Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting Multi-planar Single-Leg Drop Landing Kinetic and Dynamic Stability Outcomes from Various Wearable Inertial Measurement Unit Configurations: A Machine Learning Approach with Implications for Field-Based Injury Screening

Indy Man Kit Ho

indyho@thei.edu.hk

Technological and Higher Education Institute of Hong Kong

Daniel Hung-Kay Chow

The Education University of Hong Kong

Jialiang Gu

The Education University of Hong Kong

Lana Ružić

University of Zagreb

Article

Keywords: Inertial measurement unit (IMU), Wearable sensors, Ground reaction force, Anterior cruciate ligament (ACL), Time-to-stabilization, Dynamic Postural Stability Index, Machine learning, SHAP analysis

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11431792/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Predicting Multi-planar Single-Leg Drop Landing Kinetic and Dynamic Stability Outcomes from Various Wearable Inertial Measurement Unit Configurations: A Machine Learning Approach with Implications for Field-Based Injury Screening

Indy Man Kit Ho[1,2], Daniel Hung-Kay Chow[3], Jialiang Gu[3], Lana Ružić[2]

[1] Department of Sport and Recreation, Technological and Higher Education Institute of Hong Kong, Hong Kong SAR, China
[2] Faculty of Kinesiology, University of Zagreb, 10110 Zagreb,
[3] Department of Health and Physical Education, The Education University of Hong Kong, Hong Kong SAR, China

Abstract

Background: Non-contact lower limb injuries, including anterior cruciate ligament (ACL) tears and lateral ankle sprains, frequently occur during the sudden deceleration phase of single-leg landing. While laboratory force platforms represent the clinical gold standard for quantifying landing kinetics, the cost and spatial constraints restrict widespread on-field applications. Wearable inertial measurement units (IMUs) offer a portable alternative, but whether kinematic data from single-sensor configurations can accurately predict multi-planar impact loading and dynamic postural stability remains unclear.

Methods: Forty-eight recreationally active jumping athletes (24 males, 24 females; age: 22.4 ± 2.6 years; body mass: 64.8 ± 11.2 kg) performed standardized 30 cm single-leg drop landings (SLDL), as the baseline, and multi-planar single-leg drop landing (MSLDL) tasks. Kinematic features from a 9-sensor wearable IMU array (accelerometer, gyroscope, and combined modalities across 17 sensor configurations) were synchronized with force platform kinetic data. Nine outcome variables (body-weight normalized peak vertical Ground Reaction Force (vGRF), time to peak vGRF, vertical average/instantaneous loading rates (VALR, VILR)), dynamic postural stability indices for resultant, anteroposterior, and mediolateral directions (DPSI Total, DPSI AP, DPSI ML), and time-to-stabilization using the Ross Range method for ML and AP directions (TTS Fx and Fy). Six machine learning algorithms: ElasticNet, Support Vector Regression, Random Forest, LightGBM, XGBoost, and Multilayer Perceptron (MLP) Neural Networks were trained and evaluated using subject-isolated nested cross-validation, model optimization via grid-search, with Shapley Additive exPlanations (SHAP) analysis for feature importance decomposition.

Results: A total of 5508 ML regressors were built. ElasticNet yielded the highest predictive accuracy for VALR in SLDL (R^2 = 0.815, nRMSE = 7.98%) and MSLDL (R^2 = 0.744, nRMSE = 6.44%), alongside VILR in SLDL (R^2 = 0.804), whereas gradient boosted trees (LightGBM) led MSLDL VILR (R^2 = 0.788). Conversely, MLP Neural Networks dominated Peak vGRF/BW across both SLDL (R^2 = 0.814, nRMSE = 6.61%) and MSLDL (R^2 = 0.808, nRMSE = 6.72%), as well as dynamic postural stability (DPSI Total: R^2 = 0.641 to 0.655, nRMSE ≈ 4.1%). Sensor reduction demonstrated that a single shank placement retained over 90% of predictive accuracy retention vertical loading kinetics (Peak vGRF: R^2 = 0.730; VALR: R^2 = 0.740; VILR: R^2 = 0.770), while the thigh sensor enhanced the predictions in MSLDL loading rate. SHAP analysis demonstrated that early tibial shock kurtosis, non-landing limb jerk dynamics, and the landing shank-to-pelvis phase lag were key important kinematic features for estimating impact force and dynamic stability.

Conclusions: A single shank-mounted IMU paired with regularized linear or deep neural network regression model serves as a scientifically valid and field-deployable alternative to estimate laboratory force platforms metrics during SLDL and MSLDL conditions. Practitioners may consider applying the proper single- or 2-wearable IMU sensor placements as the monitoring or biofeedback tool for rehabilitation and injury prevention. Future studies should focus on the actual values of such IMU-based landing kinetic estimations on injury risk identification and prediction.

Keywords: Inertial measurement unit (IMU); Wearable sensors; Ground reaction force; Anterior cruciate ligament (ACL); Time-to-stabilization; Dynamic Postural Stability Index; Machine learning; SHAP analysis.

1. Introduction

Lower-limb injuries, including anterior cruciate ligament (ACL) injuries, lateral ankle sprains, and patellar tendon and patellofemoral disorders, are important concerns in team and jumping sports. Chia et al. [1] reported an overall incidence of non-contact ACL injury of 0.07 per 1000 player-hours (95% CI 0.05–0.10), with higher incidence in female than male athletes. In a cohort of high-school and collegiate soccer players, Mineta et al. [2] recorded 59 lateral ankle sprains, of which 47.5% involved indirect contact; this mechanism should be distinguished from non-contact injury. A systematic review identified some associations between landing biomechanics and patellar tendinopathy, but the evidence was insufficient to establish biomechanical predictors of its onset [3]. Kim et al. [4] also reported altered lower-limb kinematics and dynamic postural stability during single-leg drop landing (SLDL) in women with patellofemoral pain. The consequences of lower-limb injury can extend beyond initial recovery: a systematic review found that only 55% of patients returned to competitive sport after ACL reconstruction [5].

Before preventive strategies are implemented, potentially modifiable injury-related factors must be characterized. Video-based reconstruction of non-contact ACL injuries in female handball and basketball players suggests that injury can occur approximately 40 ms after initial ground contact, emphasizing the importance of early landing mechanics [6]. Compared with double-leg landing (DLL), single-leg landing (SLL) can involve greater ground reaction forces and knee loading, although these differences depend on the task [7]. Dynamic postural stability provides a complementary assessment of the ability to control the body after landing. The dynamic postural stability index (DPSI) quantifies force fluctuations during this period [8], while impaired stabilization during SLDL has been reported in athletes with chronic ankle instability [9]. Together, impact loading and stabilization measures can characterize baseline performance and inform subsequent monitoring during prevention or rehabilitation programs.

Laboratory assessment of landing biomechanics commonly combines tri-axial force platforms with synchronized three-dimensional optical motion capture to quantify ground reaction forces (GRFs), segmental kinematics, and joint-level mechanical variables. However, laboratory dependence can limit repeated assessment outside controlled settings. Portable sensing methods are therefore being investigated for out-of-laboratory ACL-related assessment, including jump-landing tasks [10].

The outcomes considered here characterize complementary aspects of landing: peak vertical ground reaction force (vGRF), vertical average loading rate (VALR), vertical instantaneous loading rate (VILR), DPSI, and time-to-stabilization (TTS). DPSI and TTS assess post-landing stabilization, but the calculated TTS depends on the chosen method and processing settings [11]. Impact forces and joint loading should not be treated as interchangeable. Cadaveric landing experiments demonstrate relationships among quadriceps force, knee flexion, and ACL strain [12], while evidence on ACL injury mechanisms supports consideration of coupled sagittal-, frontal-, and transverse-plane loading [13]. Moreover, running research shows that GRF metrics do not necessarily track internal tibial bone loading across speeds and slopes [14]. Accordingly, the present kinetic and stability outcomes are biomechanical descriptors, rather than direct measurements of tissue strain or validated prospective injury-risk predictions.

Wearable measurements may complement laboratory testing by enabling repeated movement assessment in less constrained environments. Nevertheless, transfer from a standardized laboratory protocol to real sporting situations requires validation across tasks, surfaces, and athlete populations. Thus, the present investigation evaluates estimation of laboratory-derived outcomes as a prerequisite for, rather than proof of, field deployment.

Wearable inertial measurement units (IMUs) typically combine accelerometers and gyroscopes, sometimes with magnetometers, in lightweight devices that capture segmental movement [15]. IMU-based methods have been evaluated for lower-limb kinematic estimation during treadmill walking and running [16], and for countermovement-jump assessment [17]. Tibial acceleration measurements are influenced by sensor attachment, movement characteristics, and the measurement environment, which must be considered when interpreting impact signals [18]. Machine-learning approaches

have also estimated continuous GRFs during running [19] and vGRF and knee extension moments during drop landings [20].

Estimating landing kinetics from wearable kinematics is challenging because an IMU measures acceleration and angular velocity at its attachment site, rather than whole-body center-of-mass (COM) acceleration directly. Segmental motion, sensor orientation, attachment, and impact attenuation can therefore influence the mapping from IMU signals to force-platform outcomes [18]. A sensor-reorientation study illustrates that changes in measured tibial acceleration do not necessarily produce corresponding improvements in its relationship with GRF measures [21]. These limitations motivate models that combine multiple signal features rather than relying on one acceleration peak. For example, Derie et al. [22] estimated maximal vertical loading rate during overground running from bilateral tibial accelerometer features using regression models.

Machine learning (ML) offers a range of candidate mappings between wearable kinematics and kinetic outcomes, including regularized linear regression, support vector regression (SVR), tree ensembles such as random forest (RF), XGBoost and LightGBM, and neural networks such as multilayer perceptrons (MLPs) and long short-term memory (LSTM) networks. Their performance depends on the data, prediction target, and evaluation design; methodological guidance emphasizes careful model training and validation in human movement biomechanics [23]. Shapley Additive Explanations (SHAP) can complement predictive evaluation by attributing model outputs to input features [24]. Such attributions describe the fitted model and should not be interpreted as causal effects of biomechanical variables.

Previous studies demonstrate the potential of IMUs for kinetic estimation, but their tasks and targets differ. Sun et al. [20] used eight IMUs and modular LSTM models to estimate vGRF and knee extension moments during single- and double-leg drop landings; mean vGRF R^2 was 0.88 for single-leg and 0.85 for double-leg landings. In steady-state running, Song et al. [25] found that lightweight SVD-embedding and k-nearest-neighbor regression methods could perform comparably to, or better than, LSTMs. A systematic review further documents the range of wearable-sensor ML applications in running biomechanics [26]. However, evidence from running or conventional drop landings does not establish performance during the combination of directional and rotational SLDL tasks examined here. The ability to estimate impact kinetics and dynamic stabilization outcomes during these multi-planar single-leg drop landings (MSLDL) therefore remains an important task-specific question.

Moreover, critical methodological questions remain regarding wearable sensor deployment. First, the optimal anatomical sensor placement (shank vs. foot dorsum vs. pelvis vs. thigh vs. wrist) for estimating kinetic domains has not been systematically resolved. Second, while previous investigations have attempted to predict absolute impact forces in Newtons, predicting normalized ground reaction forces (vGRF/BW) represents a biomechanically superior paradigm, as it decouples anthropometric variance, (i.e., body mass) from true neuromuscular landing compliance. Finally, no prior investigation has comprehensively benchmarked whether a small number of, or even a single IMU, can simultaneously estimate kinetic metrics at both landing impact phase (peak vGRF and VALR) and dynamic postural stability metrics during the stabilization and quiet standing phases (DPSI and TTS) across MSLDL. By illustrating the trade-off between sensor quantity (9, 7, 4, 2, and 1) and prediction accuracy, this study can provide empirical evidence and insights to scientists and practitioners for achieving an optimal balance between the practical feasibility of IMU deployments and the precisions of landing kinetics variables.

Therefore, the purpose of this study was threefold: (1) to compare six machine learning algorithms (ElasticNet, SVR, XGBoost, LightGBM, RF, and MLP) in predicting laboratory-derived landing kinetics and dynamic stability indices from selected wearable IMU kinematic features during SLDL tasks; (2) to systematically compare predictive efficacy across varying sensor quantities (9, 7, 4, 2, and 1 sensor arrays), sensor modalities (accelerometer only [A]; gyroscope only [G], and combined [A+G]) and anatomical placements (pelvis, thigh, shank, foot, and wrist) across distinct kinetic domains; and (3) to apply SHAP feature importance decomposition to the top-performing models to identify the key biomechanical determinants driving accurate predictions. Grounded in the Newton's second law (F = ma), which establishes a direct physical relationship between mass deceleration during landing and impact force, we hypothesized that the regularized

linear regression (ElasticNet) and gradient-boosted ensembles would produce strong prediction accuracy for vertical kinetic output variables when using A sensors only or combined sensor modalities (A+G). Conversely, given the non-linear and complex multi-segmental interactions and controls governing dynamic postural stabilization and the temporal characteristic of TTS (measured in seconds), we hypothesized that non-linear algorithms (i.e., neural network-based MLP and tree ensembles) would outperform linear models for dynamic stability-related metrics, with combined sensor modalities (A+G) demonstrating superiority over isolated (A) or (G) sensor configurations, albeit with lower overall prediction accuracy than vertical kinetics. Finally, we also hypothesized that the shank sensor placement would yield the best estimation for vertical loading kinetics while the foot sensor would produce superior frontal plane stability-related prediction.

2. Methods

2.1 Participants and Ethical Approval

A cohort of 48 recreationally active collegiate athletes (24 males, 24 females; mean ± standard deviation: age = 22.4 ± 2.6 years; height = 1.71 ± 0.09 m; body mass = 64.8 ± 11.2 kg; body mass index = 22.0 ± 2.4 kg/m²) volunteered to participate in this repeated-measures study. All participants engaged in competitive or structured recreational jumping-related sports (basketball, soccer, volleyball, or track and field) for a minimum of 4 hours per week over the preceding 12 months. Exclusion criteria comprised: (1) any history of major lower limb joint surgery (including ACL reconstruction, meniscal repair, or ankle ligament reconstruction); (2) any acute lower extremity musculoskeletal injury sustained within the preceding six months; (3) any diagnosed vestibular, cognitive, mental, neurological, or balance impairment; or (4) current lower limb pain or deformity that could alter natural landing mechanics. Prior to experimental testing, all participants provided written informed consent. The experimental protocol was reviewed and formally approved by the Human Subjects Ethics Committee (HSEC) of Technological and Higher Education Institute of Hong Kong, as the Institutional Review Board (IRB), strictly adhering to the ethical principles outlined in the Declaration of Helsinki.

