Applied Artificial Intelligence Research

INNOVIAIR

From Biomechanical Markers to Risk Prediction: Machine Learning Advances in Early Warning of Recurrent Acute Ankle Sprain

Yiming Wang[a], Siyu Chen[b], Zhendiao Lin[b,*]

[a]College of Life Sciences, Zhejiang Normal University, Jinhua, Zhejiang 321004, China

[b]College of Physical Education and Health Sciences, Zhejiang Normal University, Jinhua, Zhejiang 321004, China

ARTICLEINFO

Keywords:
Ankle Sprain
Chronic Ankle Instability
Biomechanical Markers
Machine Learning
Injury Risk Prediction
Wearable Sensors
Explainable Artificial Intelligence

ABSTRACT

Acute lateral ankle sprain is among the most frequent injuries in sports medicine, and its high recurrence rate and propensity toward chronic ankle instability (CAI) constitute a persistent clinical challenge. Conventional risk assessment—relying on subjective questionnaires, physical examination, and clinical experience—captures only part of the complex biomechanical and neuromuscular adaptations that follow an initial sprain. Recent advances in objective biomechanical profiling and machine learning (ML) have opened a new paradigm for individualized recurrence prediction. This review synthesizes the full translational chain from biomechanical marker identification to ML-based risk prediction. We first summarize key markers spanning gait kinetics (ground reaction forces and joint moments), proprioceptive and neuromuscular control deficits, and dynamic postural stability, highlighting the representational advantages of multimodal data fusion. We then compare mainstream ML architectures—including tree-based ensembles, recurrent networks for gait time series, and strategies for small-sample learning—and discuss the role of explainable artificial intelligence (XAI) in linking predictions to injury mechanisms. Evidence indicates that models integrating multimodal biomechanical features outperform conventional clinical scores (e.g., Ankle-GO, AUC 0.70), with few-shot learning systems achieving test accuracies of 0.89 and inertial-sensor-driven recurrent networks estimating ankle kinematics with coefficients of determination up to 0.93. Finally, we evaluate validation strategies, clinical utility in rehabilitation prescription and return-to-sport decisions, and wearable-based long-term monitoring, and we dissect the outstanding challenges of data standardization, model interpretability, annotation scarcity, and ethical governance. Emerging technologies—digital twins, federated learning, and generative AI for data augmentation—offer plausible routes toward a precise, interpretable, and equitable intelligent prevention ecosystem for recurrent ankle sprain.

1.Introduction

Acute lateral ankle sprain is one of the most common musculoskeletal injuries across athletic and general populations, and its clinical importance derives less from the index event than from what follows: a substantial proportion of first-time sprainers develop recurrent sprains, episodes of "giving way," perceived instability, and persistent functional deficits collectively termed chronic ankle instability (CAI). The stakes of recurrence are considerable. Mid-term follow-up data indicate that revision lateral ankle ligament reconstruction, although symptomatically beneficial, yields functional and pain outcomes inferior to those of primary modified Broström repair, underscoring that inadequate management of the index injury propagates toward more complex and less reversible pathology[1]. Epidemiologically, the injury burden extends beyond the ankle itself: recurrent instability alters lower-limb loading strategies in ways that redistribute mechanical demand toward the knee and hip, embedding the ankle within a whole-limb injury cascade.

CAI is increasingly recognized as a heterogeneous, multidimensional syndrome rather than a purely mechanical entity. A data-driven K-means clustering analysis of 210 participants identified five CAI subtypes—spanning perceived instability, functional instability, and combined presentations

* Corresponding author.
E-mail address: linzhendiao@acedu.ac.cn.
https://doi.org/10.65455/fsr64454
Received 20 August 2026; Received in revised form 24 August 2026; Accepted 25 August 2026; Available 28 August 2026

— without any purely mechanically unstable cluster, challenging conventional binary classification schemes and highlighting the need for precision phenotyping[2]. The burden also extends to the central nervous system: systematic evidence shows that individuals with CAI exhibit deficits in attention, inhibitory control, and visual memory that degrade postural control under dual-task conditions and plausibly elevate re-injury risk in the cognitively demanding environments of sport[3]. Yet even when effective preventive interventions exist, real-world translation falters. A randomized controlled trial of an unsupervised e-health-supported neuromuscular training programme in general practice found no reduction in one-year re-sprain rates compared with usual care (20.7% versus 24.1%; hazard ratio 1.14), largely because adherence was extremely low at 6.1%[4]. The recurrence problem is thus simultaneously biological, behavioral, and systemic.

Current clinical risk stratification relies heavily on patient-reported scales and examination-based composites such as the Ankle-GO score, the Cumberland Ankle Instability Tool (CAIT), and the Foot and Ankle Ability Measure (FAAM). In a prospective cohort of 64 patients, the two-month Ankle-GO score predicted one-year coper status—return to preinjury sport without re-injury or loss of function—with an area under the curve (AUC) of only 0.70; patients scoring above 11 points were substantially more likely to become copers, yet the overall discrimination remained moderate[5]. Such instruments are static, subjective, and insensitive to the compensatory movement strategies that precede frank recurrence. Biomechanical assessment offers a complementary, objective window: randomized trials demonstrate that targeted interventions produce measurable improvements in dynamic postural stability—Star Excursion Balance Test (SEBT) reach distances, sway velocity, and Sensory Organization Test scores—that track functional recovery more sensitively than symptom scores alone[6].

Machine learning, with its capacity to model high-dimensional, nonlinear, multimodal data, provides the analytical engine needed to convert these biomechanical signals into individualized recurrence predictions. A scoping review of 38 studies applying ML to sports injury prediction found that tree-based methods—random forests and gradient boosting—most often achieved the best discrimination, with AUC reported in 71% of studies, while cautioning that small samples and methodological heterogeneity limit generalization[7]. A broader scoping review of 97 studies in sports medicine similarly concluded that, despite strong in-sample performance, most models rely on retrospective data and internal validation and should presently be regarded as decision-support tools rather than autonomous systems[8]. Against this backdrop, this review traces the complete pathway from biomechanical marker identification to ML-based risk prediction for recurrent ankle sprain. We delineate candidate markers and their quantification (Section 2), compare algorithms and architectures suited to small biomechanical datasets (Section 3), evaluate model performance and validation rigor (Section 4), examine clinical translation into rehabilitation prescription, return-to-sport decisions, and wearable monitoring (Section 5), and close with the data, interpretability, and ethical challenges that must be resolved together with the emerging technologies—digital twins, federated learning, label-efficient learning, and generative AI—that may resolve them (Sections 6–7).

