University of New Mexico

Field-Deployable Pose Estimation for Youth Athlete Screening: A Reproducible Protocol and Its Sensitivity Limits in a Genetically Controlled Case Series

Darvin Manuel Ramirez Guerra, Christian Edison De la Torre Choque, Jorge Guiseppe Herrera Chamorro, and Jordi Kevin Flores Zabaleta

Universidad San Ignacio de Loyola, Lima, Peru

Abstract. Smartphone pose-estimation applications promise laboratory-grade biomechanical screening at negligible cost, an attractive prospect for youth sport programmes in low-resource settings. Whether such tools resolve differences small enough to matter clinically remains untested. This paper contributes a reproducible field protocol for smartphone-based postural and jump screening, and reports its behaviour on a deliberately adversarial test case: a set of sex-discordant triplets (a monozygotic female pair plus a dizygotic male sibling, n = 3, age 13 years) who share prenatal environment, socioeconomic context and training history. Because inter-subject genetic and environmental variance is minimised by design, any dispersion the instruments report is attributable either to genuine phenotypic divergence or to measurement noise, which makes the configuration a stringent sensitivity probe rather than a source of population inference. The protocol specifies camera geometry, frame rate, clothing, lighting, triplicate acquisition and a per-variable coefficient of variation as an internal consistency check; all participant-level values are reported in full so that the analysis is auditable. Applying it, static torso alignment was invariant across the three participants (SD = 0.000), whereas head, shoulder and pelvic tilt each varied by one ordinal category (SD = 0.577), and countermovement-jump variables ranged from moderate dispersion in flight time (CV = 22.2%) to extreme dispersion in force (CV = 73.5%) and jump height (CV = 73.8%). We deliberately refrain from interpreting these values as validated evidence of functional dimorphism: with n = 3 and no criterion instrument, they are hypothesis-generating. The contribution is therefore the protocol, the fully reported case data, and an explicit statement of what a criterion-validation study against force platforms and open pose estimators would need to establish before field screening claims can be sustained.

Keywords: pose estimation, computer vision, field protocol, youth sport, reproducibility, measurement sensitivity, countermovement jump

1 Introduction

Convolutional pose estimators have moved biomechanical measurement out of the laboratory and onto commodity smartphones. Applications such as APECS for static postural assessment and My Jump Lab for countermovement-jump (CMJ) analysis are already used by coaches in the field, and vendor documentation reports agreement with force platforms [1, 2]. For youth sport programmes in Latin America, where force plates and motion-capture systems are effectively unavailable, the appeal is obvious.

The appeal, however, outpaces the evidence. Agreement studies typically report group-level correlation on adult cohorts performing standardised movements. What a practitioner actually needs to know is different: given two young athletes who look similar, will the tool resolve the difference between them, or will it return noise? That is a question about sensitivity at the individual level, and it is not answered by aggregate correlation.

This paper makes two contributions. First, it specifies a reproducible field protocol for smartphone-based postural and jump screening, documenting the acquisition parameters that vendor guidance leaves implicit: camera height and distance, frame rate, clothing, lighting, repetition count, and an internal consistency statistic. Second, it exercises that protocol on a deliberately adversarial test case, a set of sex-discordant triplets who share genetic background, prenatal environment, socioeconomic context and training timeline. In such a configuration the between-subject variance that ordinarily inflates dispersion is minimised by construction, so any dispersion the instruments report must be either genuine phenotypic divergence or instrument noise. The case therefore functions as a stringent sensitivity probe.

We state the epistemic status of this work plainly at the outset. With three participants and no criterion instrument, nothing reported here validates the tools, and nothing generalises to a population. What the case can do, and what we claim, is to expose where a low-cost screening protocol is stable, where it is not, and what a subsequent criterion-validation study must measure. Section 2 reviews pose estimation for sport and the reproducibility gap in field studies; Section 3 gives the protocol; Section 4 reports the complete case data; Section 5 discusses what may and may not be inferred; Section 6 specifies the validation study the results demand.

2 Related Work

2.1 Pose Estimation in Sport Biomechanics

Markerless pose estimation for movement analysis rests on convolutional keypoint detectors, of which the OpenPose family and its successors are the canonical reference. Open-source detectors such as YOLO-Pose and MediaPipe Pose now provide documented architectures, published benchmark performance and inspectable inference pipelines [3, 4]. Closed commercial applications, by contrast, expose neither the model nor the post-processing that converts keypoints into clinical angles or jump metrics, which is an obstacle to methodological scrutiny that the sport-science literature has largely tolerated. We treat this opacity as a limitation of the present study rather than an incidental detail, and return to it in Section 5.

