ECAI 2025
I. Lynce et al. (Eds.)

doi:10.3233/FAIA251434

From Unsupervised Phenotyping to a Clinician-Ready Classifier: A Complete Pipeline for Assessing 90° Change-of-Direction Technique in Footballers

Alessandro Ghibellini[a,*], Stefano Di Paolo[b], Stefano Zaffagnini[b], Luciano Bononi[a], Maurizio Gabbrielli[a] and Francesco Della Villa[c]

[a]Dept. of Computer Science and Engineering, University of Bologna, Bologna, Italy
[b]IRCCS Rizzoli Orthopaedic Institute, Bologna, Italy
[c]Isokinetic Medical Group, Bologna, Italy

Abstract. Anterior cruciate ligament (ACL) rupture remains a leading cause of long-term absence in football and is frequently triggered by sub-optimal mechanics during rapid change-of-direction (COD) manoeuvres. To transform laboratory motion analyses into actionable field feedback, a three-stage pipeline is presented. First, 45 kinematic and kinetic variables recorded from 1,009 youth footballers during COD tests are embedded with t-SNE and clustered by agglomerative clustering (Euclidean distance, Ward linkage), revealing four force-and-control phenotypes aligned with established ACL-risk mechanisms. Second, a Random-Forest classifier recreates these phenotypes from a set of 12 features selected for both importance and ease of capture, achieving a macro-averaged $F_1 = 0.85$ (compared with 0.92 when all 45 variables are used). Third, the classifier is wrapped in a General Data Protection Regulation (GDPR) compliant application that accepts the 12 inputs, instantly assigns the athlete's phenotype, and displays tailored exercise cues. The pipeline demonstrates that laboratory-grade biomechanics can be condensed into a rapid, interpretable decision-support tool, enabling data-driven ACL injury mitigation in routine sports medicine practice.

1 Introduction

Preventing anterior cruciate ligament (ACL) injuries in football is difficult because both intrinsic factors, such as neuromuscular control, and extrinsic factors, like playing situations, interact in complex ways [9, 13]. In fact, approximately half of ACL ruptures in professional football occur without direct knee contact [11, 4]. Despite improved knowledge and structured prevention programs, the overall injury incidence has not decreased over the past two decades [18]. Two-dimensional (2-D) video scoring offers a quick, low-cost way to flag high-risk movement patterns [5]. AI techniques, including deep learning (DL) and machine learning (ML), have been used in the field of ACL injuries. Supervised works have already shown encouraging results: Schulc *et al.* reconstructed 3-D joint kinematics from single-camera match footage and achieved an AUC–ROC of 0.88 for classifying ACL-injury events [15], whereas the 30-variable field-screening battery proposed by Jauhiainen *et al.* reached an AUC–ROC of 0.63 when predicting future ACL injuries in 880 elite female athletes [10]. Complementarily, unsupervised approaches have also uncovered movement patterns linked to heightened injury risk: early-peak knee-valgus moments in change-of-direction tasks [16], distinct coordination strategies in drop vertical jumps [7], and countermovement jump strategies related to musculoskeletal injuries [1]. Importantly, these studies show that larger samples and richer feature sets generally improve model stability and generalizability. Building on this evidence, the present work leverages a high-quality dataset providing force plate and high-speed cameras data across three 90° COD trials per limb to deliver an end-to-end pipeline:

1. **Stage 1 — Unsupervised phenotyping**: 45 demographic, force, and kinematic variables from 1,009 footballers are embedded with t-SNE and clustered via agglomerative clustering (Euclidean distance, Ward linkage), yielding four interpretable force–control characteristics, defined as phenotypes.
2. **Stage 2 — Supervised phenotyping**: A Random-Forest classifier reproduces those phenotypes from only 12 high-importance, easy-to-acquire variables, achieving a macro-averaged $F_1 = 0.85$ while the full 45-variable model reaches $F_1 = 0.92$.
3. **Stage 3 — Real-time deployment**: The classifier is embedded in a General Data Protection Regulation (GDPR) compliant application that returns an athlete's phenotype and visualises personalised feedback for clinicians and players.