2.Biomechanical markers of recurrence

2.1.Gait kinetics: ground reaction forces and joint moments

Kinetic abnormalities during gait are among the most consistent biomechanical signatures of CAI. Compared with healthy controls, individuals with CAI load the posterolateral heel more heavily during the stance phase, reflecting a lateral shift of the center of pressure and elevated eversion moments during weight acceptance—a compensatory strategy that redistributes load away from the injured lateral structures yet may paradoxically increase re-injury risk by narrowing the base of dynamic support[9]. The abnormality scales with exposure in a dose-response manner: patients reporting more than five sprains within a year exhibit larger vertical ground reaction force peaks at single-leg landing, accompanied by greater ankle inversion angles and angular velocities; these kinetic features correlate significantly with anterior cruciate ligament (ACL) strain, illustrating how local ankle instability propagates mechanical demand toward the knee[10]. At the joint-moment level, the ankle acts dually as shock absorber and propulsor. During sprinting, plantar-flexor moment output is critical to dynamic stability in late stance, and altered moment patterns change energy dissipation strategies in ways plausibly linked to re-injury[11]. These kinetic markers are attractive candidates for prediction because they are objective, repeatable, sensitive to intervention, and—critically for machine learning—richly dimensional, offering dense time-series structure from which subtle pathological signatures can be learned.

2.2.Proprioceptive and neuromuscular control deficits

Deficits in proprioception and neuromuscular control are central to recurrence, and they are measurable at several physiological levels. At the spinal level, arthrogenic muscle inhibition is detectable within days of injury: patients with acute lateral sprain show a significantly reduced soleus H-reflex/maximal M-wave ratio relative to healthy controls, and the magnitude of this inhibition correlates with acute symptoms and self-reported disability, providing an early neurophysiological marker that may flag high-risk individuals before functional deficits become clinically obvious[12]. At the molecular level, dysfunction of Piezo2 mechanogated channels has been implicated in the impaired mechanosensation that underlies proprioceptive deficits after ankle sprain, linking tissue-level receptor pathology to the clinical phenotype of instability[13]. These peripheral deficits become most consequential when sensory integration is challenged: visual disruption during cutting maneuvers significantly increases ankle inversion angles in athletes after ACL reconstruction, exposing the fragility of neuromuscular control under sensory conflict—a vulnerability directly shared with lateral ankle sprain risk[14]. Central deficits compound peripheral ones, as the attentional and inhibitory impairments documented in CAI further destabilize posture during dual-task performance[3]. The layered nature of these deficits—molecular, spinal, cortical—explains why single-modality assessments underdetect risk and why marker panels spanning physiological levels are needed.

2.3.Dynamic postural stability and functional movement screening

Dynamic postural stability — the ability to regain and maintain equilibrium during movement or after perturbation—is the most clinically accessible marker domain. SEBT reach distances, particularly in the anterior and posterolateral directions, respond sensitively to targeted balance training after arthroscopic ligament repair, paralleling improvements in sway velocity and sensory integration and confirming that these functional metrics index recoverable neuromuscular capacity rather than fixed impairment[6]. Postural deficits also interact with remote central injuries: meta-analytic evidence shows that individuals with recent concussion exhibit degraded balance and gait speed, both associated with lateral ankle sprain risk, indicating crosstalk between central insults and peripheral joint stability that pure orthopedic screening misses[15]. Landing strategy constitutes another screening target with direct mechanistic interpretability — individuals with CAI who adopt increased toe-out angles at landing significantly reduce anterior talofibular ligament strain and ankle inversion angle, with a strong inverse dose relationship, identifying a precise and trainable biomechanical target for movement retraining[16].

2.4.Multimodal fusion for risk representation

No single modality captures the full risk surface of a syndrome spanning perceived, functional, and mechanical instability[2]. Gait kinetics reveal posterolateral heel overloading yet do not fully correspond to global gait-phase strategies, implying compensatory mechanisms invisible to kinetics alone[9]. Combining neurophysiological measures (H-reflex) with functional performance (SEBT) explains recovery simultaneously at the levels of neural drive and mechanical execution[6,12], and neurocognitive testing accounts for attention-dependent postural collapse that laboratory biomechanics cannot detect[3]. The practical value of fusion has been demonstrated in wearable telerehabilitation: integrating pressure insoles and inertial measurement units (IMUs) with a few-shot learning model yielded a test accuracy of 0.89 for postoperative rehabilitation assessment, significantly outperforming conventional machine-learning baselines[17]. Multimodal fusion therefore serves two predictive purposes at once—it identifies patients who appear adequate on any single dimension yet harbor latent aggregate risk, and it provides the dense feature space in which machine-learning models can discover interaction effects among neural, mechanical, and cognitive domains.

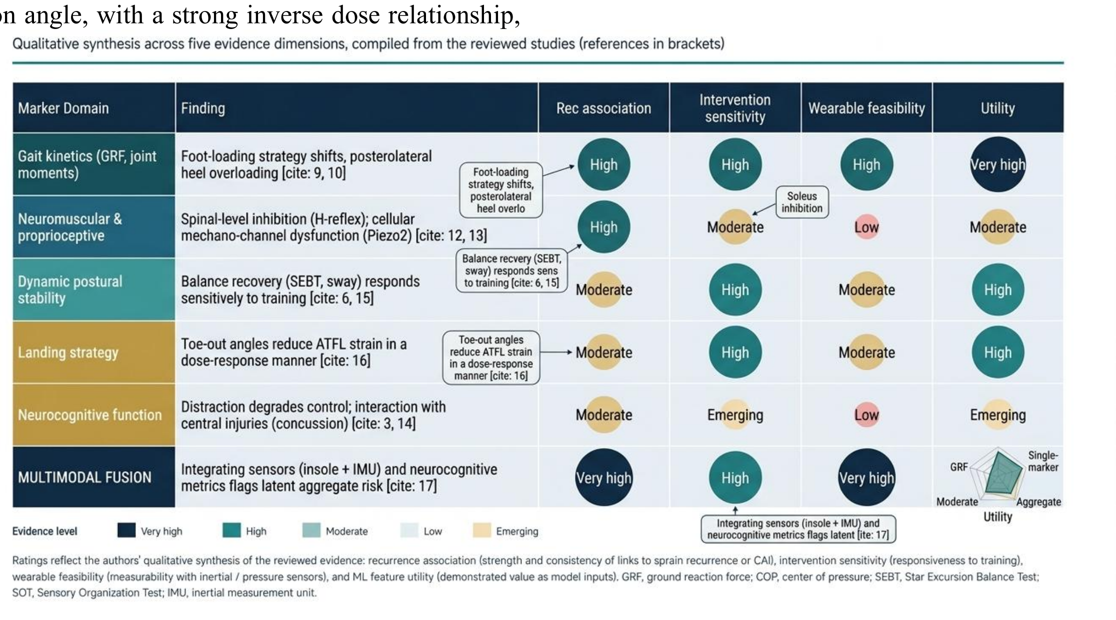

Fig 1. Evidence landscape of biomechanical marker domains for recurrent ankle sprain

3.Machine learning algorithms and architectures

Several methodological exemplars cited in this section derive from adjacent domains rather than from ankle-sprain populations—namely fall-risk prediction in older adults, postural-balance classification among the elderly, fall assessment in nursing-home residents, and generalized injury prediction in football players. None of these studies investigated ankle-sprain recurrence as an outcome; they are cited for their analytical paradigms—explainable-AI attribution, model-fairness auditing, and survival-analysis frameworks—and qualifying statements are provided at each point of use to prevent cross-domain over-extrapolation of evidence.