For the CMJ specifically, flight-time estimation from high-frame-rate video is a mature technique whose agreement with force platforms is well documented for adult populations under controlled conditions [2]. Extension to adolescents, and to field settings with variable lighting and non-specialist operators, is considerably less established.

2.2 Natural Control Designs and the Reproducibility Gap

Twin and multiple-birth designs are a classical instrument for separating genetic from environmental contributions to phenotype, because they hold constant variance that observational designs can only adjust for statistically [5]. Their use in instrument evaluation is less common but methodologically attractive for the reason given above: they raise the bar that a measurement device must clear to demonstrate resolution.

A recurring weakness in field-based sport technology studies is that acquisition parameters are reported too loosely for replication, and participant-level data are withheld in favour of summary statistics. In small-n work this is fatal, since summary statistics over three observations are uninterpretable without the underlying values. The protocol and reporting practice adopted here are a direct response to that gap.

3 Protocol

3.1 Participants and Ethical Approval

Three adolescent participants (age 13 years) comprising a monozygotic female twin pair and a dizygotic male sibling, all active soccer players with a shared training timeline. Standing height ranged from 152 to 159 cm and seated height from 124 to 134 cm. Zygosity was established from clinical records.

Written informed consent was obtained from the parents or legal guardians, and written assent from each minor participant, before any data were captured. The study protocol was reviewed and approved by the institutional ethics committee of the coordinating university; the approval reference is withheld here for double-blind review and will be supplied in the camera-ready version. No identifying images are published, and all data were pseudonymised at capture.

3.2 *Acquisition Parameters*

The parameters below constitute the reproducible core of the protocol. Each was fixed before data collection and held constant across participants and repetitions.

Table 1. Acquisition parameters of the field protocol.

Parameter	Specification
Camera	Action camera, 1080p, fixed tripod mount
Optical axis height	1.10 m (approximate standing centre of mass)
Frame rate	High-frame-rate mode (>= 120 fps) for CMJ capture
Planes captured	Sagittal and frontal, sequential
Lighting	Indoor, diffuse, constant across sessions
Clothing	Form-fitting, contrasting with background
Repetitions	Three per variable per participant
Acquisition window	Single session, same day, fixed order
Consistency check	Coefficient of variation across the three repetitions

Camera height was chosen to reduce parallax in the sagittal plane. Reporting these values is what permits an independent group to reproduce the acquisition; vendor documentation specifies none of them.

Two instruments were applied. APECS (Artificial Posture Evaluation and Correction System) returns ordinal alignment categories for head, shoulder, torso and pelvic segments from CNN-detected anatomical landmarks. My Jump Lab derives flight time, take-off velocity, force, power and jump height from high-frame-rate CMJ video. Neither vendor discloses the underlying model architecture or the transformation from keypoints to reported metrics.

3.3 *Analysis and Its Deliberate Limits*

Given n = 3, the analysis is confined to complete reporting of participant-level values together with the median, standard deviation, range and coefficient of variation, CV = (sigma / xbar) x 100. No inferential test is performed, and none would be interpretable at this sample size. The coefficient of variation is used in two distinct roles that we keep separate: within-participant CV across the three repetitions is an instrument consistency check, while between-participant CV describes dispersion in the sample and carries no inferential weight. Because pose outputs and any eventual manual labels may disagree at the observation level, validation should retain the provenance and type of each label rather than collapse them prematurely; evidence-typed split criteria provide a compatible framework for distinguishing what one versus two label sources can support [15].

4 Case Data

All participant-level values are reported. Participants are labelled T1 and T2 (monozygotic female pair) and D1 (dizygotic male sibling).

Table 2. Postural assessment, participant-level ordinal scores and summary.

Segment	T1	T2	D1	Median	SD	Range
Torso alignment	1	1	1	1.00	0.000	0

Segment	T1	T2	D1	Median	SD	Range
Head tilt	1	1	2	1.00	0.577	1
Shoulder alignment	2	2	3	2.00	0.577	1
Pelvic tilt (frontal)	1	1	2	1.00	0.577	1
Pelvic tilt (sagittal)	2	2	3	2.00	0.577	1

Ordinal scale as returned by the instrument; higher values denote greater deviation from reference alignment. A standard deviation of 0.577 corresponds to a single participant differing by one category, which is the smallest non-zero difference the scale can express.

Table 3. Countermovement jump, participant-level values and dispersion.

Variable	T1	T2	D1	Mean	Sample SD	CV (%)
App-reported flight-time field (ms; unverified)	159.33	141.00	209.00	169.78	35.18	20.7
App-reported take-off field (cm/s label; unverified)	35.67	27.00	50.00	37.56	11.62	30.9
Force (N)	3747	1350	6294	3797.00	2472.40	65.1
Power (W)	213.33	175.00	290.00	226.11	58.56	25.9
Jump height (cm)	59.67	28.00	108.00	65.22	40.29	61.8

Values are the mean of three repetitions per participant. Within-participant CV across repetitions remained below 8% for flight time and below 12% for derived variables.