2 Methods

2.1 Dataset and Pre-processing

The dataset derives from the *Cut the ACL* project, a prospective study of 1,009 competitive footballers (16.3 ± 2.8 yr, 26% female) aimed at characterising COD mechanics and ACL-risk factors [6]. Each athlete executed three maximal-speed 90° COD trials per limb inside the dedicated Green-Room (Isokinetic, Bologna, IT), a 12 m runway instrumented with three 240 Hz motion-capture cameras (Vicon, Oxford, UK) and a floor-embedded force platform (AMTI, Watertown, MA, USA). The protocol emphasises COD because the manoeuvre

* Corresponding author: alessandr.ghibellin2@unibo.it

combines high ground-reaction loads with demanding neuromuscular control, replicating key circumstances of ACL rupture [8]. Figures 1 and 2 show the raw kinematic and kinetic outputs, which are exported from cameras and the force platform into a unified database. Player-wise means across trials were discretised into tertiles (0–1–2) to place joint-angle and force data on a common ordinal scale before dimensionality reduction.

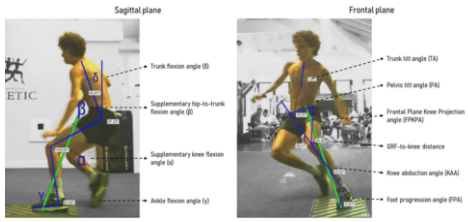

Figure 1. Instrumented "Green Room" corridor: three 240 Hz cameras capture 2-D kinematics while a force plate records ground-reaction forces during the 90° cut.

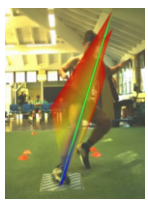

Figure 2. Ground-reaction-force vector (blue-green) and impulse surface (red–yellow) on a single COD frame.

2.2 *Stage 1 — Unsupervised Phenotyping*

The 45 continuous variables were projected onto two dimensions with t-Distributed Stochastic Neighbour Embedding (t-SNE) [17], (Scikit-learn 1.3.1, Python 3.12). Ward-linkage agglomerative clustering (Scikit-learn 1.3.1, Python 3.12) with Euclidean distance was then applied to the 2-D embedding. Perplexity was tuned with RBFOpt [3] to maximize the silhouette [14] evaluated across random seeds, whereas the number of clusters was selected in consultation with clinicians to prioritize clinical interpretability.

2.3 *Stage 2 — Supervised phenotyping*

A Random-Forest classifier (Scikit-learn 1.3.1, Python 3.12) [2] was trained on the complete 45-feature set to predict the phenotype label from stage 1. Model development followed a stratified 3-fold cross-validation scheme. The number of estimators, tested with different random states, was tuned by the optimizer RBFOpt [3]. Random Forest was preferred after brief internal benchmarking for its predictive power and low inference latency on a standard workstation, properties that make the model suitable for real-time deployment.

2.4 *Stage 3 — Real-time deployment*

Real-time deployment requires a lean input layer. Twelve variables were therefore selected by intersecting (i) the highest-ranked features in the Random-Forest importance list, computed as the mean and standard deviation of accumulation of the impurity decrease within each tree, with (ii) variables that clinicians can easily capture in the Green-Room. This compromise maintains predictive power while limiting data entry time. The graphical interface was built with Streamlit 1.30 running on Python 3.12 inside a lightweight Docker image [12]. The container bundles both the Streamlit front-end (‘app.py‘) and the serialised Random-Forest model (‘model.pkl‘), eliminating any need for external calls. Deployment occurs on an on-premise Ubuntu 22.04 server equipped with $\mathbf{2}\times$ Intel Xeon E5-2680 v4 CPUs @ 2.40 GHz and 4 GB RAM, situated behind the clinic's virtual private network (VPN). Because all computation and storage stay on this machine, raw biomechanical data and predictions remain entirely within the clinic's infrastructure, ensuring full GDPR compliance without cloud reliance.