3.1.Supervised learning for binary recurrence prediction

Recurrence warning is typically framed as binary classification over a defined time horizon—will this individual sustain another sprain within, say, twelve months? Tree-based ensembles, principally random forests and XGBoost, most frequently lead performance rankings across sports-injury studies, and AUC serves as the dominant evaluation metric[7]. Complexity, however, is not uniformly advantageous: in four reviewed studies, logistic regression outperformed machine-learning alternatives, implying that with small samples or near-linear decision structure, parsimonious statistical models remain competitive and often more robust[7]. Integrative frameworks combining machine learning with survival analysis accommodate censored observations and time-varying risk: an elite women's football decision-support system couples hazard accumulation with probability calibration to flag fatigue-related predictors and inform player-availability decisions[18]. This framework was developed for elite women's football rather than ankle-sprain cohorts; it is cited as a methodological paradigm for integrating survival analysis with machine-learned risk modeling, not as empirical evidence on ankle-sprain recurrence. Reported discrimination varies widely—from AUC ≈ 0.75 for XGBoost-based fall-risk prediction in older adults to values exceeding 0.9 in three studies—where extreme values should prompt suspicion of overfitting or data leakage rather than celebration[7,19]. Algorithm selection must therefore weigh sample size, feature dimensionality, interpretability, and deployment constraints, not leaderboard AUC alone.

3.2.Sequential modeling of gait time series with RNN/LSTM

Gait generates strongly autocorrelated time series—three-dimensional ground reaction forces, joint angles, and joint moments—whose long-range temporal dependencies defeat conventional learners that treat observations as independent. Long short-term memory (LSTM) networks, a gated recurrent architecture, are purpose-built for this structure. A hybrid LSTM–multilayer perceptron trained on foot- and shank-mounted IMU data predicted continuous ankle kinematics and kinetics during walking and running with coefficients of determination of $R^2 = 0.89 \pm 0.04$ (walking) and 0.87 ± 0.07 (running)[20]—precision adequate to expose the subtle stance-phase pressure shifts and eversion-timing abnormalities characteristic of instability[9]. In real-time kinematic estimation, LSTM models generalize strongly within individuals (root-mean-square error as low as 1.88 ± 0.02, R^2 up to 0.93 ± 0.01), suiting personalized monitoring applications; convolutional neural network (CNN)–LSTM hybrids, which jointly extract spatial and temporal structure, generalize better across individuals and therefore suit population-level warning models[21]. This division of labor—LSTM for personalization, CNN–LSTM for population generalization—offers a principled architecture-selection heuristic for future studies.

3.3.Learning under small samples: ensembles, augmentation, and few-shot paradigms

Biomechanical datasets are expensive to acquire and typically small, and this constraint shapes the feasible model space. Ensembles such as random forests and XGBoost remain robust baselines for structured features, but their ceilings are visible in neighboring domains—AUC ≈ 0.75 for fall prediction from the electronic health records of older adults—and their performance is sensitive to cohort size, injury definitions, and outcome heterogeneity[7,19]. Deep models can compete when paired with appropriate small-sample strategies. Synthetic augmentation with an attention-based dual generative adversarial network (Attention-DualGAN) expanded gait kinematic feature sets and improved downstream detection accuracy by 2.4% for convolutional LSTM models and 7.47% for temporal convolutional networks[22]. Few-shot learning goes further: trained on gait data from only 102 CAI patients, a few-shot model reached 0.89 test accuracy and 0.90 F1 score, improving accuracy by at least 0.17 over conventional machine-learning models[17]. The lesson for this field is that deep learning is viable on realistically sized biomechanical cohorts, but conditional on augmentation or sample-efficient learning paradigms; unaided deep training on small tabular datasets remains inadvisable.

3.4.Explainable AI and mechanistic insight

Clinicians rightly distrust accurate but opaque models, and explainable AI (XAI) techniques convert predictions into mechanistic hypotheses. SHAP-based attribution applied to multi-horizon fall prediction in older adults separated acute drivers of short-term risk (syncope, respiratory symptoms) from chronic accumulators of long-term risk (spondylosis, nutritional deficiency), and exposed individual risk trajectories over time[19]. Translated to ankle sprain, feature attribution can quantify the contribution of specific markers—inversion angle at touchdown, vertical ground-reaction-force peaks, peroneal activation timing—and test mechanistic claims such as the distal-to-proximal load transfer suggested by the association between landing inversion, vertical loading, and ACL strain in frequently re-injured CAI patients[10]. Posturographic machine-learning studies in older adults demonstrate the same template: interpretable classifiers built on sway parameters can rank the biomechanical features that drive risk, rendering the model's logic auditable[23]. Both exemplars derive from geriatric fall-risk research; their role here is methodological—demonstrating interpretability workflows that ankle-sprain studies can adopt—rather than evidentiary for ankle-sprain recurrence. XAI thereby bridges prediction and pathophysiology, turning statistical association into candidate intervention targets and restoring a mechanistic narrative that clinicians can act upon.

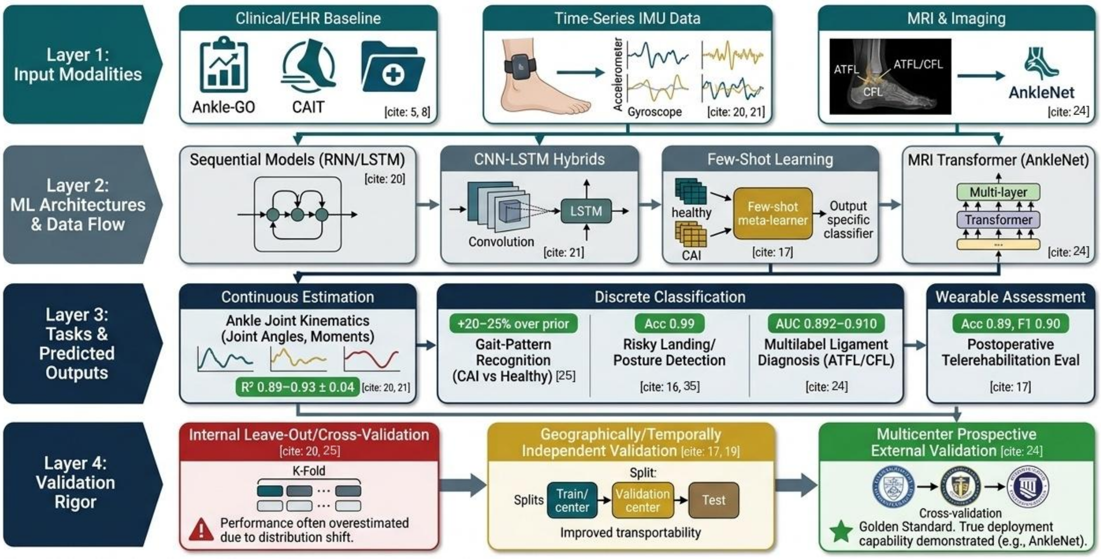

Fig 2. Machine-learning architectural taxonomy for ankle sprain risk prediction

Note: The diagram organizes input modalities, model architectures and data flow, prediction tasks and outputs, and validation rigor, spanning clinical/EHR baselines, time-series IMU data, MRI and imaging, sequential and hybrid models, few-shot learning, and MRI transformers. Metrics are not directly comparable across tasks; only the MRI-based AnkleNet underwent external validation[24].