2.2 Experimental Protocol and Single-Leg Drop Landing Paradigm

Testing was conducted in the biomechanics laboratory. Prior to data collection, participants completed a standardized dynamic warm-up consisting of 3-minute self-paced stationary cycling (60-70 RPM at 1.0 kp resistance) followed by dynamic lower limb mobility and plyometric exercises (body weight squats, forward and lateral lunges, hip rotation mobility exercises and double-leg pogo jump for 10 repetitions per each exercise). Participants were then familiarized with all SLDL tasks through structured practice trials until they demonstrated consistent, confident execution without severe loss of balance.

All experimental tasks required participants to perform SLDL from a standardized 30 cm wooden platform onto an embedded laboratory force platform. Participants stood barefoot on the wooden platform on their non-dominant leg (dominant leg for landing was defined as the preferred limb used to take off during a jumping task). To mimic a natural landing movement, instead of placing the hands on their iliac crests, subjects were allowed to use the arm movements to assist in their balance throughout the entire descent, landing impact, and stabilization sequence. There were two SLDL scenarios.

Scenario 1 – SLDL: this was regarded as the baseline reference. Participants dropped vertically off the 30-cm platform, initiating the free fall without jumping upward or stepping downward, and landed centrally on the force platform with the dominant leg. During the entire drop landing sequence, participants needed to look forward to focus on a target 4 meters apart. The SLDL was selected as a standardized baseline as SLDL onto force platforms are commonly used to quantify landing biomechanics [27].

Scenario 2 – MSLDL: participants were required to complete a total of six directional conditions (medial, lateral, antero-medial, antero-lateral, 90-degree internal rotation, and 90-degree external rotation) in a randomized order. The horizontal

distance between the near edge of the force platform and the wooden platform was set at 30% of the body height. Previous work has assessed ankle biomechanics during forward, diagonal, and lateral single-leg landings [28]. The present protocol used a study-specific horizontal distance of 30% of participant body height to accommodate the directional and rotational conditions. This distance was not treated as a replication of the earlier protocol. All MSLDL conditions were initiated by the non-dominant leg similar to the scenario 1. For medial and lateral conditions, participants hopped horizontally (medial or lateral to the dominant landing leg) without observable upward bouncing movement. For antero-medial and antero-lateral conditions, they hopped at 45 degrees diagonally (antero-medial or antero-lateral to the dominant landing leg) without observable upward bouncing movements. For 90-degree internal or external rotation drop landing, participants stood and faced 90-degree from the target. They hopped and rotated the body internally or externally towards the force platform during the flight phase and finally landed with their dominant leg with the eyes facing the same target as all other landing tasks. Besides the mentioned landing issues, participants were required to redo the rotational condition if they have completed most rotational movement already before the initiation of the hopping action. This was to assure a true rotational action rather than a "pre-rotated vertical drop landing".

Upon initial ground contact and the eccentric lower phase, participants had to maintain the motionless and stable single-leg quiet stance for 15 seconds meanwhile looking forward at the target. A trial was deemed unsuccessful and immediately discarded if the participant: (1) allowed the non-dominant contralateral foot to make contact with the floor or testing limb ('touch-down' error); (2) exhibited observable hopping, shuffling, or gross repositioning of the landing foot; (3) landing foot was fully or partly outside the force platform; or (4) failed to maintain single-leg balance for the full 15-second duration. Participants performed three successful trials for each SLDL and MSLDL conditions whereas, the scenario 1 for baseline was completed first followed by the randomized order of the six MSLDL conditions. To prevent cumulative neuromuscular fatigue, a one-minute resting period was given between trials while a 2-minute rest was provided between experimental conditions.

2.3 Instrumentation and Experimental Setup

Ground reaction forces (GRF) data were captured at a sampling rate of 1000 Hz using an embedded piezoelectric multi-component force platform (Advanced Mechanical Technology, Inc., United States) measuring 600 mm × 400 mm. Segmental kinematic data were simultaneously recorded using wireless inertial measurement units (Noraxon, Ultium Motion, Scottsdale, Arizona) operating at a continuous sampling frequency of 200 Hz. Each IMU unit integrated a high-bandwidth tri-axial accelerometer (range: ±16g), a tri-axial gyroscope (range: ±2000°/s), and a micro-controller housed within an ultra-compact casing (mass < 15 g). Nine anatomical sites were referenced from the recommended positions provided by the manufacturer's guidelines (1): the pelvis was positioned over the superior body area of the sacrum, with its center aligned approximately with the posterior superior iliac spine; (2 and 3): the thigh sensors were placed on the anterior aspect and distal half of each thigh, approximately 7.5 to 10 cm (3-4 inches) proximal to the knee joint, over a relatively stable soft tissue region with minimum muscle displacement and movement artefacts during motion; (4 and 5): the shank were positioned on the anteromedial aspects on the flat surface of the tibia, approximately to 5 to 7.5 cm (2-3 inches) below the tibial tuberosity and align with the calf muscle belly, where the bone is superficial with minimal intervening muscle bulk to optimize sensor attachment; (6 and 7): both feet (bare feet) sensors were secured over the dorsal aspect of the midfoot slightly below the ankle aligning with the second and third metatarsals; (8 and 9): both wrist sensors were positioned on the dorsal aspect of the middle hand aligning with the third metacarpals. To eliminate unnecessary sensor tilting, sensor displacement and soft-tissue vibration artifacts, all IMUs were securely attached directly on the skin of the selected body regions with the tailor-made double-sided tape.

Prior to the testing session, a dynamic sensor calibration was executed through a walking task in accordance with manufacturer specifications. Regular check on the sensor placement for any abnormalities and drifting was performed whereas another dynamic sensor calibration was performed for any abnormal drifting identified. Force platform and IMU data were synchronized via sync box with the setup based on the instruction from the manufacturer. The exported files

automatically integrated both IMU and force platform data by upscaling the IMU portion to 1000 Hz and therefore, each data file contained 15,000 rows of data points for the entire 15-second landing sequence.

2.4 Data Processing and Feature Selection

The landing phase was automatically parsed into: 1) Impact phase (Initial Contact [IC] to 250 ms); 2) Stabilization phase (0.25 to 3.0 s post-IC); 3) Quiet standing phase (from 3.0 to 15.0 s) (Fig. 1). Raw force platform and IMU signals were processed via dual-pass filtered using a zero-phase-lag, fourth-order Butterworth filter. For kinetic domain variables (i.e. vGRF/BW, VARL, VILR, DPSI) the cut-off frequency of 50 Hz was used. Regarding the TTS and G sensors variables (angular velocity), the cut-off frequency of 15 Hz was applied, whereas, 20 Hz cut-off frequency was implemented for A sensor variables (i.e., linear acceleration). All data processing, filtering, machine learning analytics and visualization were performed in Python 3 using Jupyter Notebook.

In the original research project, in accordance with clinical and sports medicine-related literature, a total of 52 potential variables including the kinetic and impact domain, postural control and stability domain, and power spectrum variables were obtained. This study was part of the original project and therefore, only nine biomechanical targets output variables relevant to the research objectives were selected for subsequent data analytics.

Peak Vertical Ground Reaction Force Normalized to Body Weight (Peak vGRF labelled as "Target_Peak_vGRF_BW"): Defined as the maximum vGRF recorded during the impact phase (0-250 ms post-IC), normalized by the participant's body weight (BW = mass × 9.81 m/s²), expressed in BW. Initial contact (IC, a.k.a. t = 0) was defined as the first instant where vGRF > 10 N threshold.

Time to peak vGRF (Target_Time_To_Peak_vGRF_s): defined as the shortest time from IC to the peak vGRF.

$$i_{\text{peak}} = \arg \max_{0 \le i < 0.25 \cdot f_s} F_z[i]$$

$$t_{\text{peak}} = \max\left(\frac{i_{\text{peak}}}{f_s}, \frac{1}{f_s}\right)$$

Vertical Average Loading Rate (VALR labelled as "Target_VALR_BW_s"): Calculated as the slope from IC to the instant of peak vGRF: VALR = (Peak vGRF_BW) / (Time to Peak vGRF), expressed in BW/s.

$$\text{VALR}_{\text{BW}} = \frac{F_{z,\text{peak}}/W}{t_{\text{peak}}} = \frac{\text{vGRF}_{\text{peak,BW}}}{t_{\text{peak}}}$$

Vertical Instantaneous Loading Rate (VILR labelled as "Target_VILR_BW_s"): Defined as the maximum single-interval slope of impact force curve, (max (dFz/dt) within the initial impact phase (≤0.25 s), expressed in BW/s [22].

$$\text{VILR}_{\text{BW}} = \frac{\text{VILR}_{\text{abs}}}{W} = \frac{1}{W} \max_{0 \le t \le 0.25\ \text{s}} \left(\frac{dF_z(t)}{dt}\right)$$

Anteroposterior Dynamic Stability Index (DPSI AP labelled as "Target_DPSI_AP_y"): The isolated sagittal plane

component of the DPSI formulation, reflecting shear instability. It was calculated by: DPSI AP = sqrt((sum(Fy²) / N_samples) / BW, where Fy represents anteroposterior component of ground reaction force at each time point.

$$\mathrm{DPSI}_{\mathrm{AP}} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(F_y[i]\right)^2}}{W}$$

Mediolateral Dynamic Stability Index (DPSI ML labelled as "Target_DPSI_ML_x"): The isolated frontal plane component of the DPSI formulation, reflecting shear instability. It was calculated by: DPSI ML = sqrt((sum(Fx²) / N_samples) / BW, where Fx represents mediolateral component of ground reaction force at each time point.

$$\mathrm{DPSI}_{\mathrm{ML}} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_x[i])^2}}{W}$$

Dynamic Postural Stability Index (DPSI Total labelled as "Target_DPSI_Total"): defined as the composite 3D dynamic postural stability index. It was evaluated over the first 3.0 seconds post-IC in accordance with the formulation of Wikstrom et al. [8]: DPSI Total = sqrt((sum(Fx²) + sum(Fy²) + sum((Fz - BW)²)) / N_samples) / BW, where Fx, Fy, and Fz represent mediolateral, anteroposterior, and vGRF at each time point, respectively.

$$\mathrm{DPSI}_{\mathrm{Total}} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[F_x[i]^2 + F_y[i]^2 + (F_z[i] - W)^2\right]}}{W}$$

6. Time-to-Stabilization (TTS) in the mediolateral and anteroposterior (TTS Fx labelled as "Target_TTS_RossRange_Fx_s" and TTS Fy labelled as "Target_TTS_RossRange_Fy_s"): were quantified using the adapted Ross Range 3rd order polynomial curve-fitting method [11, 29]. The individualized baseline stability bandwidth was established via the quiet single-leg standing phase in the final 10 seconds. The peak-to-peak force range (max – min) was extracted across two consecutive blocks of 5-second windows (5.0-10.0 s and 10.0-15.0 s). The minimum peak-to-peak range was identified as the baseline tremor amplitude, and subsequently, the stabilization threshold was defined as 50% of this minimum range to match the amplitude of rectified deviations. After that, the primary impact shear peak was determined within the initial 500 ms post-IC to isolate the landing deceleration from subsequent postural sway, and the signal from peak shear onward was rectified around the mean baseline force of the final 5-second quiet stance period. A 3rd order polynomial curve was fitted to this rectified post-peak signal using least-squares regression. TTS was defined as the earliest time point at which the fitted polynomial trajectory decayed below the threshold boundary. Let $F_\tau(t)$ denote the low-pass filtered (15 Hz) horizontal shear force along axis $\tau \in \{x, y\}$:

$$\mathrm{Range}_{\mathrm{min},\tau} = \min\left(\mathrm{ptp}\left(F_\tau[5.0\ \mathrm{s} \le t \le 10.0\ \mathrm{s}]\right),\ \mathrm{ptp}\left(F_\tau[10.0\ \mathrm{s} \le t \le 15.0\ \mathrm{s}]\right)\right)$$

$$T_\tau = \frac{\mathrm{Range}_{\mathrm{min},\tau}}{2}$$

$$F_{\mathrm{base},\tau} = \frac{1}{N_{\mathrm{quiet}}}\sum_{t=10.0\ \mathrm{s}}^{15.0\ \mathrm{s}} F_\tau\,(t)$$

$$y_\tau(t) = \left|F_\tau(t) - F_{\text{base},\tau}\right|, \quad \text{for } t \geq t_{\text{peak},\tau}$$

$$\hat{y}_\tau(t') = \beta_3(t')^3 + \beta_2(t')^2 + \beta_1(t') + \beta_0, \quad \text{where } t' = t - t_{\text{peak},\tau}$$

$$\text{TTS}_\tau = t_{\text{peak},\tau} + \min\{t' \geq 0 \mid \hat{y}_\tau(t') \leq T_\tau\}$$

2.5 IMU Kinematic Feature Engineering Pipeline

Kinematic features were extracted from the continuous IMU tri-axial acceleration and angular velocity signals across all nine IMU sensors (pelvis, thigh L/R, shank L/R, foot L/R, wrist L/R). Both accelerometer resultant ($a_{\text{res}} = \sqrt{a_x^2 + a_y^2 + a_z^2}$) and gyroscope resultant ($\omega_{\text{res}} = \sqrt{\omega_x^2 + \omega_y^2 + \omega_z^2}$) were used for calculating the segmental kinematic time-domain features (per segment, e.g., thigh) and kinetic chain shock attenuation and phase lag features along the dominant and non-dominant chains.

A total of 153 segmental kinematic features, from A (9 features per segment) and G (8 features per segment due to no TTS derived) sensors per each body segment are shown in Table 1.

Table 1 – Segmental kinematic time-domain features (per segment, e.g., thigh)

Feature name	Mathematical formula	Biomechanical meaning
Feat_[Seg]_Acc_Peak	$\max(a_{\text{res}}(t)))$	Peak impact shock
Feat_[Seg]_Acc_RMS	$\sqrt{\frac{1}{N}\sum a_{\text{res}}^2}$	Root-mean-square magnitude of segmental acceleration over the 15-second window
Feat_[Seg]_Acc_Mean	$\frac{1}{N}\sum a_{\text{res}}(t)$	Mean acceleration
Feat_[Seg]_Acc_Peak_Jerk	$\max\left(\left\|\frac{da_{\text{res}}}{dt}\right\|\right)$	Maximum rate of acceleration change, potentially reflecting abrupt impact onset or rapid deceleration
Feat_[Seg]_Acc_Jerk_RMS	$\sqrt{\frac{1}{N}\sum\left(\frac{da_{\text{res}}}{dt}\right)^2}$	Continuous fluidity of segmental motion damping
Feat_[Seg]_Acc_Kurtosis	$\frac{\frac{1}{N}\sum(a-\mu)^4}{\sigma^4}$	Sharpness of impact transients versus quiet sway background
Feat_[Seg]_Acc_Crest_Factor	$\frac{\text{Peak}}{\text{RMS}}$	Impulsiveness of landing shock relative to the participants' sustained movements
Feat_[Seg]_Acc_SampEn	$-\ln\left(\frac{A}{B}\right)$	Complexity or irregularity of segmental acceleration;

		interpreted relative to task and population
Feat_[Seg]_Acc_TTS_s	Sequential threshold	TTS as time for resultant acceleration to enter and remain within a baseline threshold band within ± 0.25 SD of quiet stance baseline
Feat_[Seg]_Gyr_[Metric]	Same metrics on ω_{res}	Rotational analogues: peak angular rate, angular-rate RMS, angular jerk, kurtosis, crest factor, and entropy.