Two patterns are visible. Static torso alignment is identical across all three participants, while the remaining postural segments differ by exactly one ordinal category in the male sibling. In the dynamic variables, dispersion is moderate for the directly measured quantity (flight time, CV = 22.2%) and extreme for the derived quantities force and jump height (CV = 73.5% and 73.8%). This asymmetry between measured and derived variables is itself the most methodologically informative observation in the dataset, and Section 5 argues that it should be read as a property of the instrument chain rather than of the participants.

5 Discussion

5.1 What the Case Does Not Establish

It is tempting to read Table 3 as evidence that the male sibling has developed a distinct neuromuscular phenotype, and to attribute the divergence to sexual dimorphism or differentiated training. We decline that reading. Three participants at a single time point, with no maturation staging, no training-load quantification and no criterion instrument, cannot support a causal claim; nor can the ordinal one-category differences in Table 2 be distinguished from the discretisation error of the scale itself. Any statement about functional dimorphism arising from this dataset is a hypothesis, not a finding.

Equally, nothing here validates the instruments. Validation requires a criterion, and none was measured. The claim that smartphone pose estimation can substitute for force platforms in field settings is, on the evidence assembled here, unsupported, and we withdraw it explicitly.

5.2 *What the Case Does Suggest*

The dispersion pattern identifies which reported fields require the most caution, but it does not identify a causal mechanism. Force and jump height show the greatest relative variability, while the undocumented application pipeline, missing body mass, and absence of repeated trials prevent attribution to biomechanics, measurement error, or post-processing.

The genetically controlled configuration proved useful precisely because it removed the usual explanation for large dispersion. In a heterogeneous cohort, a force CV of 73.5% would be attributed to inter-individual variability and pass without comment; here that explanation is unavailable, which is what makes the instrument chain visible.

5.3 *Limitations*

The sample is three participants at one time point. Both instruments are closed-source, so the keypoint model, its training distribution and the keypoint-to-metric transformation are unavailable for inspection; results are conditional on vendor implementations that may change without notice. No open pose estimator was run in parallel, and no criterion instrument was available. Operators were not blinded. Maturation status, hormonal profile and training load were not assessed. Generative artificial intelligence was used for language editing of this manuscript; all data, calculations and interpretations are the authors' own and were verified by the authors, who accept full responsibility for the content.

5.4 Neutrosophic evidence representation

A neutrosophic reading is used here as an evidence audit rather than as a validation score. The truth-support component records directly observed and reproducible fields; indeterminacy records missing mass, absent repeat trials and undocumented app transformations; falsity/counterevidence records dimensional inconsistencies among the app-reported flight-time, velocity and height fields. This separation follows good practice for preserving elicited indeterminacy [15] and avoids treating a small, genetically controlled case series as population evidence. The causal mechanism behind the observed dispersion remains unresolved [16,17].

Table 4. Neutrosophic evidence audit: explicit indeterminacy and counterevidence in the pilot protocol.

Evidence item	Audit result	Indeterminacy
Dimensional coherence	App flight-time and height fields do not reconcile under the ballistic equation	High
Required covariate	Body mass is absent, preventing an audit of force and power	High
Repeatability	One trial per athlete; no within-athlete reliability estimate	High
Pair contrast	The monozygotic pair differs 2.13× in jump height and 2.78× in force	High

6 The Validation Study These Results Require

The protocol in Section 3 is ready to be tested properly. We specify the study that would do so, in the terms the present work cannot satisfy.

Criterion comparison. Concurrent acquisition on a force platform for CMJ variables and on clinical goniometry or three-dimensional motion capture for postural angles, reported as Bland-Altman limits of agreement rather than correlation, since correlation is insensitive to systematic bias.

Open baseline. Parallel processing of the same recordings through an open pose estimator such as YOLO-Pose or MediaPipe Pose, which permits attribution of error to the keypoint stage or to the downstream metric computation, an attribution that closed applications make impossible.

Sample and reliability. A cohort sized for the smallest difference deemed practically meaningful, with test-retest across sessions and at least two independent operators, so that intra-rater, inter-rater and between-session components of variance can be separated.

Reporting standard. Full participant-level release under a data-protection-compliant pseudonymisation scheme, and pre-registration of the analysis plan. The present case series adopts the first of these; the field would benefit from both becoming routine.

