this works. Global lifting and biomechanical refinement remove pose artifacts before small joint errors become large angle errors.

What the risk model learns. Repeated exposure to near-extreme mechanics and high variability is more informative than one average delivery.

How to use it. Coaches can threshold scores to trigger workload changes, mechanics review, or medical follow-up in settings without motion capture.

Signal	Source	Role
P90	mechanics	stress exposure
Range/CV	mechanics	variability
History	records	baseline risk

Next Steps

- Expand validation beyond 13 professional pitchers.
- Improve shoulder-center localization and monocular depth.
- Test prospective screening decisions in real workflows.

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.

Acknowledgements