Kinetic chain shock attenuation (dB) for both dominant and non-dominant sides, were to quantify the reduction in peak segmental acceleration from distal to proximal sites. A higher positive value shows greater attenuation of the selected impact transient during proximal transmission whereas values close to 0 suggest limited attenuation were calculated. The inter-segmental phase lag features (ms) were to quantify the time interval between selected distal and proximal peak-acceleration events. A shorter lag reflects more rapid proximal transmission of the impact-related acceleration transient while a longer lag may indicate delayed or distributed loading, potentially consistent with greater joint compliance.

$$\text{Feat_Shock_Atten_Shank_to_Thigh_dB} = 20\log_{10}\left(\frac{\max(a_{\text{shank}})}{\max(a_{\text{thigh}})}\right)$$

$$\text{Feat_Shock_Atten_Thigh_to_Pelvis_dB} = 20\log_{10}\left(\frac{\max(a_{\text{thigh}})}{\max(a_{\text{pelvis}})}\right)$$

$$\text{Feat_Shock_Atten_Total_Foot_to_Pelvis_dB} = 20\log_{10}\left(\frac{\max(a_{\text{foot}})}{\max(a_{\text{pelvis}})}\right)$$

$$\text{Feat_Phase_Lag_Foot_to_Pelvis_ms} = \left(t_{\text{peak}}(\text{Pelvis}) - t_{\text{peak}}(\text{Foot})\right) \times 1000$$

$$\text{Feat_Phase_Lag_Shank_to_Pelvis_ms} = \left(t_{\text{peak}}(\text{Pelvis}) - t_{\text{peak}}(\text{Shank})\right) \times 1000$$

2.6 Machine Learning Algorithms and Cross-Validation Structure

Six ML algorithms were performed to predict selected scalar target variables derived from kinetic data: (1) Elastic Net Regression (ElasticNet): a regularized generalized linear model that combines L1 (Lasso) and L2 (Ridge) penalties to address multicollinearity issues among kinematic features and promote feature sparsity; (2) Support Vector Regression (SVR): a kernel-based algorithm using a radial basis function (RBF) kernel to model non-linear relationships in a high-dimensional feature space with an epsilon-insensitive loss function; (3) Random Forest (RF): An ensemble bagging method comprising regression decision trees trained on bootstrapped subsets of the data; (4) LightGBM (Gradient Boosted Decision Trees): a leaf-wise gradient-boosted decision-tree framework that uses histogram-based splitting and gradient-based one-side sampling (GOSS) for efficient modeling of non-linear tabular data; (5) XGBoost, a scalable gradient-boosted decision-tree algorithm; and; (6) a multilayer perceptron (MLP): a feedforward artificial neural network comprising two hidden layers of 128 and 64 neurons, rectified linear unit (ReLU) activations, batch normalization, and Adam optimization. Both ElasticNet, SVR, RF, LightGBM and XGBoost were implemented using scikit-learn in Python 3 within a Jupyter Notebook environment, whereas the MLP was performed using PyTorch.

To minimize overfitting and prevent data leakage problems, models were trained and evaluated using subject-isolated nested cross-validation (CV) with a 5-by-5 outer-inner scheme. Among the 48 participants, 80% (n=38 or 39) were assigned to the training and tuning sets while the remaining 20% (n=9 or 10) were held out as the test set in each outer fold. Within each outer-training iteration, an inner 5-fold group K-fold CV procedure was used for hyperparameter optimization via grid search were performed exclusively within the outer-training data, thereby preventing information from the held-out test set from altering model development. Following inner fold optimization, the best-performing configuration was retrained on the pooled outer training set and evaluated on the completely unseen held-out test set.

2.7 Feature Preprocessing, Model Optimization and Performance Metrics

Traditional linear transformation, such as min-max scaling can be sensitive to extreme transient impact shock spikes and can potentially compress the range of typical non-outlier value. Therefore, continuous kinematic inputs for ElasticNet and SVR were scaled using "RobustScaler", which centers each feature on its median and scales it according to its interquartile range. Tree-based algorithms (RF, LightGBM and XGBoost) were trained using unscaled features as tree-based splitting and histogram-based binning do not require feature scaling.

To balance model optimization, robustness and computational efficiency, grid search was conducted over selected anchor points hyperparameter ranges:

ElasticNet: Regularization penalty $\alpha \in \{0.01,0.1,1.0,10.0\}$ and L1 mixing ratio $\rho \in \{0.2,0.5,0.8\}$ were evaluated across 12 combinations to balance feature sparsity against collinear predictors.

Support Vector Regression (SVR): An RBF kernel with an automatically variance-scaled coefficient (γ = 'scale') was adopted with regularization parameter $C \in \{1.0,10.0\}$ and epsilon insensitivity threshold $\varepsilon \in \{0.01,0.1\}$ across 4 states.

Random Forest: Tree depth was defined by $d_{max} \in \{8,14\}$ with minimum split criteria $s_{min} \in \{2,5\}$, for evaluating 4 configurations using 100 estimators and 80% bootstrap subsampling (max_samples=0.8).

LightGBM: Leaf-wise structural complexity was tuned using maximum leaves $L \in \{15,31\}$ and learning rate (shrinkage factor) $\eta \in \{0.05,0.1\}$ across 4 configurations with fixed 100 boosting iterations.

XGBoost: Tree complexity was tuned via maximum depth $d_{max} \in \{3,5\}$, learning rate $\eta \in \{0.05,0.1\}$, and L2 leaf regularization $\lambda \in \{1.0,5.0\}$ across 8 combinations using 100 trees.

Regarding the MLP, a similar nested CV pipeline with outer 5-fold and inner 3-fold group K-fold CV was used to balance model optimization capacity and reasonable computational efficiency. A discrete grid search evaluated three combinations as follow:

1. A heavily regularized, compact network ($d_{hidden} = 64 \rightarrow 32$, learning rate $\eta = 1 \times 10^{-3}$, dropout $p = 0.20$, weight decay $\lambda = 1 \times 10^{-3}$);
2. A balanced standard network ($d_{hidden} = 128 \rightarrow 64$, $\eta = 1 \times 10^{-3}$, $p = 0.10$, $\lambda = 1 \times 10^{-4}$); and
3. A high-capacity, low-regularization network ($d_{hidden} = 128 \rightarrow 64$, $\eta = 2 \times 10^{-3}$, $p = 0.05$, $\lambda = 1 \times 10^{-5}$).

Model predictive performance was quantified using coefficient of determination (R^2), root mean squared error (RMSE), normalized RMSE as nRMSE (expressed as a percentage of the target range of mean), and mean absolute error (MAE).

2.8 Model Explainability by SHAP

To enhance model interpretability and reduce the 'black-box' nature of the ML models used, Shapley Additive Explanations (SHAP) analyses were conducted for the top-performing models [24]. Because the best-performing algorithms spanned different ML families including tree ensembles, decision trees, regularized linear models, kernel-based models, and deep

neural networks, model-specific SHAP explainer methods were used. TreeExplainer was applied to RF, XGBoost, and LightGBM [30], whereas LinearExplainer was applied to ElasticNet. Likewise, shap.KernelExplainer was used for SVR, while the MLP was evaluated using GradientExplainer, which uses expected gradients, an extension of integrated gradients [31]. The global feature importance was ranked using the mean absolute SHAP value (mean |SHAP|) and visualized via beeswarm plots concurrently showing the bidirectional effects.

3. Results

The original project contained a total of 1,576 experimental trials while the 1008 trials (for seven SLDL conditions in two different scenarios and 3 trials per each condition) were successfully recorded and processed across the 48 participants (7 x 3 x 48). All kinetic and kinematic data were verified to assure no abnormalities including missing values, suspected frozen values, abnormal data drifting, exaggerated or unreasonable values according to typical value ranges from previous literature. The final machine learning dataset comprised 2 scenarios, 9 target kinetic and postural control features, 17 sensor configurations (9 sensors; 7 sensors excluded both wrists; 4 sensors on pelvis and landing thigh, shank, and foot; 2 sensors for landing and non-landing feet, landing thigh and foot, landing thigh and shank, landing shank and foot, pelvis and landing shank, pelvis and landing foot, pelvis and landing wrist, pelvis and non-landing wrist; 1 sensor on pelvis/landing foot/landing thigh/landing shank/landing or non-landing wrists) and 3 sensor modalities (A only, G only, A+G sensors) across 6 ML algorithms (2 x 9 x 17 x 3 x 6) constructed a total of 5,508 ML models.

3.1 Machine Learning Predictive Accuracy Across Kinetic and Postural Control Metrics

Table 2. The best and mean model performance for predicting the kinetic variables across 6 machine learning methods in Scenario 1: single-leg drop landing condition.

Biomechanical Target	Predictive Metric	ElasticNet	SVR	Random Forest	LightGBM	XGBoost	MLP Neural Net
Phase 1: Impact Phase (0–250 ms)							
Peak vGRF (BW*)	**Best R^2**	0.798	0.709	0.739	0.762	0.752	**0.814**
	Best MAE	**0.223**	0.266	0.255	0.237	0.257	0.306
	Best RMSE	**0.281**	0.338	0.320	0.305	0.311	0.390
	Best nRMSE (%)	9.76	11.72	11.10	10.59	10.81	**6.61**
	Mean R^2 (±SD)	0.4696 ± 0.2565	0.4014 ± 0.2343	0.4476 ± 0.2186	0.4166 ± 0.2275	0.4159 ± 0.2429	0.5540 ± 0.1833
	Mean MAE (±SD)	0.3529 ± 0.0869	0.3781 ± 0.0763	0.3646 ± 0.0754	0.3736 ± 0.0764	0.3742 ± 0.0782	0.4552 ± 0.0963
	Mean RMSE (±SD)	0.4425 ± 0.1101	0.4751 ± 0.0941	0.4559 ± 0.0930	0.4683 ± 0.0962	0.4678 ± 0.1004	0.5917 ± 0.1241
	Mean nRMSE (%) (±SD)	15.3576 ± 3.8198	16.4884 ± 3.2652	15.8218 ± 3.2273	16.2555 ± 3.3407	16.2353 ± 3.4853	10.0304 ± 2.1037
Time to Peak vGRF (s*)	**Best R^2**	0.424	**0.628**	0.578	0.625	0.599	0.460
	Best MAE	0.007	**0.005**	0.006	**0.005**	0.006	0.006
	Best RMSE	0.009	**0.007**	0.008	**0.007**	0.007	0.011
	Best nRMSE (%)	13.38	10.75	11.46	10.80	11.17	**6.58**
	Mean R^2 (±SD)	0.2246 ± 0.1574	0.2794 ± 0.2135	0.3023 ± 0.2345	0.2867 ± 0.2461	0.2647 ± 0.2657	0.2413 ± 0.1568
	Mean MAE (±SD)	0.0080 ± 0.0009	0.0077 ± 0.0012	0.0076 ± 0.0014	0.0078 ± 0.0015	0.0078 ± 0.0014	0.0086 ± 0.0015
	Mean RMSE (±SD)	0.0104 ± 0.0010	0.0099 ± 0.0015	0.0097 ± 0.0016	0.0098 ± 0.0017	0.0100 ± 0.0018	0.0135 ± 0.0014
	Mean nRMSE (%) (±SD)	15.4482 ± 1.5514	14.8025 ± 2.2184	14.5294 ± 2.4183	14.6731 ± 2.5503	14.8804 ± 2.6917	7.7527 ± 0.7983
VALR (BW/s*)	**Best R^2**	**0.815**	0.713	0.757	0.681	0.800	0.699
	Best MAE	**8.791**	10.195	9.477	10.478	8.935	12.262
	Best RMSE	**11.118**	13.839	12.715	14.584	11.550	20.250
	Best nRMSE (%)	**7.98**	9.94	9.13	10.47	8.29	4.88
	Mean R^2 (±SD)	0.4653 ± 0.2657	0.3877 ± 0.2018	0.4203 ± 0.2710	0.3815 ± 0.2125	0.4118 ± 0.2879	0.4107 ± 0.2000
	Mean MAE (±SD)	13.4800 ± 3.1652	14.0934 ± 2.4012	13.8285 ± 3.1403	14.4813 ± 2.7311	13.8796 ± 3.2684	18.1923 ± 3.8401
	Mean RMSE (±SD)	18.2883 ± 4.7161	19.9254 ± 3.3518	19.1136 ± 4.6196	19.9967 ± 3.5355	19.1902 ± 4.9130	27.9540 ± 4.7710
	Mean nRMSE (%) (±SD)	13.1300 ± 3.3854	14.3043 ± 2.4061	13.7216 ± 3.3165	14.3561 ± 2.5377	13.7769 ± 3.5271	6.7347 ± 1.1499