4.Model performance, validation, and clinical utility

4.1.Validation strategies: internal cross-validation and external cohorts

K-fold and nested cross-validation dominate the field and are pragmatically appropriate for single-center, small-sample biomechanical studies—for example, the nested k-fold schemes used to evaluate LSTM-based ankle prediction across walking and running modes[20]. Internal validation, however, systematically overestimates performance when preprocessing or feature selection leaks outcome information, and it cannot detect failure under distribution shift. External validation on temporally, geographically, or institutionally independent cohorts is the decisive test. The transformer-based AnkleNet, validated retrospectively on MRI data from three independent centers, detected lateral and medial ligament injuries with AUCs of 0.910 and 0.892, respectively, outperforming internal comparators and demonstrating that cross-institutional robustness is achievable when deliberately engineered[24]. Such examples remain rare: most sports-injury models still rely on retrospective, internally validated data with little calibration reporting[8], leaving their clinical transportability unproven. Multicenter prospective external validation is the field's most urgent methodological need, and its absence should temper enthusiasm for any individual model's reported performance.

4.2.Discrimination metrics and warning thresholds

Sensitivity, specificity, and AUC carry distinct implications for tiered warning thresholds. High sensitivity suits screening applications, ensuring high-risk individuals are not missed; high specificity limits overtreatment and conserves rehabilitation resources. Conventional scores illustrate the ceiling of unimodal assessment (Ankle-GO AUC = 0.70[5]), whereas multimodal biomechanical-ML systems can deliver balanced performance—accuracy 0.89, recall 0.88, specificity 0.89, and F1 0.90 in the wearable few-shot system[17] — enabling graded probability thresholds that convert continuous biomechanical abnormality into actionable low-, intermediate-, and high-risk strata. Two cautions deserve emphasis. First, calibration, not merely discrimination, determines whether predicted probabilities can be trusted at a chosen threshold, and calibration reporting is near-universally absent from current sports-injury literature[8]. Second, threshold choice is a value judgment about the relative costs of false alarms and missed events; it should be made explicit and revisited as deployment contexts change.

4.3.Comparison with conventional clinical scores

Direct comparisons favor biomechanics-informed machine learning. Against the Ankle-GO's AUC of 0.70[5], a graph neural network with attention-based reinforcement learning (GaitNet+ARL) improved recognition of abnormal CAI gait patterns by 20-25% over existing methods[25], and GAN-augmented detectors added further gains of 2.4-7.47% in accuracy[22]. The same pattern holds in adjacent medical domains: a deep-learning radiomics pipeline fusing ultrasound-derived features with clinical data outperformed manual ultrasound assessment of hepatobiliary disease (AUC 0.93 versus 0.79), further illustrating how multimodal deep models surpass conventional unimodal assessment[26]. These advantages arise because machine-learning models extract deep temporal-spatial features — the evolution of plantar pressure across stance, moment-by-moment joint loading, intersegmental coordination — that subjective item-based scales structurally cannot capture. The shift is from expert impression to data-driven quantification, and from static snapshots to dynamic risk trajectories.

4.4.Generalizability across populations and injury subtypes

Generalizability is the principal barrier to deployment. Models trained on narrow cohorts—28 male CAI patients and 28 male controls[9], or ten healthy participants in a multimodal capture dataset[27] — may degrade in female athletes, adolescents, or older adults whose anthropometrics, gait strategies, and injury mechanisms differ systematically. Subtype heterogeneity compounds the problem: a single undifferentiated model trained across the five CAI clusters identified by data-driven phenotyping[2] risks averaging away subtype-specific signatures—neuromuscular-control-dominant risk in functional instability versus laxity-dominant risk in mechanical presentations. Context matters as well: systematic evidence on football boot design shows that footwear-surface interactions shift injury location between ankle and knee, demonstrating that identical movement biomechanics carry different risk meaning across equipment and playing-surface contexts[28]. Remedies include large multicenter datasets with explicit subtype labels, transfer learning, and domain-adaptation techniques that recalibrate models to new populations with limited local data.

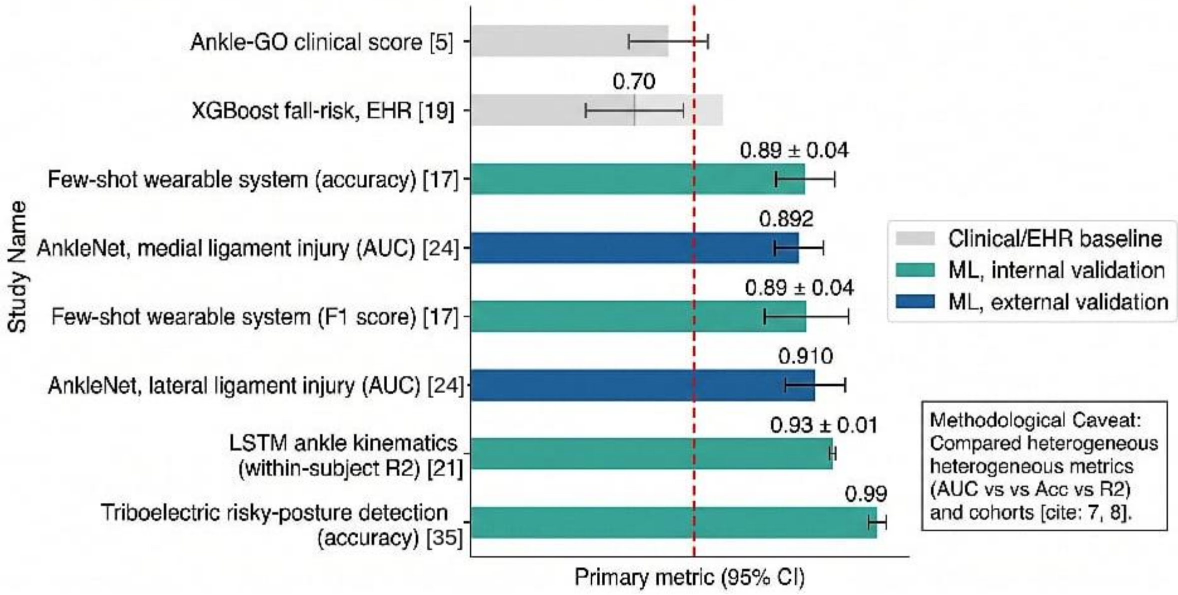

Fig 3. Reported model performance versus the clinical benchmark

Note: The dashed line marks the Ankle-GO clinical score (AUC 0.70)[5]. Bars are grouped by validation rigor; metrics differ in nature (AUC, accuracy, F1, R2) and derive from different cohorts, so the figure is not a head-to-head comparison[7,8].