3 Results

3.1 *Phenotype Characteristics*

After the phase was adjusted, 1,002 footballers were successfully clustered into 4 groups (t-SNE optimal parameters: p = 497, s= 88), raising the silhouette from 0.33 to 0.42 (Figure 3). Each group entails specific biomechanical characteristics, the phenotype:

- **Cluster 0: Low-Force / Low-Control (LF–LC).** Athletes display limited hip and knee flexion, pronounced knee valgus, and the lowest ground reaction peaks, indicating both weak braking capacity and sub-optimal alignment.
- **Cluster 1: Low-Force / High-Control (LF–HC).** Good frontal and sagittal plane stability offsets modest braking forces; players absorb load through cleaner hip knee mechanics, suggesting room to improve performance without large injury-risk penalties.
- **Cluster 2: High-Force / Low-Control (HF–LC).** Large ground reaction peaks coincide with knee valgus and excessive trunk lean. This phenotype reflects a power-oriented strategy that may elevate joint loading unless neuromuscular control is addressed.
- **Cluster 3: High-Force / High-Control (HF–HC).** Athletes generate the highest braking forces while maintaining stable lower limb and trunk kinematics; the combination typifies an efficient, lower-risk cutting strategy.

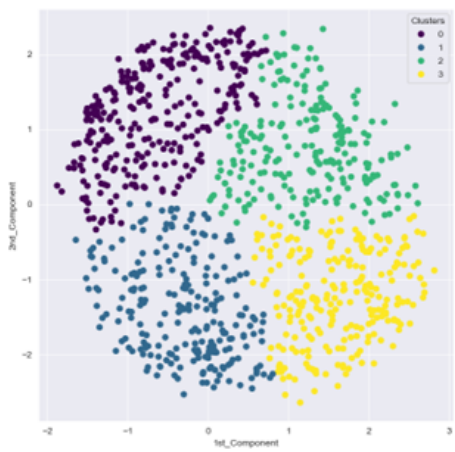

Figure 3. t-SNE embedding coloured by the four biomechanical phenotypes. Each dot represents a player.

3.2 *Classifier Performance*

Twelve variables, those in Figure 4, were retained for deployment because they sit near the top of the Random-Forest importance

ranking *and* can be captured in a low amount of time during the test in Green Room. Table 1 reports the 3-fold stratified cross-

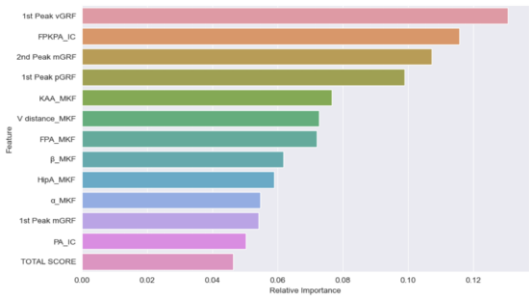

Figure 4. Relative importance of the twelve features selected for real-time use.Abbreviations: FPKPA = Frontal-Plane Knee Projection Angle, IC = initial contact, PA = pelvis angle, MKF = maximum-knee-flexion frame, vGRF/pGRF/mGRF = vertical/anterior–posterior / medio-lateral ground-reaction force, KAA = Knee-Abduction Angle, *V distance* = knee–GRF-vector distance. Total Score = Sum of the evaluations given by the doctor during the test.

validated average scores of a Random Forest ($n\_estimators = 560, random\_state = 831$) trained on the full 45-variable set. Separate experiments across random_state values 0–1000 confirmed robustness, with a standard deviation of 0.01 in F1. In contrast, Table 2 shows performance when only the twelve deployable variables are supplied. Using all 45 predictors raises macro-averaged F1 from

Table 1. Random-Forest average scores (45 features,3-fold stratified CV).