VILR (BW/s*)	**Best R²**	**0.804**	0.768	0.763	0.678	0.755	0.776
	Best MAE	**29.651**	31.741	31.791	35.302	30.777	35.599
	Best RMSE	**37.697**	41.036	41.451	48.331	42.215	47.088
	Best nRMSE (%)	9.71	10.57	10.68	12.45	10.87	**7.94**
	Mean R² (±SD)	0.4232 ± 0.2936	0.3332 ± 0.2538	0.4107 ± 0.2693	0.3488 ± 0.2518	0.3796 ± 0.2872	0.4070 ± 0.2696
	Mean MAE (±SD)	48.4657 ± 12.9736	51.9176 ± 10.4652	49.1842 ± 12.0925	51.4897 ± 10.8343	49.7751 ± 12.6680	55.3975 ± 13.8354
	Mean RMSE (±SD)	62.4363 ± 17.2914	68.2779 ± 13.6692	63.6190 ± 15.4681	67.4702 ± 13.5442	65.2328 ± 16.0581	74.4107 ± 18.2169
	Mean nRMSE (%) (±SD)	16.0839 ± 4.4545	17.5871 ± 3.5200	16.3876 ± 3.9845	17.3794 ± 3.4892	16.8037 ± 4.1369	12.5455 ± 3.0714
Phase 2: Dynamic Postural Stability Phase (0–3.0 s)							
DPSI AP (Fy*)	**Best R²**	0.142	0.094	0.108	0.129	0.030	**0.509**
	Best MAE	**0.006**	**0.006**	**0.006**	**0.006**	**0.006**	0.014
	Best RMSE	**0.007**	**0.007**	**0.007**	**0.007**	0.008	0.019
	Best nRMSE (%)	18.18	18.68	18.53	18.32	19.33	**14.02**
	Mean R² (±SD)	-0.0057 ± 0.0354	-0.1582 ± 0.1040	-0.1271 ± 0.0824	-0.1620 ± 0.1074	-0.2295 ± 0.1020	0.2171 ± 0.1100
	Mean MAE (±SD)	0.0062 ± 0.0001	0.0067 ± 0.0004	0.0066 ± 0.0003	0.0067 ± 0.0003	0.0069 ± 0.0004	0.0184 ± 0.0017
	Mean RMSE (±SD)	0.0078 ± 0.0002	0.0084 ± 0.0004	0.0083 ± 0.0003	0.0084 ± 0.0004	0.0086 ± 0.0004	0.0239 ± 0.0018
	Mean nRMSE (%) (±SD)	19.6786 ± 0.3582	21.0988 ± 0.9594	20.8210 ± 0.7767	21.1331 ± 1.0019	21.7433 ± 0.9064	17.6616 ± 1.2913
DPSI ML (Fx*)	**Best R²**	0.029	0.206	0.207	0.128	0.155	**0.478**
	Best MAE	**0.004**	**0.004**	**0.004**	**0.004**	**0.004**	0.017
	Best RMSE	0.005	**0.004**	**0.004**	0.005	0.005	0.025
	Best nRMSE (%)	19.44	17.59	**17.57**	18.42	18.14	19.58
	Mean R² (±SD)	0.0019 ± 0.0119	0.0294 ± 0.0846	0.0455 ± 0.0898	-0.0108 ± 0.0889	-0.0476 ± 0.1151	0.1509 ± 0.1000
	Mean MAE (±SD)	0.0040 ± 0.0000	0.0039 ± 0.0002	0.0039 ± 0.0002	0.0040 ± 0.0002	0.0041 ± 0.0002	0.0249 ± 0.0027
	Mean RMSE (±SD)	0.0050 ± 0.0000	0.0049 ± 0.0002	0.0049 ± 0.0002	0.0050 ± 0.0002	0.0051 ± 0.0003	0.0315 ± 0.0019
	Mean nRMSE (%) (±SD)	19.7129 ± 0.1178	19.4227 ± 0.8395	19.2582 ± 0.9001	19.8208 ± 0.8584	20.1673 ± 1.1017	24.9192 ± 1.5176
DPSI Total Index	**Best R²**	0.477	0.487	0.460	0.472	0.480	**0.641**
	Best MAE	0.021	0.020	0.019	0.021	**0.018**	0.026

	Best RMSE	0.029	**0.028**	0.029	0.029	0.029	0.049
	Best nRMSE (%)	11.07	10.97	11.26	11.12	11.04	**4.17**
	Mean R^2 (±SD)	0.2388 ± 0.1602	0.2557 ± 0.1530	0.2000 ± 0.1654	0.1641 ± 0.1422	0.1733 ± 0.1952	0.4091 ± 0.1383
	Mean MAE (±SD)	0.0245 ± 0.0034	0.0244 ± 0.0030	0.0249 ± 0.0034	0.0261 ± 0.0027	0.0255 ± 0.0037	0.0381 ± 0.0057
	Mean RMSE (±SD)	0.0343 ± 0.0036	0.0339 ± 0.0035	0.0352 ± 0.0036	0.0360 ± 0.0031	0.0357 ± 0.0042	0.0624 ± 0.0072
	Mean nRMSE (%) (±SD)	13.2878 ± 1.3899	13.1425 ± 1.3524	13.6247 ± 1.3989	13.9492 ± 1.1980	13.8267 ± 1.6398	5.3141 ± 0.6119
Phase 3: Postural Stabilization & Baseline Phase (3.0–15.0 s)							
TTS Ross Range Fx (s*)	**Best R^2**	**0.386**	0.333	0.256	0.246	0.254	0.315
	Best MAE	**0.477**	0.492	0.524	0.527	0.521	0.643
	Best RMSE	**0.596**	0.621	0.656	0.660	0.657	1.086
	Best nRMSE (%)	14.72	15.34	16.20	16.31	16.22	**8.01**
	Mean R^2 (±SD)	0.0932 ± 0.1350	0.0400 ± 0.1267	0.0413 ± 0.1075	0.0302 ± 0.1024	-0.0483 ± 0.1177	0.1594 ± 0.0658
	Mean MAE (±SD)	0.5714 ± 0.0453	0.5884 ± 0.0429	0.5899 ± 0.0321	0.5984 ± 0.0320	0.6163 ± 0.0367	0.7092 ± 0.0293
	Mean RMSE (±SD)	0.7225 ± 0.0544	0.7438 ± 0.0502	0.7438 ± 0.0422	0.7482 ± 0.0401	0.7777 ± 0.0448	1.2022 ± 0.0477
	Mean nRMSE (%) (±SD)	17.8382 ± 1.3442	18.3655 ± 1.2401	18.3649 ± 1.0423	18.4745 ± 0.9899	19.2033 ± 1.1053	8.8694 ± 0.3521
TTS Ross Range Fy (s*)	**Best R^2**	0.102	0.115	0.118	0.091	0.076	**0.147**
	Best MAE	0.163	**0.162**	0.166	0.170	0.167	0.247
	Best RMSE	0.225	**0.223**	**0.223**	0.226	0.228	0.452
	Best nRMSE (%)	13.72	13.62	13.59	13.80	13.91	**4.05**
	Mean R^2 (±SD)	-0.0095 ± 0.0491	-0.0339 ± 0.0723	-0.0721 ± 0.0998	-0.0657 ± 0.0852	-0.1120 ± 0.0920	0.0738 ± 0.0393
	Mean MAE (±SD)	0.1713 ± 0.0039	0.1751 ± 0.0076	0.1801 ± 0.0073	0.1806 ± 0.0085	0.1841 ± 0.0082	0.2593 ± 0.0086
	Mean RMSE (±SD)	0.2384 ± 0.0058	0.2412 ± 0.0084	0.2455 ± 0.0114	0.2448 ± 0.0098	0.2501 ± 0.0103	0.4713 ± 0.0101
	Mean nRMSE (%) (±SD)	14.5376 ± 0.3546	14.7075 ± 0.5142	14.9702 ± 0.6962	14.9296 ± 0.5983	15.2506 ± 0.6299	4.2196 ± 0.0896

*Note: vGRF, vertical ground reaction force; BW, body weight; VALR, vertical average loading rate; VILR, vertical instantaneous loading rate; DPSI, dynamic postural stability index; AP, anterior–posterior axis (Fy); ML, medial–lateral axis (Fx); TTS, time-to-stabilization; R^2, coefficient of determination; MAE, mean absolute error; RMSE, root mean square error; nRMSE, normalized root mean square error (range-normalized, %); SVR, support vector regression; MLP, multilayer perceptron neural network. 'Best' denotes the peak predictive capacity achieved by the optimal wearable sensor configuration and hyperparameter set. 'Mean (±SD)' indicates the architecture's omnibus stability across all tested sensor subsets and signal modalities. Bold text denotes the champion algorithm yielding the highest peak R^2 for each respective target; *the units are applied to MAE and RMSE*

The predictive performance of the six selected ML models across the 9 targets variables is summarized in Table 2. Among the 9 target variables in the baseline SLDL condition (scenario 1), the ElasticNet yielded the strongest predictive accuracy for VALR variable (R^2 = 0.815; MAE = 8.791 BW/s; RMSE = 11.118; nRMSE = 7.98%) whereas, MLP achieved the highest predictive accuracy (in terms of coefficient of determination, R^2) across more than half of the evaluated landing metrics including the normalized Peak vGRF (R^2 = 0.814; MAE = 0.306 BW; RMSE = 0.390; nRMSE = 6.61%), DPSI AP (R^2 = 0.509; MAE = 0.014; RMSE = 0.019; nRMSE = 14.02%), DPSI ML (R^2 = 0.478; MAE = 0.017; RMSE = 0.025; nRMSE = 19.58%), DPSI Total (R^2 = 0.641; MAE = 0.026; RMSE = 0.049; nRMSE = 4.17%), and TTS Ross in AP (Fy) direction (R^2 = 0.147; MAE = 0.247 s; RMSE = 0.452; nRMSE = 4.05%). Regarding the metrics with strong predictive accuracy ($R^2 \geq 0.7$), only those under kinetic domains including normalized Peak vGRF, VALR and VILR achieved peak R^2 of 0.814, 0.815 and 0.804 respectively (Table 2).

Table 3. Out-of-fold predictive performance across six candidate machine learning architectures evaluated across nine biomechanical target variables during multi-planar single-leg drop landings (Scenario 2: MSLDL).

Biomechanical Target	Predictive Metric	ElasticNet	SVR	Random Forest	LightGBM	XGBoost	MLP Neural Net
Phase 1: Impact Phase (0–250 ms)							
Peak vGRF (BW*)	**Best R^2**	0.794	0.742	0.752	0.773	0.772	**0.808**
	Best MAE	0.270	0.292	0.291	0.279	0.283	0.302
	Best RMSE	0.348	0.390	0.382	0.366	0.367	0.397
	Best nRMSE (%)	7.59	8.50	8.33	7.98	7.99	6.72
	Mean R^2 (±SD)	0.3613 ± 0.3818	0.3819 ± 0.2432	0.3848 ± 0.2633	0.3654 ± 0.2814	0.3768 ± 0.2629	0.5534 ± 0.1859
	Mean MAE (±SD)	0.4478 ± 0.1106	0.4548 ± 0.0962	0.4561 ± 0.1034	0.4643 ± 0.1103	0.4596 ± 0.1031	0.4546 ± 0.0981
	Mean RMSE (±SD)	0.5898 ± 0.1715	0.5914 ± 0.1225	0.5877 ± 0.1331	0.5956 ± 0.1410	0.5916 ± 0.1336	0.5918 ± 0.1259
	Mean nRMSE (%) (±SD)	12.8506 ± 3.7356	12.8857 ± 2.6676	12.8049 ± 2.8987	12.9775 ± 3.0721	12.8904 ± 2.9099	10.0316 ± 2.1331
Time to Peak vGRF (s*)	**Best R^2**	0.376	0.550	**0.583**	0.574	0.547	0.432
	Best MAE	0.007	0.005	0.006	0.005	0.006	0.007
	Best RMSE	0.011	0.009	0.009	0.009	0.009	0.012
	Best nRMSE (%)	7.55	6.41	6.17	6.24	6.43	6.74
	Mean R^2 (±SD)	0.1798 ± 0.1338	0.2860 ± 0.2074	0.2628 ± 0.2198	0.2809 ± 0.2315	0.2743 ± 0.2048	0.2395 ± 0.1584
	Mean MAE (±SD)	0.0086 ± 0.0009	0.0075 ± 0.0016	0.0078 ± 0.0016	0.0078 ± 0.0018	0.0078 ± 0.0016	0.0086 ± 0.0015
	Mean RMSE (±SD)	0.0127 ± 0.0010	0.0118 ± 0.0017	0.0119 ± 0.0018	0.0118 ± 0.0019	0.0118 ± 0.0017	0.0135 ± 0.0014
	Mean nRMSE (%) (±SD)	8.6259 ± 0.7145	7.9947 ± 1.1525	8.1161 ± 1.2153	8.0029 ± 1.2867	8.0627 ± 1.1346	7.7618 ± 0.8025
VALR (BW/s*)	**Best R^2**	**0.744**	0.645	0.697	0.712	0.683	0.680
	Best MAE	12.204	13.071	13.008	12.760	13.349	13.111
	Best RMSE	17.981	21.153	19.538	19.070	19.996	20.878
	Best nRMSE (%)	6.44	7.58	7.00	6.83	7.16	5.03
	Mean R^2 (±SD)	0.3546 ± 0.2984	0.3611 ± 0.2223	0.3490 ± 0.2514	0.3436 ± 0.2614	0.3344 ± 0.2380	0.4113 ± 0.2018
	Mean MAE (±SD)	19.5935 ± 4.5668	19.1242 ± 4.3561	19.6652 ± 4.4871	19.8503 ± 4.7004	19.8503 ± 4.3086	18.1776 ± 3.9130
	Mean RMSE (±SD)	27.7992 ± 6.4694	27.9573 ± 4.9444	28.1034 ± 5.6277	28.1856 ± 5.8211	28.5020 ± 5.2379	27.9315 ± 4.8079