5.Clinical translation

5.1.Individualized rehabilitation prescription

Prediction matters only if it changes treatment. Marker-informed stratification enables phenotype-targeted prescriptions that replace one-size-fits-all protocols. Patients flagged for posterolateral heel overloading receive neuromuscular training directed at rearfoot inversion control during weight acceptance[9]. Those with documented soleus arthrogenic inhibition receive targeted activation strategies—neuromuscular electrical stimulation or biofeedback—rather than generic strengthening, because the deficit is neural drive, not contractile capacity[12]. Landing-retraining protocols exploit the dose-dependent reduction of anterior talofibular ligament strain with increased toe-out angles[16]. Robotic platforms can deliver such prescriptions adaptively: a dual-agent multi-model reinforcement learning framework achieves human–machine co-adaptation to time-varying interaction dynamics, allowing assistance to evolve with patient performance[29], and Bayesian-optimization-based variable impedance control personalizes ankle-robot assistance in real time, improving movement speed, accuracy, and task efficiency[30]. Adjunct neuromodulation is supported by meta-analytic evidence that transcranial direct current stimulation combined with foot-core training improves balance beyond exercise alone in CAI[31]. The endpoint is a composite prescription assembled from the patient's specific deficit profile rather than from diagnostic category alone.

5.2.Return-to-sport decision support

Return-to-sport (RTS) decisions remain among the most consequential and least standardized judgments in sports medicine. Definitions range from "any participation" to full competitive return, and standardizing this terminology is a prerequisite for comparable evidence and consistent decision-making[32]. The modest predictive power of clinical composites (Ankle-GO AUC 0.70[5]) motivates decision-support systems that integrate individualized recurrence risk, sport-specific physical and cognitive demands, and biological healing windows. Practice frameworks such as the control–chaos continuum operationalize progression from highly controlled drills to chaotic, game-like exposure, with strength and power diagnostics supporting stage advancement[33]; evolved versions add visual–cognitive and attentional challenges and decision-making under simulated match pressure, directly probing the neurocognitive deficits implicated in re-injury[3,34]. The field must also reckon with implementation reality: the failure of the unsupervised e-health trial, in which adherence of 6.1% nullified a biologically plausible intervention[4], warns that decision-support systems fail without behavioral scaffolding—reminders, feedback loops, and clinician engagement—that keeps patients inside the programme.

5.3.Wearable sensing and out-of-hospital monitoring

Wearable sensors convert episodic laboratory assessment into continuous ecological surveillance, moving prevention upstream of the clinic. A self-powered triboelectric nanofiber ankle platform classifies risky ankle postures with 99% accuracy while harvesting enough biomechanical energy to drive small electronics—an enabling prototype for maintenance-free long-term monitoring[35]. In telerehabilitation, integrated pressure-insole and IMU systems with few-shot learning deliver clinician-grade quantitative assessment in the home[17]. Hybrid integration strategies—combining electrochemical, mechanical, and optical sensing—extend monitoring to physiological context, though engineering obstacles persist: sensor durability versus wearer comfort, power budgets for continuous operation, and robust multi-sensor fusion to reject environmental and motion artifacts[36]. Integrated frameworks spanning acute assessment, emergency triage, and rehabilitation guidance are already being articulated, proposing wearable-driven decision pathways across the entire injury-management continuum[37]; analogous wearable machine-learning pipelines are maturing in adjacent domains—for example, continuous blood-pressure estimation and cardiovascular risk assessment from PPG signals—further validating the feasibility of sensor-driven, out-of-hospital risk monitoring[38]. Digital care models are maturing in parallel: a noninferiority randomized trial after modified Broström surgery showed that an app-based ankle-specific training system achieved noninferior functional outcomes to face-to-face physiotherapy at lower cost[39], while the adherence failure of unsupervised programmes[4] defines the central design problem—closing the loop among patient engagement, clinician oversight, and algorithmic feedback so that monitoring translates into sustained behavior change.

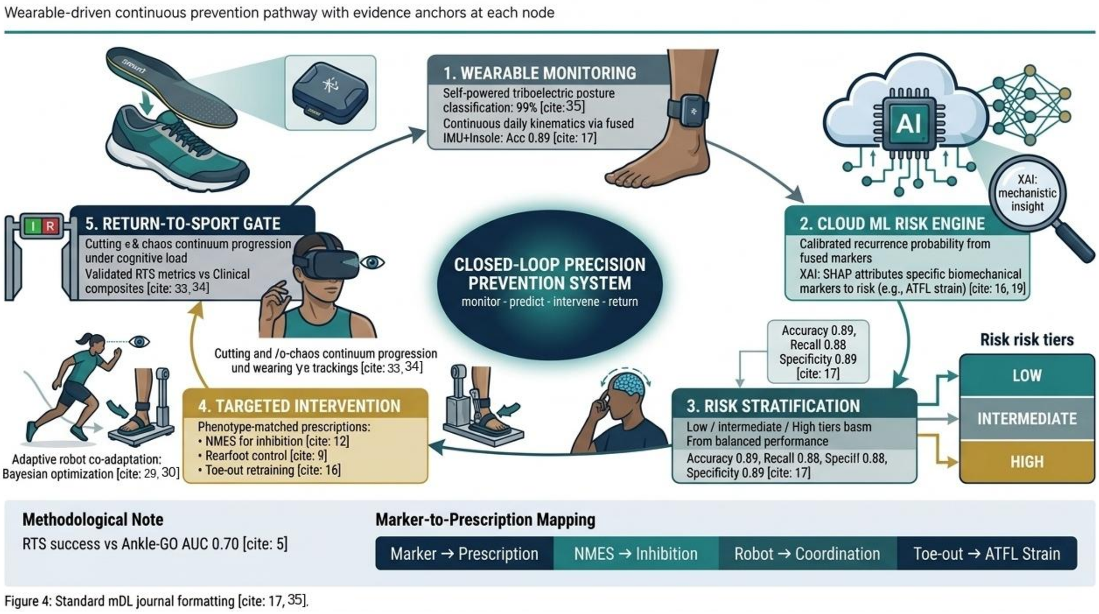

Fig 4. The wearable-ML clinical closed loop for precision ankle sprain prevention

Note: The clinical closed loop: wearable-driven continuous prevention from daily-life monitoring through machine-learning risk stratification and phenotype-matched intervention to return-to-sport clearance, with evidence anchors at each node[5,9,12,16,17,19,29,30,33–35] and a marker-to-prescription mapping band. NMES, neuromuscular electrical stimulation.