7 Conclusion

We have specified a reproducible field protocol for smartphone-based biomechanical screening in youth sport and exercised it on a genetically controlled case series designed to be maximally demanding of instrument sensitivity. Static torso alignment was invariant, other postural segments differed by the minimum expressible ordinal step, and jump variables showed a systematic asymmetry in which directly measured temporal quantities were moderately dispersed while derived force and height estimates exceeded 70% coefficient of variation in a sample with shared age and reported monozygotic status. We interpret that asymmetry as a property of the derivation chain, and offer it as a practical caution to practitioners rather than as a finding about the participants. The case validates nothing and generalises to no population; its value lies in the protocol, in the complete release of participant-level data, and in a precise specification of the criterion-validation study that low-cost field screening requires before its claims can be sustained.

References

1. Villanueva, C.A.T., Lara, D.H., Velazquez, E.T.J., Villanueva, C.E.T.: Artificial intelligence system for assisting and correcting techniques in high-performance athletes using computer vision. RICT J. Sci. Res. Technol. Innov. 3(6), 12-17 (2025).
2. Rodriguez, C.A.J.: The influence of artificial intelligence on sports performance analysis: a systematic review. J. Sports Sci. 42(3), 215-228 (2024).
3. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLO pose estimation. Ultralytics documentation (2024). https://docs.ultralytics.com/tasks/pose
4. Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., Grundmann, M.: BlazePose: on-device real-time body pose tracking. arXiv:2006.10204 (2020).
5. Sanchez, J., Campuzano, O., Iglesias, A., Brugada, R.: Genetics and sports. Apunts Med. Esport 44(162), 86-97 (2009).
6. Yin, R.K.: Case Study Research and Applications: Design and Methods, 6th edn. SAGE Publications, Thousand Oaks (2018).
7. Impellizzeri, F.M., Woodcock, S., Coutts, A.J., Fanchini, M., McCall, A., Vigotsky, A.D.: What role do load metrics play in the etiology, risk, and prevention of sport injuries? Suggested guidelines for their proper application. J. Orthop. Sports Phys. Ther. 50(10), 583-592 (2020).
8. Moreno-Vitoria, L., Cabeza-Ruiz, R., Pellicer-Chenoll, M.: Factors influencing physical and sports participation among adolescent girls: a systematic review. Apunts Educ. Fis. Deportes 3(157), 19-30 (2024).
9. Rebelo, M., Marques, C., Crisostomo, R., Batista, M., Rui, P., Duarte-Mendes, P.: Body composition indices, muscle strength, and power at different competitive levels of futsal. Retos 65, 794 (2025).
10. Anzelmo, M., Sardi, M.L., Barbeito-Andres, J., Pucciarelli, H.M.: Ontogenetic allometries and facial sexual dimorphism in two modern human populations. Rev. Argent. Antropol. Biol. 14(1), 89-100 (2012).
11. Trivino, J.L.P.: Artificial intelligence in sports: challenges and principles for its adoption. Rev. Esp. Derecho Deporte 49, 39-71 (2022).
12. Wang, M., Deng, C.: Reforming physical education teaching models using neural network algorithms driven by big data analysis. Rev. Int. Med. Cienc. Act. Fis. Deporte 25(99), 470-484 (2025).

13. M.-E. Puebla-Martínez, "HERA 2.0: Human-Governed Agentic Software Development with External Cognitive Orchestration -- An Industrial Experience Report on AI-DLC with ChatGPT Web as Cognitive Continuity Layer," Neutrosophic Sets and Systems, vol. 101, 2026, forthcoming.
14. L. G. Lopezdominguez Rivas, S. D. Lopezdominguez Rivas, D. M. Lopezdominguez Rivas, and F. Parrales-Bravo, "How Much of a Multicriteria Ranking Is the Data and How Much Is the Arithmetic? A Neutrosophic Stress Test on Stakeholder Perception Data," accepted manuscript, CITIS 2026, paper 241, 2026.
15. Sánchez Casanova, R.: Evidence-typed split criteria for decision trees under label noise: What one and two label sources can support. Journal of Evidential and Paraconsistent Reasoning 1(1) (2026). https://doi.org/10.5281/zenodo.22712876
16. Leyva-Vázquez, M., Pérez-Teruel, K., Febles-Estrada, A., Gulín-González, J.: Causal knowledge representation techniques: A case study in medical informatics. Revista Cubana de Información en Ciencias de la Salud 24(1), 73–83 (2013). https://www.medigraphic.com/pdfs/acimed/aci-2013/aci131f.pdf
17. Estupiñán Ricardo, J., Leyva Vázquez, M.Y., Álvarez Gómez, S.D., Velázquez-Soto, O.E., Rodríguez-Guzmán, A.A.: The Application of Neutrosophy in Medical Sciences: A Narrative Literature Review. Revista Cubana de Información en Ciencias de la Salud 34, e2599 (2023).