	Mean nRMSE (%) (±SD)	9.9598 ± 2.3181	10.0157 ± 1.7711	10.0692 ± 2.0165	10.0978 ± 2.0854	10.2110 ± 1.8771	6.7288 ± 1.1580
VILR (BW/s*)	**Best R²**	0.780	0.727	0.785	**0.788**	0.784	0.776
	Best MAE	35.415	38.499	35.258	34.779	35.322	34.993
	Best RMSE	47.072	52.434	46.534	46.180	46.621	47.057
	Best nRMSE (%)	8.19	9.12	8.09	8.03	8.11	7.93
	Mean R² (±SD)	0.3671 ± 0.3689	0.3798 ± 0.2630	0.4060 ± 0.2902	0.3859 ± 0.3044	0.3948 ± 0.2933	0.4101 ± 0.2675
	Mean MAE (±SD)	58.0670 ± 15.1222	57.9917 ± 13.6464	57.3410 ± 15.4933	58.2550 ± 15.9660	57.7172 ± 15.3674	55.1755 ± 13.6837
	Mean RMSE (±SD)	76.6234 ± 22.4752	77.1104 ± 17.2911	74.7449 ± 19.9071	75.9080 ± 20.5770	75.4795 ± 19.9766	74.2301 ± 18.1238
	Mean nRMSE (%) (±SD)	13.3275 ± 3.9093	13.4120 ± 3.0068	13.0000 ± 3.4626	13.2025 ± 3.5785	13.1284 ± 3.4741	12.5139 ± 3.0551
Phase 2: Dynamic Postural Stability Phase (0–3.0 s)							
DPSI AP (Fy*)	**Best R²**	0.364	0.455	0.420	0.450	0.457	**0.513**
	Best MAE	0.021	0.019	0.020	0.018	0.019	0.013
	Best RMSE	0.025	0.023	0.024	0.023	0.023	0.019
	Best nRMSE (%)	20.01	18.53	19.11	18.62	18.49	13.97
	Mean R² (±SD)	0.1246 ± 0.1009	0.1316 ± 0.1705	0.1562 ± 0.1311	0.1305 ± 0.1567	0.1600 ± 0.1371	0.2157 ± 0.1107
	Mean MAE (±SD)	0.0254 ± 0.0018	0.0235 ± 0.0024	0.0241 ± 0.0024	0.0240 ± 0.0028	0.0240 ± 0.0026	0.0184 ± 0.0017
	Mean RMSE (±SD)	0.0295 ± 0.0017	0.0293 ± 0.0029	0.0289 ± 0.0023	0.0293 ± 0.0027	0.0288 ± 0.0024	0.0239 ± 0.0018
	Mean nRMSE (%) (±SD)	23.4406 ± 1.3892	23.2706 ± 2.3462	22.9829 ± 1.8235	23.3037 ± 2.1594	22.9216 ± 1.9231	17.6771 ± 1.2991
DPSI ML (Fx*)	**Best R²**	0.194	0.259	0.246	0.271	0.250	**0.471**
	Best MAE	0.026	0.024	0.025	0.024	0.024	0.017
	Best RMSE	0.030	0.029	0.029	0.029	0.029	0.025
	Best nRMSE (%)	24.64	23.62	23.83	23.43	23.77	19.71
	Mean R² (±SD)	0.0557 ± 0.0500	0.0198 ± 0.1155	0.0805 ± 0.0823	0.0346 ± 0.1052	0.0729 ± 0.0834	0.1511 ± 0.0980
	Mean MAE (±SD)	0.0289 ± 0.0011	0.0274 ± 0.0019	0.0277 ± 0.0016	0.0278 ± 0.0019	0.0278 ± 0.0017	0.0249 ± 0.0026
	Mean RMSE (±SD)	0.0327 ± 0.0009	0.0333 ± 0.0020	0.0323 ± 0.0015	0.0331 ± 0.0018	0.0324 ± 0.0015	0.0315 ± 0.0019
	Mean nRMSE (%) (±SD)	26.6602 ± 0.7187	27.1247 ± 1.6249	26.2910 ± 1.1900	26.9255 ± 1.4969	26.3990 ± 1.2032	24.9202 ± 1.4901
DPSI Total Index	**Best R²**	0.555	0.575	0.505	0.551	0.569	**0.654**

	Best MAE	0.024	0.022	0.025	0.024	0.024	0.025
	Best RMSE	0.031	0.030	0.033	0.031	0.031	0.048
	Best nRMSE (%)	9.38	9.17	9.90	9.42	9.23	4.09
	Mean R^2 (±SD)	0.1889 ± 0.3365	0.2526 ± 0.1563	0.2222 ± 0.1510	0.2053 ± 0.1726	0.2387 ± 0.1569	0.4099 ± 0.1390
	Mean MAE (±SD)	0.0307 ± 0.0036	0.0306 ± 0.0032	0.0313 ± 0.0029	0.0317 ± 0.0034	0.0310 ± 0.0031	0.0380 ± 0.0058
	Mean RMSE (±SD)	0.0415 ± 0.0076	0.0403 ± 0.0043	0.0411 ± 0.0040	0.0415 ± 0.0046	0.0406 ± 0.0043	0.0623 ± 0.0072
	Mean nRMSE (%) (±SD)	12.4625 ± 2.2893	12.0906 ± 1.2987	12.3467 ± 1.2148	12.4629 ± 1.3851	12.2049 ± 1.2925	5.3098 ± 0.6148
Phase 3: Postural Stabilization & Baseline Phase (3.0–15.0 s)							
TTS Ross Range Fx (s*)	**Best R^2**	0.162	0.121	0.174	0.197	0.178	**0.324**
	Best MAE	0.428	0.387	0.417	0.404	0.408	0.628
	Best RMSE	0.625	0.640	0.621	0.612	0.619	1.079
	Best nRMSE (%)	6.95	7.12	6.90	6.81	6.88	7.96
	Mean R^2 (±SD)	0.0275 ± 0.0498	0.0194 ± 0.0438	0.0138 ± 0.0665	-0.0131 ± 0.0858	0.0137 ± 0.0681	0.1556 ± 0.0674
	Mean MAE (±SD)	0.4531 ± 0.0086	0.4127 ± 0.0107	0.4566 ± 0.0169	0.4630 ± 0.0223	0.4515 ± 0.0177	0.7101 ± 0.0320
	Mean RMSE (±SD)	0.6734 ± 0.0174	0.6763 ± 0.0152	0.6780 ± 0.0232	0.6869 ± 0.0296	0.6780 ± 0.0237	1.2049 ± 0.0488
	Mean nRMSE (%) (±SD)	7.4867 ± 0.1934	7.5184 ± 0.1696	7.5373 ± 0.2580	7.6371 ± 0.3292	7.5375 ± 0.2631	8.8886 ± 0.3601
TTS Ross Range Fy (s*)	**Best R^2**	0.301	0.247	0.321	0.304	**0.333**	0.164
	Best MAE	0.233	0.227	0.226	0.228	0.223	0.243
	Best RMSE	0.333	0.346	0.329	0.333	0.326	0.448
	Best nRMSE (%)	5.78	6.01	5.70	5.77	5.65	4.01
	Mean R^2 (±SD)	0.1243 ± 0.0848	0.0847 ± 0.0856	0.1109 ± 0.0987	0.0756 ± 0.1091	0.1219 ± 0.0965	0.0743 ± 0.0407
	Mean MAE (±SD)	0.2608 ± 0.0110	0.2508 ± 0.0112	0.2631 ± 0.0158	0.2680 ± 0.0187	0.2595 ± 0.0154	0.2596 ± 0.0085
	Mean RMSE (±SD)	0.3726 ± 0.0183	0.3810 ± 0.0181	0.3753 ± 0.0211	0.3826 ± 0.0230	0.3730 ± 0.0208	0.4711 ± 0.0104
	Mean nRMSE (%) (±SD)	6.4676 ± 0.3171	6.6127 ± 0.3131	6.5145 ± 0.3665	6.6406 ± 0.3991	6.4743 ± 0.3608	4.2184 ± 0.0938

*Note: MSLDL, multi-planar single-leg drop landing; vGRF, vertical ground reaction force; BW, body weight; VALR, vertical average loading rate; VILR, vertical instantaneous loading rate; DPSI, dynamic postural stability index; AP, anterior–posterior axis (Fy); ML, medial–lateral axis (Fx); TTS, time-to-stabilization; R^2, coefficient of determination; MAE, mean absolute error; RMSE, root mean square error; nRMSE, normalized root mean square error (range-normalized, %); SVR, support vector regression; MLP, multilayer perceptron neural network. 'Best' denotes the peak predictive capacity achieved by the optimal wearable sensor configuration and hyperparameter set. 'Mean (±SD)' indicates the architecture's omnibus stability across all tested sensor subsets and signal modalities. Bold text denotes the champion algorithm yielding the highest peak R^2 for each respective target; *the units are applied to MAE and RMS*

For the MSLDL conditions (scenario 2), the MLP outperformed other ML models in more than half of the target variables including the normalized Peak vGRF (R^2 = 0.808; MAE = 0.302 BW; RMSE = 0.397; nRMSE = 6.72%), DPSI AP (R^2 = 0.513; MAE = 0.013; RMSE = 0.019; nRMSE = 13.97%), DPSI ML (R^2 = 0.471; MAE = 0.017; RMSE = 0.025; nRMSE = 19.71%), DPSI Total (R^2 = 0.654; MAE = 0.025; RMSE = 0.048; nRMSE = 4.09%), and TTS Ross in ML (Fx) direction (R^2 = 0.324; MAE = 0.628 s; RMSE = 1.079; nRMSE = 7.96%). RF yielded the best model in Time to Peak with R^2 of 0.583 only. For the VALR and VILR, the top-performing ML models are ElasticNet (R^2 = 0.744) and LightGBM (R^2 = 0.788) respectively. Similar to the SLDL condition, MSLDL showed only strong predictive accuracy ($R^2 \geq 0.7$) in kinetic domains variables including Peak vGRF, VALR and VILR (Table 3).

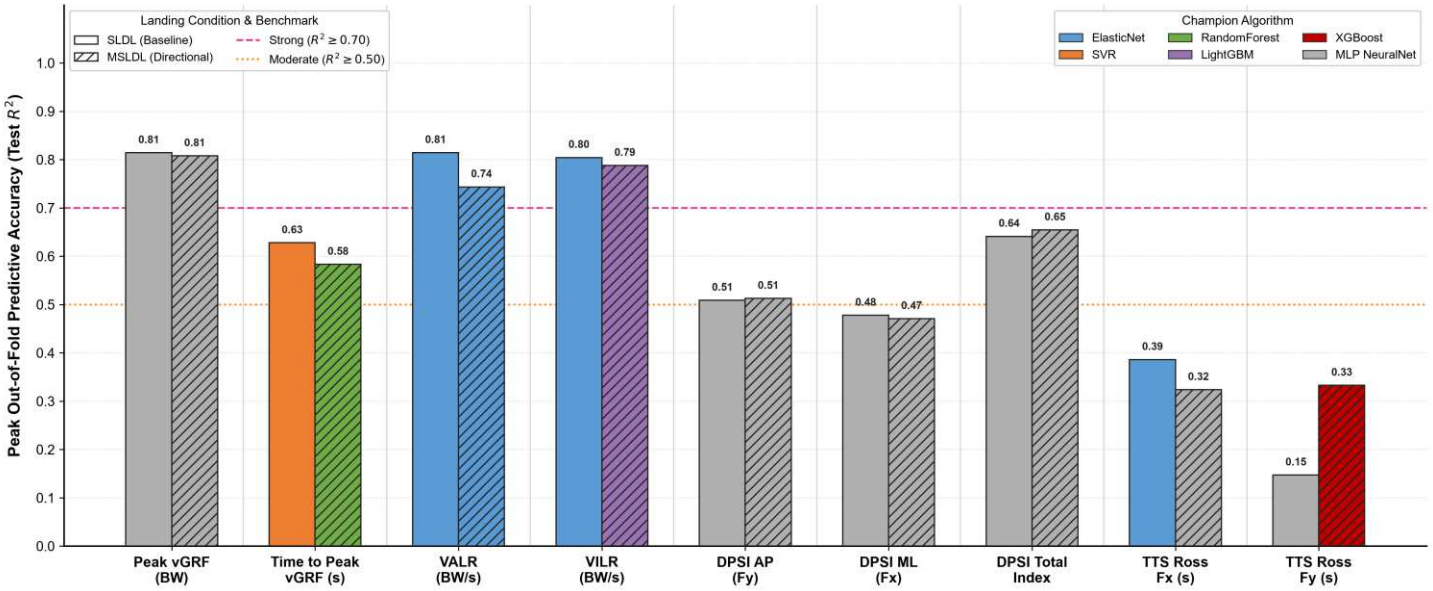
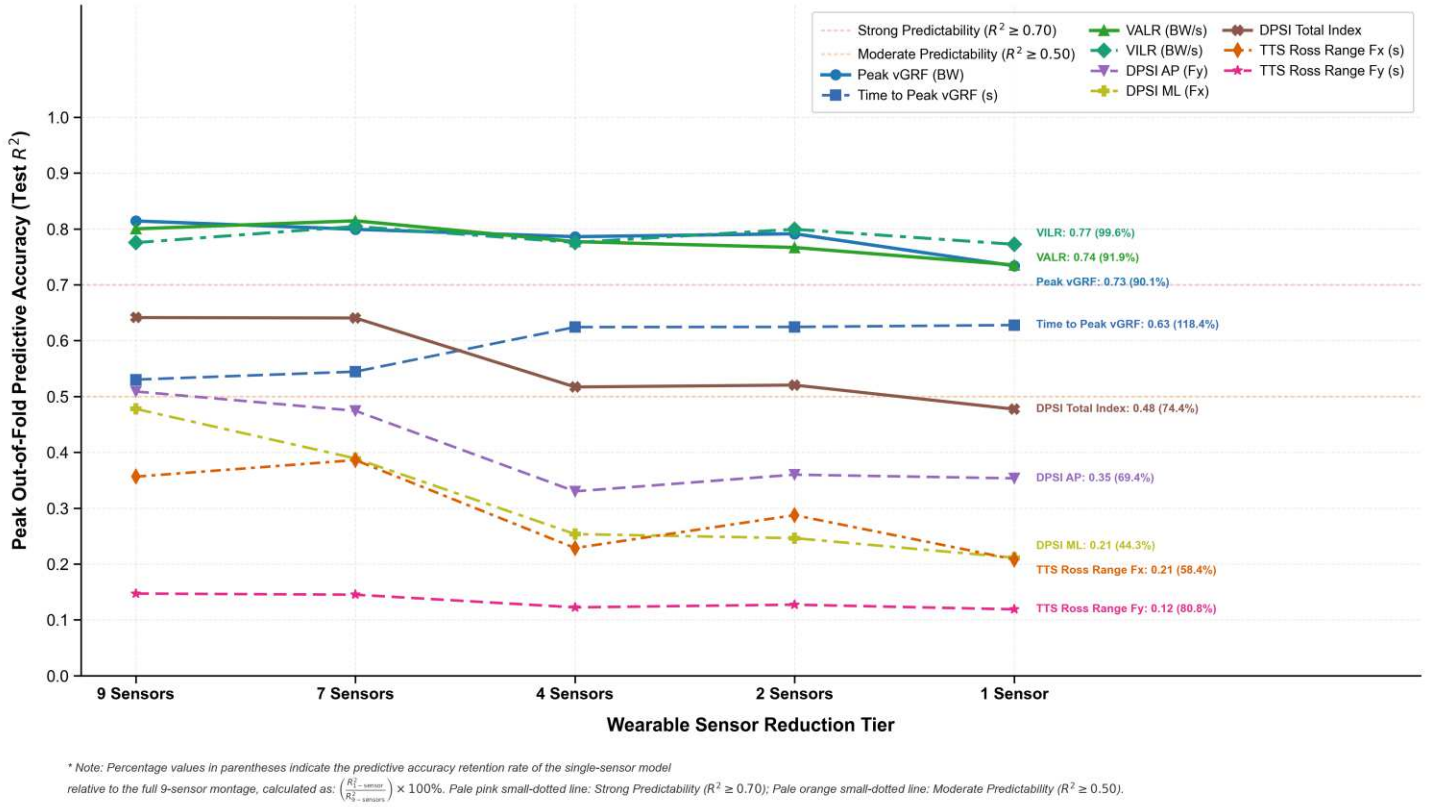