Phenotype	Precision	Recall	F1	Support
LF–LC (0)	0.95	0.94	0.94	95
LF–HC (1)	0.91	0.93	0.92	80
HF–LC (2)	0.91	0.92	0.92	76
HF–HC (3)	0.91	0.90	0.91	80

Table 2. Random-Forest average scores (12 features,3-fold stratified CV).

Phenotype	Precision	Recall	F1	Support
LF–LC (0)	0.87	0.88	0.87	95
LF–HC (1)	0.87	0.85	0.86	80
HF–LC (2)	0.83	0.79	0.81	76
HF–HC (3)	0.83	0.88	0.85	80

0.85 to 0.92, a 7-point gain that nevertheless demands more data entry time. The twelve-feature model, therefore, represents a pragmatic compromise: its accuracy remains within acceptable clinical bounds while cutting input burden to a level feasible for routine Green-Room COD assessments. The corresponding confusion matrix for the twelve-feature model is provided in Figure 5.

3.3 Clinical Workflow

The application supports the execution of a compact seven-step sequence (Figure 6):

1. **Check-in & warm-up** — the athlete registers and completes a 5-min dynamic warm-up while demographics (age, sex) are logged.
2. **COD trials** — six maximal-speed 90° cuts (three per limb) are executed over the embedded force plate while cameras record kinematics.

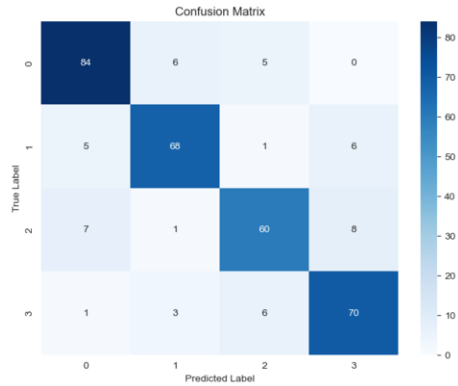

Figure 5. Confusion matrix for the twelve-feature classifier.

3. **Manual feature extraction** force and video files are opened in Gpem Screen Editor (Gpem s.r.l., Genova, Italy); the clinician reads the twelve required values (e.g. $vGRF_{max}$, $pGRF_{max}$, $FPKPA_{IC}$) and enters them into the form.
4. **Real-time classification** — pressing *Submit* invokes the Random-Forest, which returns the phenotype (LF, LC, HF, HC).
5. **Immediate feedback** an avatar with ground-reaction force arrow, and a 2×2 force–control grid appear (Figure 6b); phenotype-specific recommendations are displayed at the bottom for clinician-athlete discussion.
6. **Data archiving** an anonymised record (hashed ID, features, label) is appended to the research database, enabling periodic model retraining.

Figure 6. Application interface: (a) Before submitting, (b) After submitting.

4 Conclusion

The present work demonstrates that laboratory-grade biomechanical data can be condensed into a clinically practical tool. A Random-Forest using all forty-five available variables achieved a macro-averaged $F_1 = 0.92$; constraining the model to the twelve features F_1 to 0.85, an acceptable seven-point decline given the workflow efficiency. Embedding the classifier in an on-premise application allows seamless integration into the existing Green-Room protocol and yields an immediate, interpretable force–control phenotype for each athlete. By merging unsupervised phenotyping, a lightweight classifier, and a GDPR-compliant application, the pipeline provides a realistic path toward data-driven ACL-injury prevention in football. Future work will couple phenotype assignments to prospective injury outcomes, assess the efficacy of phenotype-specific exercise prescriptions, and automate feature extraction to eliminate manual input entirely.

References

[1] M. B. Bird, Q. Mi, K. J. Koltun, M. Lovalekar, B. J. Martin, A. Fain, A. Bannister, A. Vera Cruz, T. L. Doyle, and B. C. Nindl. Unsupervised clustering techniques identify movement strategies in the countermovement jump associated with musculoskeletal injury risk during us marine corps officer candidates school. *Frontiers in physiology*, 13: 868002, 2022.