6.Challenges and ethical considerations

Four bottlenecks currently constrain translation. First, data standardization: sampling rates, signal-processing pipelines, derived-variable definitions, inclusion criteria, and outcome definitions vary across laboratories, and this heterogeneity—documented across 97 sports-medicine AI studies[8]—defeats multicenter aggregation exactly as it does in orthopedic AI more broadly[40]. Consensus protocols covering acquisition, annotation, preprocessing, and sharing are prerequisites for scale. Second, the black-box problem: without interpretability, clinicians cannot tell whether a high-risk flag reflects ligamentous laxity, neuromuscular inhibition, or measurement artifact, and prediction remains correlational rather than causal—a lesson drawn from geriatric fall-prediction research rather than ankle-sprain cohorts[19,23]. XAI methods such as SHAP help, but mechanistic validation must become a reporting norm rather than an optional supplement. Third, annotation scarcity: recurrence is a delayed, episodic outcome whose reliable labeling demands long, intensive, expensive prospective follow-up; retrospective shortcuts invite misclassification, and even intermediate phenotypes resist crisp labeling given CAI's subtyped heterogeneity[2,8]. Fourth, accountability and ethics: erroneous warnings carry real costs—unnecessary activity restriction and anxiety at one extreme, missed prevention before severe re-injury at the other—yet responsibility among algorithm developers, device manufacturers, and clinicians remains legally under-specified, and current consensus positions AI as adjunctive rather than autonomous[8,40]. Continuous biomechanical monitoring raises additional privacy concerns, and models trained on unrepresentative cohorts risk inequitable performance; federated evaluations in nursing-home fall-risk assessment have already demonstrated age-related unfairness, making fairness auditing and privacy-preserving computation mandatory design features rather than afterthoughts[41].

7.Future directions

The three technology families below are not a loosely assembled list: each responds to specific bottlenecks identified in Section 6, and we organize this section around that mapping (Table 1). Digital twins respond chiefly to the black-box challenge by rendering predictions mechanistically testable; federated learning responds to the data-standardization impasse by enabling multicenter collaboration without centralizing raw data; label-efficient learning—weakly and self-supervised paradigms—together with generative AI responds directly to annotation scarcity; and the governance framework described at the end of this section responds to the accountability, privacy, and equity challenge.

Table 1. Mapping between the four translation bottlenecks identified in Section 6 and the technology families proposed to resolve them in Section 7

Bottleneck identified in Section 6	Targeted technology family (Section 7)	Mechanism of response
Data standardization deficits	Federated learning	Multicenter training exchanges only encrypted parameters, bypassing the privacy and legal barriers to data aggregation; because gradient aggregation presupposes harmonized feature definitions and outcome encodings, federated consortia create concrete institutional pressure to adopt the consensus protocols called for in Section 6.
Black-box predictions	Digital twins (with explainable AI)	Patient-specific in-silico simulation makes predictions mechanistically inspectable and testable against physiology, converting correlational flags into causal hypotheses.
Annotation scarcity	Label-efficient learning and generative AI	Programmatic weak labeling, self-supervised pretraining on unlabeled sensor streams, few-shot learning, and GAN/diffusion-based synthetic augmentation reduce dependence on prospectively labeled recurrence outcomes.
Accountability, privacy, and equity	Governance framework and fairness auditing	Privacy-preserving computation, fairness audits, and explicit human–AI responsibility allocation under an interdisciplinary engineering standard.

Responding first to the black-box challenge, digital twins—patient-specific dynamic simulations integrating physiology, biomechanics, and environmental context—promise individualized in-silico experimentation: domain-adaptive LSTM models fusing wearable data already estimate bone stresses at the metatarsals and calcaneus during running, noninvasively flagging high-risk tissue[42], and physics-informed approaches harnessing nonlinear mechanics could render ankle instability during landing interpretable in real time[43]. Twin-based decision aids have already reached randomized evaluation in knee osteoarthritis shared decision-making, providing a clinical template for ankle applications[44].

Federated learning responds to the data-standardization bottleneck indirectly but decisively. Because only encrypted parameters or gradients are exchanged, multicenter model training no longer requires pooling raw biomechanical data, thereby bypassing the privacy and legal barriers that have so far prevented the aggregation of heterogeneously acquired datasets; federated prompt-learning frameworks achieve efficient convergence across centers even in low-resource settings[45], and quantum-assisted federated variants coupled with 5G-connected digital twins point toward real-time distributed model updating[46]. Federated learning does not abolish the need for standardization—gradient aggregation presupposes harmonized feature definitions and outcome encodings—but participation in a federated consortium creates concrete institutional pressure to adopt the consensus acquisition and annotation protocols called for in Section 6, so that federated infrastructure and standardization efforts become mutually reinforcing. Although federated training does not guarantee superior predictive accuracy, as multicenter evaluation in nursing-home fall-risk assessment has shown[41], its value for privacy, legal compliance, and the construction of large trustworthy databases is decisive for a field starved of multicenter data.

Label-efficient learning offers the most direct countermeasure to annotation scarcity. Weak supervision replaces hand-labeled outcomes with programmatic labeling functions that encode clinical heuristics—rule-based thresholds on inversion angle, ground-reaction-force asymmetry, or sway velocity—and generative models of these noisy labeling sources can produce probabilistic training labels at scale[47]. The three canonical weakly supervised regimes—incomplete, inexact, and inaccurate supervision—map naturally onto ankle-sprain datasets, in which only a fraction of prospective cohorts is completely labeled, intermediate phenotypes are coarsely defined, and self-reported recurrence is noisy[48]. Self-supervised pretraining is complementary: contrastive predictive coding over unlabeled wearable inertial streams learns transferable gait representations that sharply reduce the labeled-sample requirement for downstream risk classification[49]. Together with the few-shot and augmentation strategies reviewed in Section 3.3, label-efficient learning converts annotation scarcity from a hard ceiling into a manageable engineering constraint.

Generative AI complements label-efficient supervision from the synthesis side: GAN-based gait augmentation already improves instability detection[22], large clinical language models demonstrate zero-shot forecasting of patient trajectories from noisy electronic records[50], and fairness-aware diffusion models such as FairDiffusion aim to keep synthetic data representative rather than bias-amplifying[51]—though generative approaches bring their own challenges of bias, privacy, hallucination, and subgroup-dependent quality that require active management[52].

The accountability, privacy, and equity bottleneck, finally, is addressed not by any single model class but by governance. Realizing these gains requires an interdisciplinary engineering standard—harmonized acquisition, subtype-aware annotation, mandatory external validation and calibration reporting, interpretability requirements, fairness and privacy audits, and explicit human–AI responsibility allocation—co-authored by sports medicine, biomechanics, data science, engineering, and regulators[8,53].

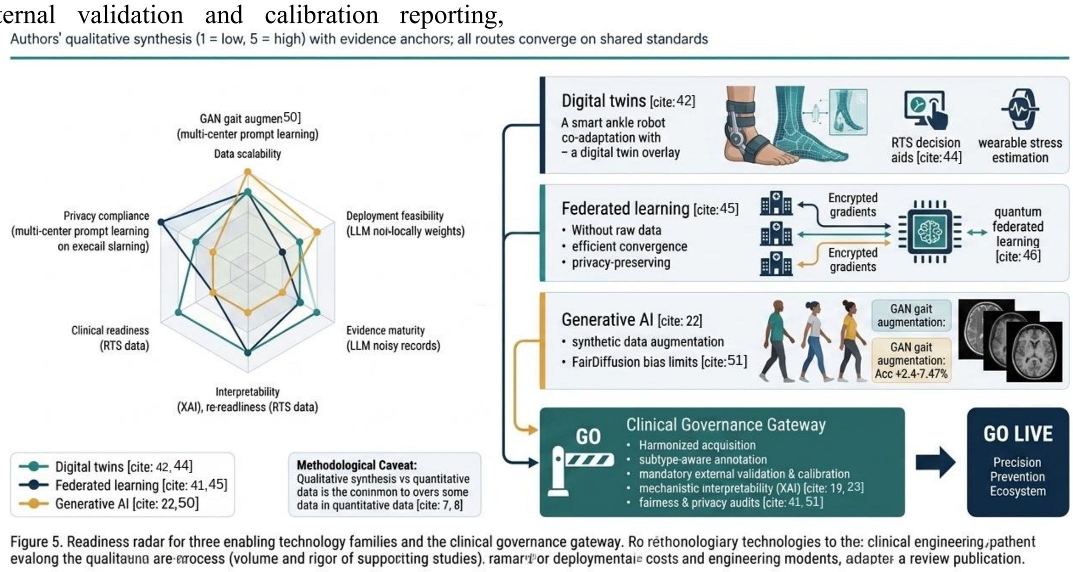

Fig 5.Readiness radar for three enabling technology families and the clinical governance gateway