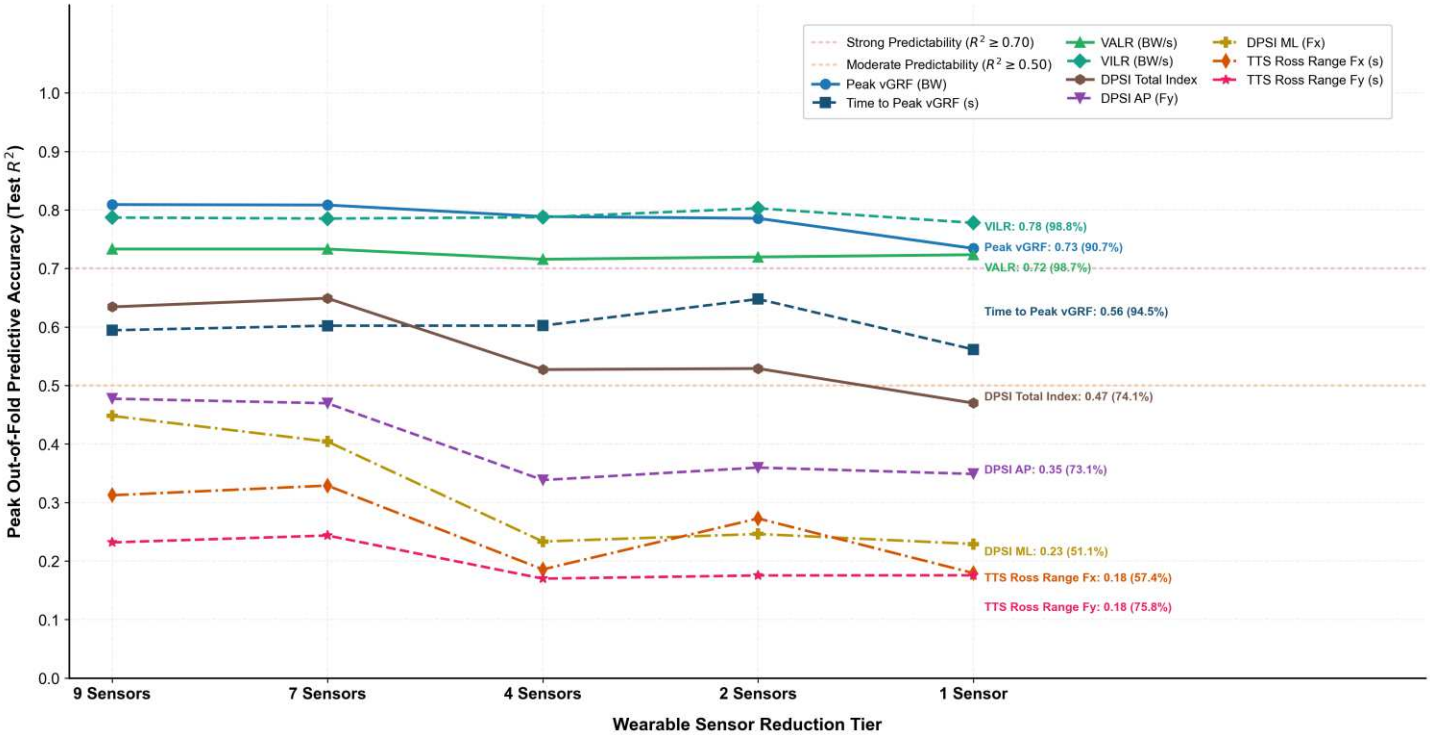
Figure 1 – The comparison of peak predictive accuracy between scenario 1 (single-leg drop landing) and 2 (multi-planar single-leg drop landing)

When comparing the predictive accuracy between scenario 1 (SLDL) and 2 (MSLDL), most performance are similar and comparable (Fig. 1). Regarding the kinetic domain (Peak vGRF, Time to Peak, VALR and VILR), the R^2 values of the best-performing models in SLDL are slightly higher than those of MSLDL. In contrast, the prediction of several postural control and stability-related metrics including DPSI in AP direction (Fy), DPSI Total, and TTS in AP direction (Fy) are more accurate in MSLDL than that of SLDL condition.

3.2 Quantities of Sensors and Predictive Accuracy Trade Off



** Note: Percentage values in parentheses indicate the predictive accuracy retention rate of the single-sensor model relative to the full 9-sensor montage, calculated as: $\left(\frac{R^2_{1-\text{sensor}}}{R^2_{9-\text{sensors}}}\right) \times 100\%$. Pale pink small-dotted line: Strong Predictability ($R^2 \geq 0.70$); Pale orange small-dotted line: Moderate Predictability ($R^2 \geq 0.50$).*



** Note: Percentage values in parentheses indicate the predictive accuracy retention rate of the single-sensor model relative to the full 9-sensor montage, calculated as: $(R^2_{1-\text{sensor}}/R^2_{9-\text{sensors}}) \times 100\%$. Pale pink small-dotted line: Strong Predictability ($R^2 \geq 0.70$); Pale orange small-dotted line: Moderate Predictability ($R^2 \geq 0.50$).*

Figure 2 – The peak predictive accuracy and retention percentage of single IMU sensor (when compared to 9 sensors) in top: scenario 1, single-leg drop landing condition and bottom: scenario 2, multi-planar drop landing condition.

Regarding the kinetic domain variables in both SLDL and MSLDL condition including Peak vGRF, Time to Peak vGRF, VALR, and VILR, single IMU sensor achieved overall higher than 90% accuracy retention when compared to 9 sensors used (SLDL: 90.1%, 118.4%, 91.9%, and 99.6% respectively; MSLDL: 90.7%, 94.5%, 98.7%, and 98.8% respectively) whereas, except the Time to Peak vGRF, all yielded R^2 value higher than 0.7 showing strong predictive accuracy (Peak vGRF: $R^2 = 0.73$; VALR: $R^2 = 0.74$; VILR: $R^2 = 0.77$ for SLDL; Peak vGRF: $R^2 = 0.73$; VALR: $R^2 = 0.72$; VILR: $R^2 = 0.78$ for MSLDL). For other postural control and stability-related metrics, the predictive accuracy retentions (1 vs. 9 sensors) are between 44.3 to 80.8% (Figure 2).

3.3 Sensor Placement Modality Comparison

A direct comparison between anatomical sensor locations (distal shank versus midfoot dorsum) revealed pronounced domain-specific disparities (Table 2). When using the flagship LightGBM model:

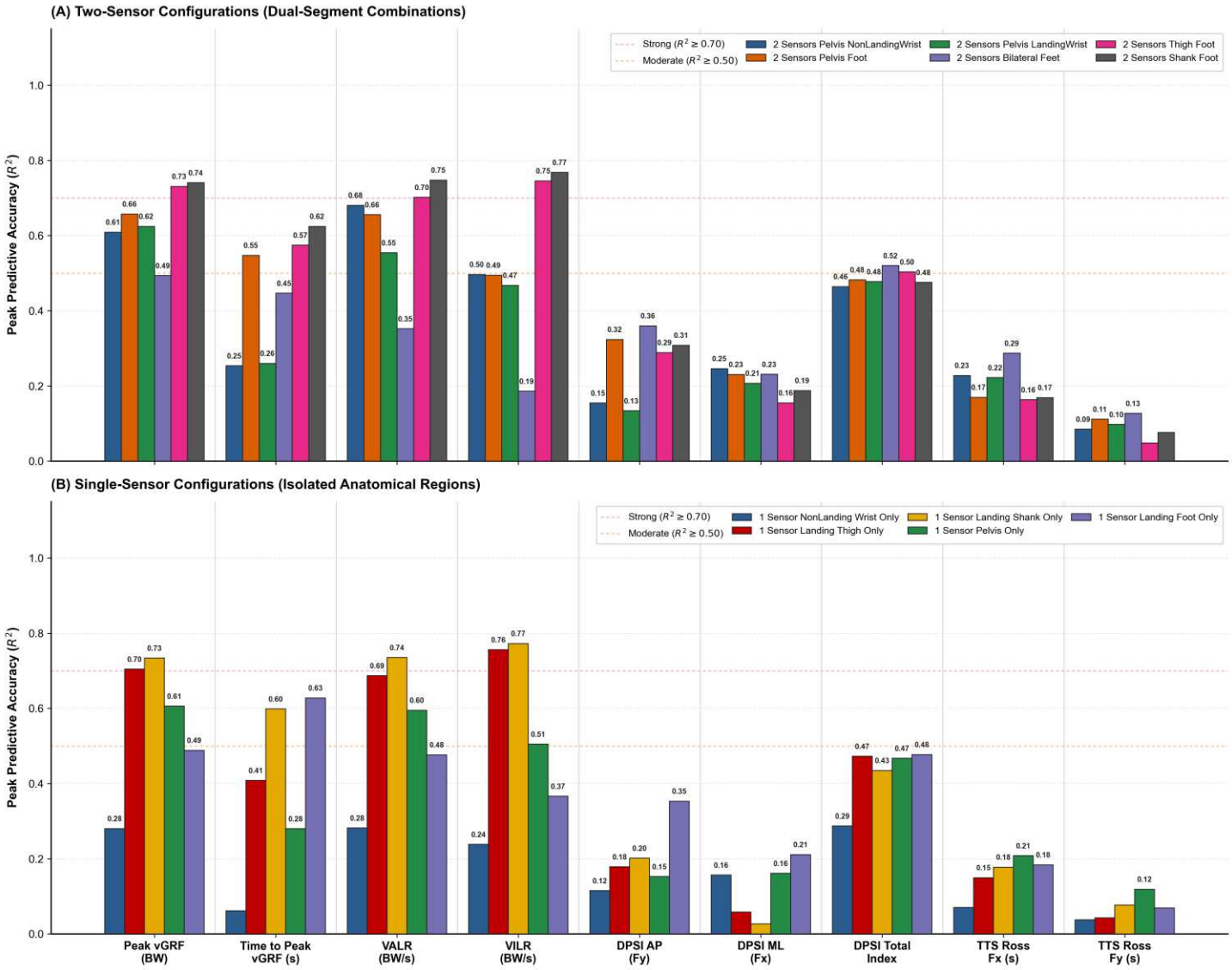


Figure 3 – The peak predictive accuracy of different sensor placement configurations in scenario 1, single-leg drop landing condition; Top: results about 2-sensor configuration; Bottom: results about 1-sensor only

In the 2-sensor configuration (Figure 3), the dominant shank-foot combo yielded the best predictive accuracy across kinetic-related metrics including Peak vGRF ($R^2 = 0.74$), Time to Peak vGRF ($R^2 = 0.62$), VALR ($R^2 = 0.75$), and VILR

($R^2 = 0.77$). For postural control and stability-related variables, all except DPSI in ML direction (Fx) were best performed with bilateral feet placement whereas, only the DPSI Total produced a moderate predictive accuracy ($R^2 = 0.52$).

For the single-sensor configuration, the dominant-shank produced the best predictive accuracy in Peak vGRF ($R^2 = 0.73$), VARL ($R^2 = 0.74$), and VILR ($R^2 = 0.77$). The landing-foot placement yielded the top-performing predictive accuracy in Time to Peak ($R^2 = 0.63$), DPSI in AP ($R^2 = 0.35$), DPSI in ML ($R^2 = 0.21$), and DPSI Total ($R^2 = 0.48$). Regarding the TTS-related metrics, although pelvis placement performed the best in both ML (Fx) and AP (Fy) directions, the R^2 values are only 0.21 and 0.12 respectively.

The MSLDL condition shows a similar results and trends as the SLDL condition except that the landing thigh-foot placement for 2-sensor configuration produced the best predictive accuracy in VALR ($R^2 = 0.72$) and VILR ($R^2 = 0.78$). Likewise for the single-sensor configuration, the landing thigh placement also yielded the best prediction in both VALR ($R^2 = 0.72$) and VILR ($R^2 = 0.78$).

For the sensor modalities, either accelerometer (A) only or combined (A+G) sensors yielded the best predictive accuracy across the 9 selected output variables. A sensor only achieved the best-predictive performance in VALR ($R^2 = 0.81$), DPSI Total ($R^2 = 0.64$), and TTS Fx ($R^2 = 0.39$ whereas, combined (A+G) produced the top-performing accuracy in Peak vGRF ($R^2 = 0.81$), Time to Peak ($R^2 = 0.63$), VILR ($R^2 = 0.80$), DPSI AP ($R^2 = 0.51$), DPSI ML ($R^2 = 0.48$), DPSI Total ($R^2 = 0.64$), and TTS Fy ($R^2 = 0.15$).

