

# A MULTIMODAL DIGITAL TWIN FOR INJURY RISK PREDICTION IN ELITE FEMALE FOOTBALL

Canal B<sup>1,2</sup>, Dueñas Y<sup>2</sup>, Ramon A<sup>2</sup>, Jarrin P<sup>3,4</sup>, Yanguas X<sup>3,4</sup>, Gonzalez JR<sup>1,5</sup>, Isus L<sup>2</sup>, Rodas G<sup>3,4</sup>



<sup>1</sup>Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain

<sup>2</sup>Made of Genes, Barcelona, Spain

<sup>3</sup>Medical Department of Football Club Barcelona (FIFA Medical Center of Excellence) Barcelona, Spain

<sup>4</sup>Barça Innovation Hub, Health & Wellness Area, Barcelona, Spain.

<sup>5</sup>CIBER in Epidemiology and Public Health (CIBERESP), Barcelona, Spain

## Introduction and purpose

Precision sports medicine in elite football is constrained by small cohorts, extreme class imbalance, and the multifactorial nature of injury and performance adaptation. Digital avatars (human digital twins) provide a pathway to personalised, actionable decision support by integrating stable “innate” predisposition with dynamic longitudinal monitoring. To our knowledge, no digital-twin system has yet been prospectively deployed in elite football to integrate genetics, multi-omics/medical screening, workload and psychosocial/contextual data into a single actionable framework. This approach transitions from the individual identification of omics biomarkers toward an integrated predictive model (1), allowing for a holistic view of the athlete. We aimed to design and prospectively validate a digital-twin system for elite football, developed using multicohort retrospective longitudinal data from both male and female squads across several professional teams, modelling sex-specific effects and deployed prospectively during the 2025/26 season. Results are reported as an illustrative women’s-team use case (FC Barcelona Women).

## Methods

A stakeholder-driven data architecture and a bespoke injury ontology were co-designed with medical and performance staff to harmonise phenotypes and standardise injury definitions across squads. Determinants were curated from clinical literature and routine club workflows, then organised into complementary domains: i) genetics, via validated polygenic risk scores relevant to lower-limb muscle/tendon/ligament phenotypes (2); ii) routine longitudinal medical screening integrated into club workflows (clinical history, laboratory biomarkers/biochemistry, body composition, and targeted physiological/metabolomic panels); iii) daily external load from GPS; iv) wellness, recovery, and psychological assessments; and v) contextual/exposome information. Modelling quantified the marginal contribution of each layer, followed by integration via phenotype-specific weighting to preserve interpretability (Figure 1). The framework incorporates female-specific variables and domains (including hormonal and menstrual-cycle data) to ensure a robust women’s model, thereby mitigating the historical underrepresentation of women in existing models (3). Predictive modeling was implemented using an XGBoost framework, which was rigorously optimized via GridSearchCV to maximize Precision-AUC. Model outputs were co-designed to be actionable: athlete-level short-horizon risk, key drivers, and scenario-based recommendations aligned with training and match preparation. Performance was evaluated using time-aware validation and precision-recall metrics.

## Results

In the women’s-team use case, data from 47 professional players using multi-season GPS data (>30,000 sessions over 8 seasons) linked to >75 medically-curated non-contact muscle time-loss injuries. Using a 7-day forecasting horizon, the workload-driven model integrated player profiles, session typology, and injury history, outperforming random baselines by 29.6% (AUPRC 0.0135 vs. 0.01042). At the operational threshold, the system captured 33% of all injury events (recall), identifying drivers such as acute load variability and reduced chronic work capacity. Early functional testing with sports-medicine and performance staff confirmed that risk visualisations and recommended levers were actionable in planning discussions. Furthermore, baseline predisposition and screening profiles were integrated at the athlete level to support personalised interpretation of workload-related alerts, allowing staff to prioritise follow-up assessments based on each athlete’s unique risk profile.

## Conclusions

A multimodal digital-twin approach is feasible in elite football and enables the translation of heterogeneous data into clinically actionable decision support. Multicohort development mitigates small-n constraints while preserving athlete-level personalisation and sex-specific modelling. This work supports the transition from isolated biomarker discovery toward integrated, short-horizon precision models that directly inform weekly injury-risk management in elite female football.

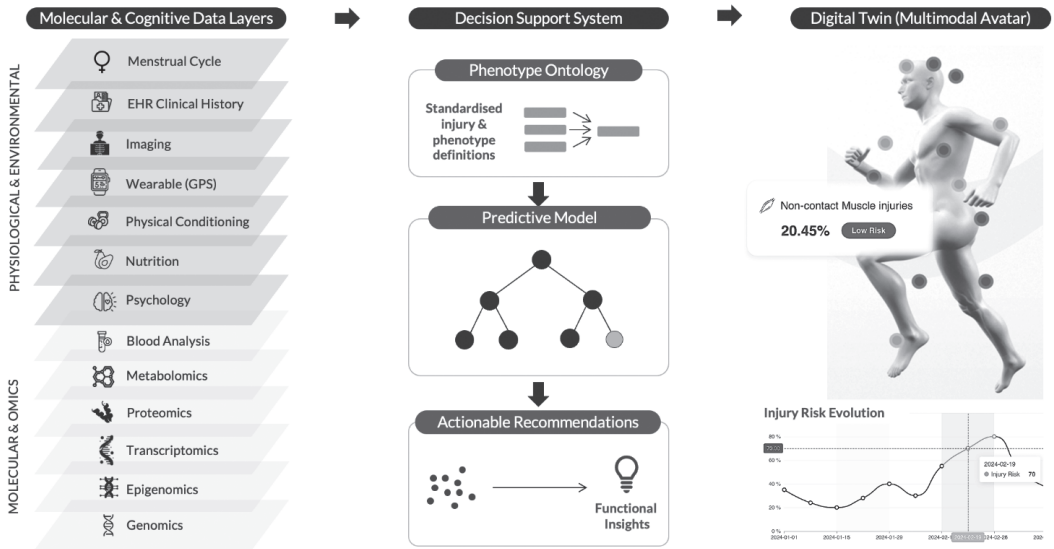


Figure 1. Multimodal digital-twin workflow for actionable injury-risk decision support in elite football.

## References

1. Rodas G, González JR. Sportomics: Futbol Club Barcelona's approach to personalized injury prevention. *Apunts Sports Medicine* 2024; 59 (222): 100445.
2. Ding H, Deng Q, Guo Z. Genetic and epigenetic determinants of injury risk [...]: toward precision sports medicine. *Gene* 2025; 980: 149957.
3. Sun L, Luo L, Yang Y, Wangchuan Q, Luo J. Gender differences in the incidence and types of sports injuries among female athletes [...]. *Journal of Nursing Management* 2025; 2025: 6416101.