8.Conclusions

The convergence of objective biomechanical phenotyping and machine learning is reshaping recurrent ankle sprain prevention. Kinetic abnormalities, neuromuscular inhibition, and postural-control deficits provide mechanistically grounded markers; LSTM- and CNN-class models extract predictive structure from gait time series at near-laboratory fidelity; few-shot and augmentation strategies make deep learning viable on realistically sized cohorts; and explainable-AI techniques reconnect prediction to mechanism. Coupled biomechanical–ML models already surpass conventional clinical scores in controlled evaluations. The decisive challenges are no longer algorithmic novelty but data standardization, prospective external validation, calibrated and interpretable risk estimates, equitable deployment, and—above all—closed-loop integration with interventions to which patients actually adhere. If these conditions are met, the field can complete its transition from experience-driven management of established instability to data-driven, precision prevention of recurrence.

References

[1] HOLLANDER J J, KERKHOFFS G M M J. Editorial Commentary: Revision Lateral Ankle Ligament Reconstruction Improves Symptoms, but Outcomes Reflect the Revision Context. Arthroscopy, 2026, 42(6): 1050-1051.

[2] LUAN L, ADAMS R, PAPPAS E, et al. Unraveling Chronic Ankle Instability: A Data-Driven Clustering Approach to Redefine Subtypes and Improve Diagnosis. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2026, 34: 894-905.

[3] CORLÙY H, SCHAMPHELEER E, MARICOT A, et al. Neurocognitive Deficits Related to Ligamentous Ankle Injuries: A Systematic Review. Sports Medicine Open, 2026, 12(1): 17.

[4] MAILUHU A K, VERHAGEN E A, VAN OCHTEN J, et al. E-health intervention for preventing recurrent ankle sprains: a randomised controlled trial in general practice. British Journal of General Practice, 2024, 74(738): e56-e62.

[5] PICOT B, FOURCHET F, RAULINE G, et al. Ankle-GO score is associated with the probability of becoming coper after lateral ankle sprain: a 1-year prospective cohort study. British Journal of Sports Medicine, 2024, 58(19): 1115-1122.

[6] ZHAO Y, PANG L, GE Y, et al. Contralateral Balance Training Improves Postural Control and Functional Recovery After Arthroscopic Anterior Talofibular Ligament Repair: A Prospective, Randomized, Single-Blind Trial. Sports Medicine Open, 2026, 12(1): 104.

[7] LECKEY C, VAN DYK N, DOHERTY C, et al. Machine learning approaches to injury risk prediction in sport: a scoping review with evidence synthesis. British Journal of Sports Medicine, 2025, 59(7): 491-500.

[8] LINDSKOG J, HEDER TERNELL K, YU Y, et al. Artificial intelligence and machine learning in sports medicine: mapping clinical tasks and assessing clinical maturity - a scoping review. BMC Medical Informatics and Decision Making, 2026, 26(1): 212.

[9] ELABD O M, ELABD A M, EL-AZEZ M S A, et al. Impact of chronic ankle instability on gait loading strategy in individuals with chronic ankle instability: a comparative study. Journal of NeuroEngineering and Rehabilitation, 2024, 21(1): 185.

[10] ZHANG Z, WEI S, SHI H, et al. Association between lower extremity movement patterns and ACL loading in CAI patients across varied ankle sprain frequencies within a year. Journal of NeuroEngineering and Rehabilitation, 2025, 22(1): 72.

[11] CROTTY E D, FURLONG L A M, HARRISON A J. Ankle and Plantar Flexor Muscle–Tendon Unit Function in Sprinters: A Narrative Review. Sports Medicine, 2024, 54(3): 585-606.

[12] KIM K M, KIM J S, NEEDLE A R. Soleus arthrogenic muscle inhibition following acute lateral ankle sprain correlates with symptoms and ankle disability but not with postural control. Journal of Sport and Health Science, 2024, 13(4): 559-568.

[13] WANG J, CHEN J, ZHAO Y, et al. Anatomy changes, signalling pathways, and clinical treatment after ankle sprain. Bone & Joint Research, 2026, 15(4): 425-436.

[14] YU P, YU Y, SUN J, et al. Effect of visual disruption on the lower limb kinematic and kinetic characteristics in athletes after anterior cruciate ligament reconstruction. Frontiers in Bioengineering and Biotechnology, 2026, 14: 1767092.

[15] CHOU T Y, HUANG Y L, LEUNG W, et al. Does prior concussion lead to biomechanical alterations associated with lateral ankle sprain and

anterior cruciate ligament injury? A systematic review and meta-analysis. British Journal of Sports Medicine, 2023, 57(23): 1509-1515.

[16] ZHU X, WEI F, LI S, et al. Toe-out landing reduces anterior talofibular ligament strain while maintains calcaneofibular ligament strain in people with chronic ankle instability. Journal of Sport and Health Science, 2025, 14: 101035.

[17] LIU H, XIE Z, GUO R, et al. Development and Validation of a Wearable Intelligent Telerehabilitation Device for Postoperative Rehabilitation of Chronic Ankle Instability Using a Portable Integrated Sensors System and Few-Shot Learning Algorithm. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2026, 34: 416-425.

[18] HUTH M, CANAL-SIMÓN B, FERRER E, et al. Injury prediction in elite women's football: an integrative machine learning-based decision-support framework. npj Digital Medicine, 2026.

[19] KHANI M, FRIEDLAND D R, WIDLANSKY M, et al. Explainable AI reveals temporal risk pathways in fall prediction: Extracting clinical insights from multi-horizon machine learning models. GeroScience, 2026.

[20] XIANG L, GU Y, GAO Z, et al. Integrating an LSTM framework for predicting ankle joint biomechanics during gait using inertial sensors. Computers in Biology and Medicine, 2024, 170: 108016.

[21] WANG L, SONG P, STONE T, et al. Ankle Kinematics Estimation Using Artificial Neural Network and Multimodal IMU Data. IEEE Journal of Biomedical and Health Informatics, 2025, 29(4): 2617-2628.

[22] LIU X, ZHANG Y, JIAO Y, et al. DAGAN-Based Gait Features Augmentation for Ankle Instability Detection. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2025, 33: 2793-2805.