3.3 SHAP Analysis

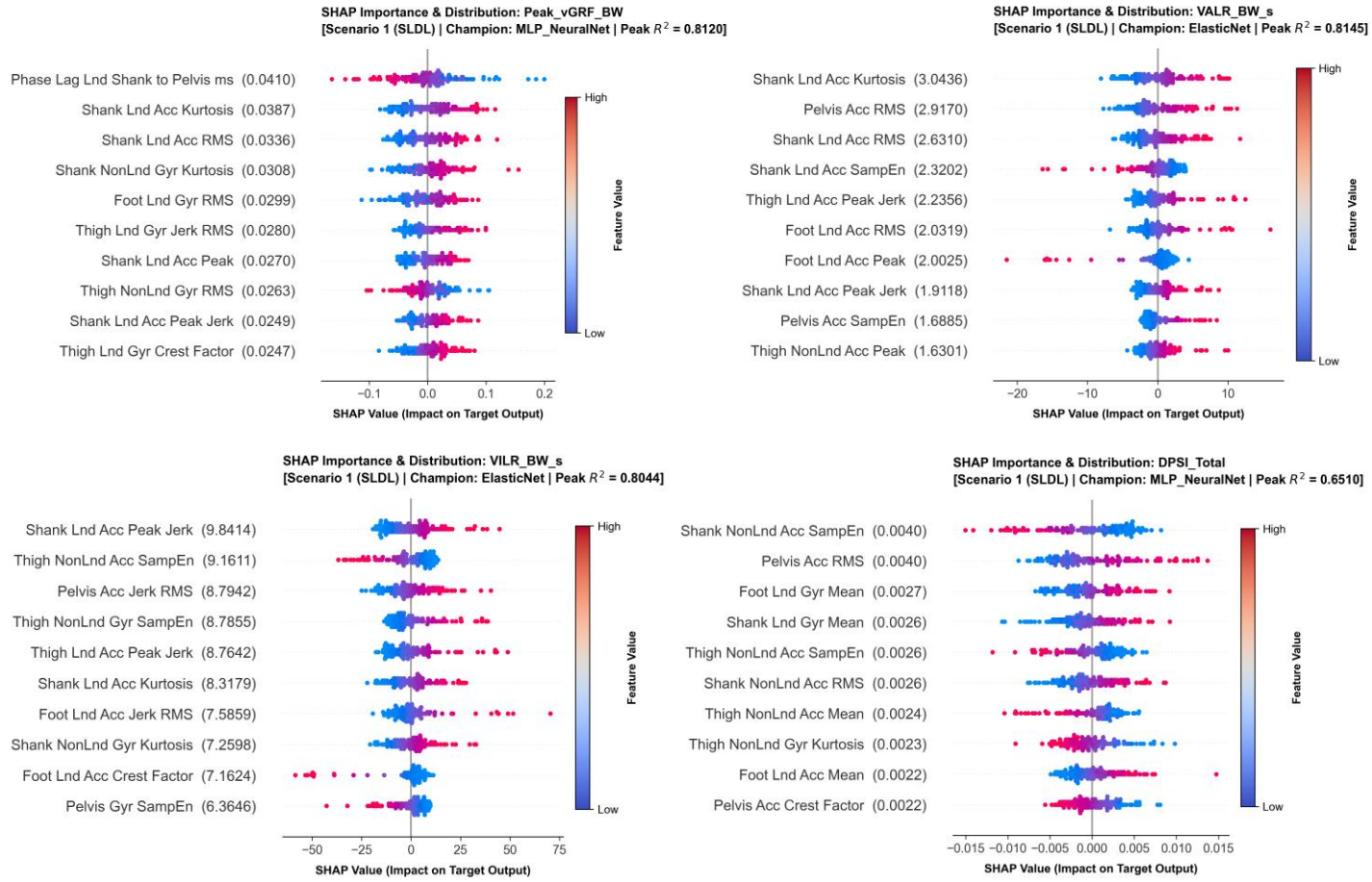


Figure 4 – SHAP analyses for Peak vGRF, VALR, VILR, and DPSI Total for scenario 1; SHAP values are shown inside the bracket

Figure 4 illustrates the SHAP visualization and value using beeswarm plot for Peak vGRF, VALR, VILR, and DPSI Total in SLDL condition. For Peak vGRF, the top three most important features are the phase lag of landing shank to pelvis, landing shank accelerometer kurtosis, and landing shank accelerometer RMS with SHAP values of 0.041, 0.0387, and 0.0336 respectively. Regarding VALR, the top 3 most impactful features are landing shank accelerometer kurtosis, pelvis accelerometer RMS, and landing shank accelerometer RMS with SHAP values of 3.0436, 2.9170, and 2.6310 respectively. For VILR, the top three important features are landing shank accelerometer peak jerk (SHAP = 9.8414), non-landing thigh accelerometer sample entropy (SHAP = 9.1611), and pelvis accelerometer jerk RMS (SHAP = 8.7942). Likewise, the most impactful features for predicting DPSI Total are non-landing shank accelerometer sample entropy (SHAP = 0.0040), pelvis accelerometer RMS (SHAP = 0.0040), and landing foot gyroscope mean (SHAP = 0.0027).

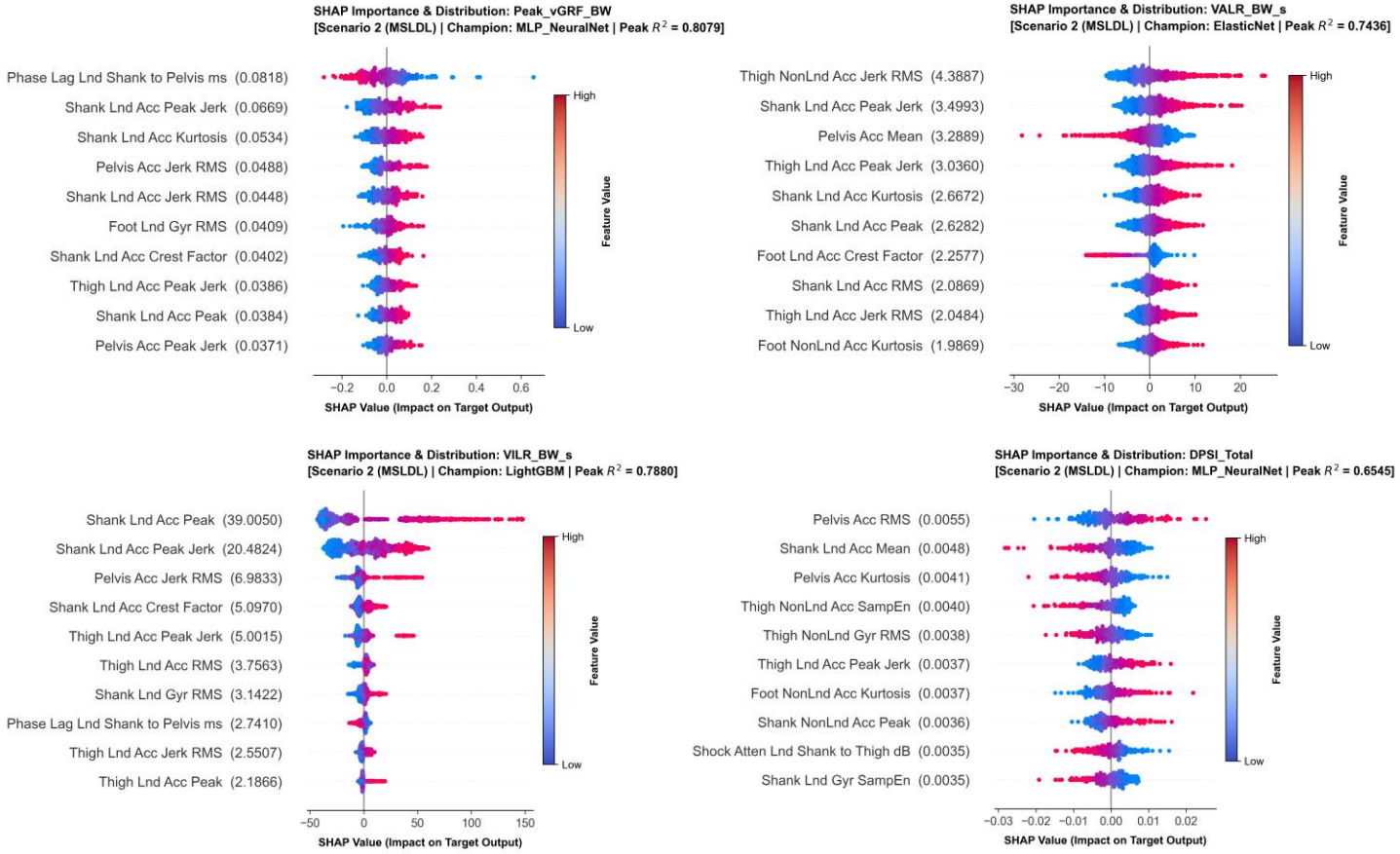


Figure 5 – SHAP analyses for Peak vGRF, VALR, VILR, and DPSI Total for scenario 2; SHAP values are shown inside the bracket

For predicting kinetic and postural control related variables in MSLDL condition, Figure 5 shows that phase lag of landing shank and pelvis (SHAP = 0.0818) followed by landing shank accelerometer peak jerk (SHAP = 0.0669) and landing shank accelerometer kurtosis (SHAP = 0.0534) are the three most impactful predictors to predict Peak vGRF using MLP model. Regarding the VALR, the three most impactful predictors are non-landing thigh accelerometer jerk RMS (SHAP = 4.3887), landing shank accelerometer peak jerk (SHAP = 3.4993), and pelvis accelerometer mean (SHAP = 3.2889). For the VILR, SHAP analysis on LightGBM model reveals the three most impactful predictors as landing shank accelerometer peak (SHAP = 39.005), landing shank accelerometer peak jerk (SHAP = 20.4824), and pelvis accelerometer jerk RMS (SHAP = 6.9833). Regarding the prediction of DPSI Total, the three most impactful input features are pelvis accelerometer RMS (SHAP = 0.0055), landing shank accelerometer mean (0.0048), and pelvis accelerometer kurtosis (SHAP = 0.0041).

In both SLDL and MSLDL scenarios, the reverse color presentation in the SHAP beeswarm visualizations (Figures 4 and 5) indicated that lower the phase lag values, the higher the impact to estimate the Peak vGRF it was.

4. Discussion

The primary aim of this study was to compare the performance of ML algorithms for predicting single-leg drop landing (SLDL) and multi-planar single-leg drop landing (MSLDL) kinetics, including Peak vGRF, vertical loading rates (VALR and VILR), as well as dynamic postural stability (DPSI) and time-to-stabilization variables (TTS) using IMU data. Sensor configurations with anatomical placements, and modalities (A, G, combined A+G) were pooled for finding the best-performing models. By systematically evaluating 5,508 models across 48 collegiate athletes, our findings provide novel insights into wearable biomechanical characteristics: (1) linear and gradient-boosted models as well as the deep learning neural network (ElasticNet, LightGBM, and MLP) accurately estimate transient vertical landing impact kinetics (Peak vGRF, VALR, VILR; $R^2 = 0.744–0.815$), whereas MLP also excels at dynamic stability indices (DPSI Total; $R^2 = 0.641–0.655$); (2) a single-mounted IMU sensor on the landing shank retained over 90% of the predictive capacity when comparing a full 9-sensor configuration for vertical loading kinetics (i.e., Peak vGRF, VALR, VILR). Either A sensor only or combined (A+G) modality could produce the best-performing prediction accuracy on various target landing metrics; and (3) explainable SHAP analysis and decomposition confirms that most model predictions were highly influenced by early-impact jerk dynamics (e.g., landing shank peak jerk), and the landing compliance by inter-segmental phase lags (e.g., phase lag between landing shank and pelvis).

4.1 Diversified Machine Learning Performance Across Kinetic Domains

The prediction accuracy was diverse across ML families depending on the mechanical domain being evaluated. For Phase 1 impact loading metrics (during 0–250 ms), when considering both R^2 and nRMSE, ElasticNet achieved the highest prediction capability and accuracy for VALR in SLDL ($R^2 = 0.815$, nRMSE $= 7.98\%$) and MSLDL ($R^2 = 0.744$, nRMSE $= 6.44\%$), alongside VILR in SLDL ($R^2 = 0.804$). This can be considered in relation to Newton's law ($F = ma$): whole-body force depends on CoM acceleration, whereas a shank-mounted IMU measures local segmental acceleration; the relationship between these quantities must therefore be learned and validated. ElasticNet is a regularized linear model with combined L_1/L_2 penalties that may capture useful impact-related information while limiting coefficient instability from correlated inputs. In contrast, for MSLDL VILR metric, tree-based ensembles (LightGBM: $R^2 = 0.788$; Random Forest: $R^2 = 0.785$) outperformed linear regression. Multi-planar landings introduce rotational shear and variable touchdown angles that potentially create non-linear conditions. The advantages of tree ensembles are compatible with broader benchmarking evidence showing strong performance on medium-sized tabular datasets, but those benchmarks do not establish a mechanism for the present landing results or eliminate overfitting concerns [32]. The performance of regularized linear models is also consistent with Bogaert et al. [33], who used a single sacral IMU and Lasso regression to predict running contact time, active and impact peaks, and impulse. Their impact peak error was greater than their active peak error, illustrating that even within one task the prediction target affects model performance. Derie et al. [22] evaluated ElasticNet, LASSO, and gradient-boosted regression using tibial acceleration features during overground running, with the lowest mean absolute errors obtained by gradient boosting. These running studies support comparing linear and nonlinear estimators, but their sensors, targets, and validation settings differ from the current SLDL/MSLDL experiment. The present findings extend the question of outcome dependent model performance to SLDL and MSLDL by showing that ElasticNet was particularly effective for VALR, whereas LightGBM provided an advantage for the more non-linear VILR outcome estimation in MSLDL tasks.

Likewise, MLP neural network produced superior predictive performance for Peak vGRF across both SLDL ($R^2 = 0.814$, best nRMSE $= 6.61\%$) and MSLDL ($R^2 = 0.808$, best nRMSE $= 6.72\%$), while demonstrating the lowest omnibus error dispersion across sensor configurations (mean nRMSE $\approx 10.03\%$, compared to 15% to 16% in linear and tree-based models). A possible explanation for this outcome-specific advantage is that the MLP represented interactions among the available kinematic features more effectively than the other fitted models. This remains a model-level interpretation, not evidence that a specific physical mechanism was learned. Landing experiments demonstrate contributions from multiple joints of lower limbs to energy absorption [34], but the current predictions do not separately identify those joint contributions.

Direct wearable landing comparisons nevertheless support the feasibility of neural-network estimation. Sun et al. [20] estimated vGRF and knee extension moments during single- and double-leg drop landings using modular LSTM models. Cerfoglio et al. [35] used three IMUs on the sacrum, thigh, and shank with a nonlinear autoregressive network to estimate GRF and knee-moment waveforms during the first landing of a vertical drop jump. Their study included 11 participants and therefore provides proof-of-concept evidence rather than a definitive performance benchmark for the present multi-planar task. Conversely, Song et al. [25] found that lightweight SVD-embedding and k-nearest-neighbor regression could match or outperform LSTMs during running. Taken together, these comparisons support selecting models by the target outcome and validation performance rather than assuming that greater architectural complexity is always advantageous.

Regarding the landing and stabilization phase (0–3.0 s), MLP neural network substantially outperformed linear and tree-based models under both conditions. In the SLDL condition, none of the linear or conventional ML algorithms achieved best predictive performance for dynamic postural stability (ElasticNet, SVR, RF, LightGBM, and XGBoost yielded $R^2 \leq 0.207$ for DPSI AP and ML) whereas, MLP achieved $R^2 = 0.509$ for DPSI AP, $R^2 = 0.478$ for DPSI ML, and $R^2 = 0.641$ (nRMSE $= 4.17\%$) for DPSI Total. In the MSLDL condition, the MLP likewise showed superior predictive capabilities across DPSI Total ($R^2 = 0.655$, nRMSE $= 4.09\%$), DPSI AP ($R^2 = 0.513$), and DPSI ML ($R^2 = 0.471$). The greater predictive performance of the MLP may reflect how the available features represent coordinated motion during the stabilization interval. Postural control involves interacting sensorimotor processes rather than one isolated joint response [36]. However, wearable acceleration, angular velocity, and force platform stability measures represent different physical quantities and require outcome-specific validation, as emphasized in a systematic review of standing balance assessment [37]. The MLP may have captured nonlinear associations among these measured features, but the study did not directly test the proposed neural control explanation. Its superiority for the present DPSI outcomes should therefore not be generalized to every balance measure, sensor configuration, or population.