[2] L. Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[3] A. Costa and G. Nannicini. Rbfopt: an open-source library for black-box optimization with costly function evaluations. *Mathematical Programming Computation*, 10:597–629, 2018.

[4] F. Della Villa, M. Buckthorpe, A. Grassi, A. Nabiuzzi, F. Tosarelli, S. Zaffagnini, and S. Della Villa. Systematic video analysis of acl injuries in professional male football (soccer): injury mechanisms, situational patterns and biomechanics study on 134 consecutive cases. *British journal of sports medicine*, 54(23):1423–1432, 2020.

[5] F. Della Villa, S. Di Paolo, D. Santagati, E. Della Croce, N. F. Lopomo, A. Grassi, and S. Zaffagnini. A 2d video-analysis scoring system of 90° change of direction technique identifies football players with high knee abduction moment. *Knee Surgery, Sports Traumatology, Arthroscopy*, 30(11):3616–3625, 2022.

[6] F. Della Villa, S. Di Paolo, M. Crepaldi, P. Santin, I. Menditto, L. Pirli Capitani, L. Boldrini, L. Ciampone, G. Vassura, A. Bortolami, et al. Kinematics of 90° change of direction in young football players: Insights for acl injury prevention from the cuttheacl study on 6008 trials. *Knee surgery, sports traumatology, arthroscopy*, 32(10):2666–2678, 2024.

[7] C. A. DiCesare, A. A. Minai, M. A. Riley, K. R. Ford, T. E. Hewett, and G. D. Myer. Distinct coordination strategies associated with the drop vertical jump task. *Medicine and science in sports and exercise*, 52(5): 1088, 2020.

[8] T. Dos' Santos, C. Thomas, A. McBurnie, P. Comfort, and P. A. Jones. Biomechanical determinants of performance and injury risk during cutting: a performance-injury conflict? *Sports Medicine*, 51:1983–1998, 2021.

[9] T. E. Hewett, G. D. Myer, K. R. Ford, M. V. Paterno, and C. E. Quatman. Mechanisms, prediction, and prevention of acl injuries: Cut risk with three sharpened and validated tools. *Journal of Orthopaedic Research*, 34(11):1843–1855, 2016.

[10] S. Jauhiainen, J.-P. Kauppi, T. Krosshaug, R. Bahr, J. Bartsch, and S. Äyrämö. Predicting acl injury using machine learning on data from an extensive screening test battery of 880 female elite athletes. *The American Journal of Sports Medicine*, 50(11):2917–2924, 2022.

[11] S. Lucarno, M. Zago, M. Buckthorpe, A. Grassi, F. Tosarelli, R. Smith, and F. Della Villa. Systematic video analysis of anterior cruciate ligament injuries in professional female soccer players. *The American Journal of Sports Medicine*, 49(7):1794–1802, 2021.

[12] D. Merkel et al. Docker: lightweight linux containers for consistent development and deployment. volume 239, page 2, 2014.

[13] T. Nessler, L. Denney, and J. Sampley. Acl injury prevention: what does research tell us? *Current reviews in musculoskeletal medicine*, 10: 281–288, 2017.

[14] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[15] A. Schulc, C. B. Leite, M. Csákvári, L. Lattermann, M. F. Zgoda, E. M. Farina, C. Lattermann, Z. Tősér, and G. Merkely. Identifying anterior cruciate ligament injuries through automated video analysis of in-game motion patterns. *Orthopaedic journal of sports medicine*, 12(3): 23259671231221579, 2024.

[16] H. B. Sigurðsson and K. Briem. Cluster analysis successfully identifies clinically meaningful knee valgus moment patterns: frequency of early peaks reflects sex-specific acl injury incidence. *Journal of Experimental Orthopaedics*, 6:1–12, 2019.

[17] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[18] M. Waldén, M. Hägglund, H. Magnusson, and J. Ekstrand. Acl injuries in men's professional football: a 15-year prospective study on time trends and return-to-play rates reveals only 65% of players still play at the top level 3 years after acl rupture. *British journal of sports medicine*, 50(12):744–750, 2016.