[23] LIANG H W, AMERI R, BAND S, et al. Fall risk classification with posturographic parameters in community-dwelling older adults: a machine learning and explainable artificial intelligence approach. Journal of NeuroEngineering and Rehabilitation, 2024, 21(1): 15.

[24] YIN R, CHEN H, WANG C, et al. Transformer-Based Multilabel Deep Learning Model Is Efficient for Detecting Ankle Lateral and Medial Ligament Injuries on Magnetic Resonance Imaging and Improving Clinicians' Diagnostic Accuracy for Rotational Chronic Ankle Instability. Arthroscopy, 2025, 41(3): 574.

[25] GU H, YEN S C, FOLMAR E, et al. GaitNet+ARL: A Deep Learning Algorithm for Interpretable Gait Analysis of Chronic Ankle Instability. IEEE Journal of Biomedical and Health Informatics, 2024, 28(7): 3918-3927.

[26] KANG H, LI X, FENG M, et al. Applying radiomics and deep learning to investigations into seafarers' health status. Applied Artificial Intelligence Research, 2025, 1(3).

[27] HOSSAIN M S B, NGUYEN S, DRANETZ J, et al. A Wearable Motion Capture Dataset for Gait Analysis Using IMUs and Shank-Mounted Egocentric Cameras. Scientific Data, 2026.

[28] YANG F, HUANG J, WEI Q, et al. The Impact of Football Boot Design on Lower Limb Injury Risk: A Systematic Review. Sports Medicine Open, 2026, 12(1): 65.

[29] AN Y, LI Y, WANG H, et al. Human-Machine Co-Adaptation for Robot-Assisted Rehabilitation via Dual-Agent Multiple Model Reinforcement Learning (DAMMRL). IEEE Transactions on Neural Networks and Learning Systems, 2026: 1-15.

[30] MANOHARAN G, LEE H. User-Adaptive Variable Impedance Control Using Bayesian Optimization for Robot-Aided Ankle Rehabilitation. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2025, 33: 3355-3366.

[31] ZENG Z, XU Z, SHEN B, et al. Does transcranial direct current stimulation enhance sensorimotor recovery in chronic ankle instability? A systematic review and meta-analysis. Journal of NeuroEngineering and Rehabilitation, 2026, 23(1): 112.

[32] ROSINSKY P, MATZKIN E, ROSSI M J. Standardizing Definitions for Reporting Return to Sport and Athlete Levels in Sports Medicine Outcomes Research. Arthroscopy, 2026, 42(4): 691-694.

[33] TABERNER M, SPENCER N, MURPHY B, et al. Progressing On-Court Rehabilitation After Injury: The Control-Chaos Continuum Adapted to Basketball. Journal of Orthopaedic & Sports Physical Therapy, 2023, 53(9): 498-509.

[34] TABERNER M, ALLEN T, O'KEEFE J, et al. Evolving the Control-Chaos Continuum: Part 2—Shifting "Attention" to Progress On-Pitch Rehabilitation. Journal of Orthopaedic & Sports Physical Therapy, 2025, 55(3): 162-172.

[35] AKTER A, RANI G M, APU M M H, et al. Toward intelligent wearables: engineering self-powered triboelectric nanofiber sensors for the classification of risky ankle postures. Materials Horizons, 2026, 13(15): 7674-7689.

[36] ZHANG Y, ZHENG X T, ZHANG X, et al. Hybrid Integration of Wearable Devices for Physiological Monitoring. Chemical Reviews, 2024, 124(18): 10386-10434.

[37] XU M, LANG S, WANG J, et al. Emergency Clinical Decision for Sports Injury Management: A Wearable Sensor-Driven Framework from Training to Rehabilitation. Biosensors, 2026, 16(4): 205.

[38] XU Z. Continuous blood pressure estimation and cardiovascular risk assessment based on wearable PPG signals. Applied Artificial Intelligence Research, 2026, 2(1).

[39] GENG X, JIN Y, WANG X, et al. Digital Physiotherapeutic Ankle-Specific Training System for Patients With Chronic Ankle Instability Following Modified Brostrom Surgery: Noninferiority Randomized Controlled Trial at a Tertiary Grade A Trauma Center in China. JMIR mHealth and uHealth, 2025, 13: e78307.

[40] HAN F, HUANG X, WANG X, et al. Artificial Intelligence in Orthopedic Surgery: Current Applications, Challenges, and Future Directions. MedComm, 2025, 6(7): e70260.

[41] NANEVSKI I, JÄGER S, SCHULTE-ALTHOFF M, et al. The Potential of AI in Nursing Care: Multicenter Evaluation in Fall Risk Assessment. Journal of Medical Internet Research, 2025, 27: e71034.

[42] XIANG L, GU Y, DENG K, et al. Integrating personalized shape prediction, biomechanical modeling, and wearables for bone stress prediction in runners. npj Digital Medicine, 2025, 8(1): 276.

[43] NATARAJ C, HEMMATI S. Harnessing Nonlinear Mechanics to Transform Medical Diagnostics. Applied Mechanics Reviews, 2025, 77(6): 061001.

[44] JAYAKUMAR P, RATHOUZ P J, LIN E, et al. Shared decision making using digital twins in knee osteoarthritis care: a randomized clinical trial of an AI-enabled decision aid versus education alone on decision quality, physical function, and user experience. eClinicalMedicine, 2025, 89: 103545.

[45] LIN X, CHEN Y, WU J, et al. Federated generative prompt learning with vision foundation models: universal efficient multi-center medical image analysis. npj Digital Medicine, 2026.

[46] QU Z, LI Y, LIU B, et al. DTQFL: A Digital Twin-Assisted Quantum Federated Learning Algorithm for Intelligent Diagnosis in 5G Mobile Network. IEEE Journal of Biomedical and Health Informatics, 2026, 30(1): 17-26.

[47] RATNER A, BACH S H, EHRENBERG H, et al. Snorkel: rapid training data creation with weak supervision. Proceedings of the VLDB Endowment, 2017, 11(3): 269-282.

[48] ZHOU Z H. A brief introduction to weakly supervised learning. National Science Review, 2018, 5(1): 44-53.

[49] HARESAMUDRAM H, ESSA I, PLÖTZ T. Contrastive predictive coding for human activity recognition. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2021, 5(2): 1-26.

[50] MAKAROV N, BORDUKOVA M, QUENGDAENG P, et al. Large language models forecast patient health trajectories enabling digital twins. npj Digital Medicine, 2025, 8(1): 588.

[51] LUO Y, KHAN M O, WEN C, et al. FairDiffusion: Enhancing equity in latent diffusion models via fair Bayesian perturbation. Science Advances, 2025, 11(14): eads4593.

[52] TEMPLIN T, PEREZ M W, SYLVIA S, et al. Addressing 6 challenges in generative AI for digital health: A scoping review. PLOS Digital Health, 2024, 3(5): e0000503.

[53] PANAYIDES A S, CHEN H, FILIPOVIC N D, et al. Position Paper: Artificial Intelligence in Medical Image Analysis: Advances, Clinical Translation, and Emerging Frontiers. IEEE Journal of Biomedical and Health Informatics, 2026, 30(2): 1187-1202.