For outcomes derived from the quiet stance phase, particularly the TTS-related features (3.0 to 15.0 s), the models demonstrated only low to moderate predictability ($R^2 = 0.324$–0.386 for F_x; $R^2 = 0.147$–0.333 for F_y). TTS derived using the Ross Range implementation depends on the baseline force bandwidth and the fitted post-landing trajectory. Methodological research shows that calculation method, sample rate, filtering, and trial duration can affect TTS outcomes [11]. Consequently, lower prediction performance may reflect both the information available in the IMU features and the sensitivity of the reference outcome whereas, the present analysis cannot separate these explanations. Akiyama et al. [38] illustrate why stability outcomes should not be treated as interchangeable: their CNN classified margin of stability categories during a multi-condition treadmill gait experiment with accuracies above 90%, while the margin of stability correlated weakly with other gait stability indices. Those classification results in nine healthy men are not directly comparable to continuous TTS regression after landing, but they support interpreting wearable stability performance in relation to the precise target definition. Validation of standing balance IMU measures likewise requires attention to the physical quantity and reference measure being compared [37].

4.2 Implications from Sensor Placement

Despite the strong predictive capabilities on certain kinetic metrics, it is also crucial for optimizing the practical feasibility for field application and deployment. In this regard, the optimal trade-off between required sensor quantity and predictive accuracy must be identified. Our study demonstrated over 90% accuracy retention for vertical impact kinetics in SLDL condition: 90.1% for Peak vGRF ($R^2 = 0.73$), 91.9% for VALR ($R^2 = 0.74$), and 99.6% for VILR ($R^2 = 0.77$) using a single shank-mounted IMU when compared to 9 sensors placement. A potential advantage of tibial placement is that the sensor can be secured close to the underlying bone, although attachment, soft-tissue motion, and orientation remain measurement considerations [18]. Derie et al. [22] provide task adjacent evidence that shin acceleration features can predict vertical loading rate during running. Nevertheless, the review by Xiang et al. [39] emphasizes that tibial acceleration, external GRFs, and internal tibial loading are related but distinct measurements. Thus, the current shank result supports

estimation of the specified force platform outcomes within the tested conditions whereas, it does not establish that one tibial sensor universally measures tissue loading or injury risk.

Conversely, midfoot placement performed poorly for Peak vGRF ($R^2 = 0.49$). Sensor coupling and local foot motion are possible explanations, but these were not measured directly. Attachment-related sensitivity should therefore be treated as a hypothesis rather than an established reason for the lower performance [18]. Nevertheless, it is worth noting that the foot sensor produced more meaningful estimation in frontal plane stability (DPSI ML single-sensor $R^2 = 0.21$ vs. 0.06 for shank). Directional landing research documents task dependent ankle kinematics [28], providing a rationale for studying foot-derived features. However, the present comparison does not establish that subtalar motion was the primary determinant of mediolateral GRF or DPSI.

For dynamic postural stability, single-sensor configurations exhibited only moderate accuracy retention (DPSI Total $R^2 = 0.48$, 74.4% retention). Peak DPSI estimation ($R^2 = 0.641$–0.655), required multi-segment lower body configurations particularly 7 or 9 sensors setups in our study. These results suggest that multi-segment measurements retained information relevant to the current force-based stability targets. Standing balance research commonly uses a lower back sensor, sometimes supplemented by limb sensors, but placement and sensor number depend on the task and intended measure [37]. The present advantage of seven- or nine-sensor configurations should therefore be viewed as a result for the tested DPSI targets, rather than evidence that more sensors always improve balance assessment. The distributed sensorimotor basis of postural control provides physiological context [36], while a deployable configuration still requires a separate assessment of accuracy, repeatability, setup burden, and target population performance.

Despite a similar trend in MSLDL observed, it is worth noting for some minor discrepancies when compared to SLDL condition. Both the landing thigh in single-sensor and thigh-foot combo in 2-sensor configurations achieved comparable or better predictive accuracy for loading rates (VALR: $R^2 = 0.72$; VILR: $R^2 = 0.78$) in MSLDL. A plausible interpretation is that additional directional and rotational components changed which segmental features were most informative. Multidirectional landing research has shown that ankle kinematics vary with landing direction and ankle instability status [28] and therefore, this supports task-specific assessment, not a direct explanation of the present thigh-sensor ranking. In the present MSLDL VILR model, four of the ten most influential features were landing-thigh accelerometer-related. This model attribution is consistent with useful thigh-derived information, but does not by itself establish increased eccentric hip work or a particular shock dissipation mechanism. Sensor choice should therefore be based on the specific movement and target outcome, and the proposed physical explanation should be tested using synchronized joint-level biomechanics.

4.3 Contributions from Sensor Modalities: Accelerometers vs. Gyroscopes

Apart from the quantity and location of sensor placement, our results from various sensor modalities confirmed that isolated accelerometers (A) or combined (A+G) consistently outperformed gyroscopes alone (G). For landing kinetic metrics during the impact phase (Peak vGRF, VALR, VILR), A-sensor-only models achieved $R^2 = 0.79$–0.81, closely matching the combined A+G ($R^2 = 0.80$–0.81), whereas G-sensor-only models achieved lower predictive accuracies ($R^2 = 0.48$–0.61). This pattern is compatible with the impulse momentum relationship ($F\Delta t = m\Delta v$), whereas gyroscopes measure rotational motion. Nevertheless, the impulse momentum relationship applies to the relevant system dynamics whereas, a local segment acceleration is not a direct measurement of whole body force or internal tissue loading [14].

In contrast, G sensors provided valuable information for estimating dynamic stability metrics. By combining accelerometers with gyroscopes, our study yielded the highest predictive accuracy for DPSI AP ($R^2 = 0.51$ for A+G vs. 0.35 for A-only) and DPSI ML ($R^2 = 0.48$ for A+G vs. 0.33 for A-only). Post-landing stabilization involves both translational and rotational segment motions, so accelerometers and gyroscopes can provide complementary rather than interchangeable measurements [15]. Direct landing evidence is available from Chaaban et al. [40]: in 26 healthy young adults performing a double-limb jump-landing task, multiple feature linear models combining accelerometer and gyroscope inputs improved prediction of

peak vGRF and knee biomechanics compared with single feature approaches. Their reported multiple feature models spanned R^2 values of 0.68–0.94 across different outcomes. However, their trial-wise cross-validation allowed records from the same participant in different folds, whereas the present methods specify subject-isolated nested validation; their accuracy values should not be treated as a head-to-head benchmark. Their targets were also impact and knee variables, not DPSI or TTS. Thus, the study supports the potential complementarity of modalities during landing, while the specific A+G advantage for the current stability targets is evidence from the present experiment.

4.4 SHAP Analyses and Decomposition of Model Feature Importance

Beyond predictive accuracy, SHAP analysis identified features that contributed strongly to the fitted models. Across SLDL and MSLDL, landing-shank-to-pelvis phase lag and landing-shank acceleration kurtosis were prominent features for Peak vGRF and VALR. Acceleration kurtosis describes the shape of the signal distribution, while inter-segmental phase lag describes the timing of measured segmental responses. The beeswarm patterns indicated associations between these features and predicted impact outcomes. Landing research has demonstrated that lower-limb joints contribute differently to energy absorption [34], providing a biomechanical context for considering multi-segment behavior. However, statistical feature attribution is not equivalent to identifying a causal physical mechanism. Janzing et al. [41] explain that feature relevance calculations depend on distinctions between observational and interventional distributions. Accordingly, the current SHAP rankings identify contributions within fitted predictive models [24], not causal effects of changing joint compliance, landing stiffness, or segmental coordination. Those proposed mechanisms require targeted biomechanical or interventional validation.

In MSLDL, rotational and lateral shearing forces upon landing may increase the importance of non-landing limb kinematics and multi-axial jerk. For VALR, non-landing thigh jerk RMS appeared to be the top predictor (SHAP $= 4.3887$), followed by landing shank peak jerk (SHAP $= 3.4993$) and pelvis mean acceleration (SHAP $= 3.2889$). When athletes land with multi-planar directional perturbation, the non-dominant (i.e., contralateral) limb may need to react dynamically to help counterbalance to achieve the equilibrium. It is speculated that rapid corrections of the non-landing limb provide compensatory torques across the pelvis that can potentially modulate stance-leg loading rates. Future studies in this regard are warranted to clarify the relationship between non-landing limb kinematics, joint moments and loading rates of MSLDL condition. Similarly, for DPSI Total, pelvis accelerometer RMS (SHAP $= 0.0040–0.0055$) and non-landing segment entropy were primary contributors. Postural control coordinates body segments to maintain orientation and equilibrium [36], but a larger pelvic acceleration feature cannot be interpreted as poorer control without considering the task and target definition. The contribution of pelvis-derived features is therefore an association within the current fitted models, not proof of a deficient stabilization strategy [41].

4.5 Limitations and Future Research Directions

Several limitations should be acknowledged. First, the standardized 30 cm platform and pre-determined landing directions do not reproduce unanticipated or contact-related sporting situations. Second, discrete feature engineering summarizes continuous signals, and alternative waveform-based models should be evaluated using comparable targets and validation procedures rather than presumed superior. Third, validation design affects the interpretation of prediction accuracy. Saeb et al. [42] showed that record-wise splitting can produce overly optimistic estimates when the intended application is prediction in new participants. The subject-isolated nested design described in the present methods addresses this concern within the available sample, but does not replace independent testing across sessions, devices, or external cohorts. This distinction also limits numerical comparisons with studies using different splitting schemes [23]. Finally, the participants were healthy collegiate athletes. Clinical use after ACL or ankle injury, responsiveness to rehabilitation-related change, and performance

in field settings require separate validation. Neither predictive accuracy nor SHAP feature importance establishes prospective injury risk prediction.

5. Practical and Clinical Applications in Sports Injury Prevention

The ML predictive analytics in this study provides empirical evidence and insights for sports medicine practitioners, and strength and conditioning specialists for bridging the gap between laboratory biomechanics and on-field screening. By demonstrating that a single shank-mounted IMU retains over 90% predictive accuracy for Peak vGRF/BW ($R^2 = 0.73$) and VALR ($R^2 = 0.74$), practitioners can screen the single-leg landing biomechanics using single-mounted sensor (shank, thigh, or foot depending on the metrics) in the field quickly. Despite the only moderate prediction accuracy for dynamic postural control metric (i.e., DPSI Total), 2-sensor (bilateral feet placement) can be an option to provide a non-specific SLDL or MSLDL postural control and stability assessment as a gross estimation. Coaches or team medical practitioners can pay more attention to jumping athletes flagged to have "stiff landing" or "poor postural control" with the predicted landing kinetic profiles for further detailed biomechanical assessments or landing-specific neuromuscular injury prevention training prior to competitive seasons.

Similarly, for the rehabilitation progression or return-to-sport decision making, injured athletes such as those following ACL reconstruction, patellar tendinopathy, or lateral ankle sprains may be preliminarily screened via single thigh/shank IMU sensor for revealing the single-leg landing qualities including impact and loading rates, or via 2-IMU-sensors on bilateral feet for supplementing the observation of SLDL/MSLDL postural control and stability progress. Instead of relying on the IMU-based screening alone, therapists and clinicians may also make use of this instant SLDL or MSLDL assessment to supplement their qualitative judgments as the real-time biofeedback to enhance the neuromuscular and motor control training effectiveness. By combining the qualitative clinical biomechanics and the use of IMU-based objective SLDL biomechanical analyses, a more complete biomechanical landing profile can be built before making any clinical decisions. Future studies should verify if the estimated landing kinetics are valid to reflect neuromuscular fatigue, in particular, the monitoring and tracking of landing compliance after cumulative fatigue. Furthermore, more studies regarding the use of IMU-estimated landing kinetics and postural control metrics for lower limb sports injury risk estimation are warranted. For further extending the applications in injured populations, real-time feedback, and prospective injury-risk screening, more validations beyond the healthy individuals or athletes in the laboratory setting should be made.

6. Conclusions

This study showed that single-mounted (shank or thigh) wearable IMU sensor combined with ML or neural network predict laboratory force platform landing kinetics with strong accuracy, and dynamic stability indices during both SLDL and MSLDL tasks with moderate accuracy. SHAP analyses identified jerk-related and inter-segmental timing features that contributed to the fitted predictions; these associations do not by themselves establish shock-attenuation mechanisms. Our findings establish an accessible and scalable wearable methodology for the potential applications in on-field sports injury risk profiling and objective rehabilitation monitoring in athletic populations.

Declarations

Ethics Approval and Consent to Participate: The study protocol was approved by the Institutional Review Board and Human Subjects Ethics Committee of the host institution, Technological and Higher Education Institute of Hong Kong (HSEC Ref. No:HE2025-01). Written informed consent was obtained from all individual participants prior to participation.

Consent for Publication: Not applicable.

Availability of Data and Materials: The rectified tabular machine learning datasets and analysis code generated during the current study are available from the corresponding author upon reasonable academic request.

Competing Interests: The authors declare that they have no competing financial or non-financial interests.

Funding: Nil.

Authors' Contributions: Conceptualization: [Indy M.K. Ho, Daniel H.K. Chow, Lana Ružić]; Methodology: [Indy M.K. Ho, Daniel H.K. Chow]; Software and Machine Learning Pipeline: [Indy M.K. Ho, Jialiang Gu]; Formal Analysis and Investigation: [Indy M.K. Ho, Jialiang Gu]; Writing – Original Draft Preparation: [Indy M.K. Ho]; Writing – Review & Editing: [Indy M.K. Ho, Daniel H.K. Chow, Lana Ružić]; Supervision: [Indy M.K. Ho, Daniel H.K. Chow]. All authors read and approved the final manuscript.

Use of Generative AI and AI-Assisted Technologies: During the preparation of this work, the authors utilized Gemini (Google) to assist with Python machine learning and neural network scripting syntax (specifically implementing cross-validation pipelines, hyperparameter grid tuning, and SHAP analysis) and Perplexity for structural copyediting and academic word checks for rigorousness, conciseness, and precision. The tool was not used to synthesize, generate, or manipulate raw experimental data, nor did it substitute for scientific interpretation. All computational scripts, statistical results, and final manuscript text were critically evaluated, independently verified, and approved by the authors, who take full academic responsibility for the integrity of the published work.

References

1. Chia, L. et al. Non-contact anterior cruciate ligament injury epidemiology in team-ball sports: a systematic review with meta-analysis by sex, age, sport, participation level and exposure type. *Sports Med.* **52,** 2447–2467 (2022). https://doi.org/10.1007/s40279-022-01697-w

2. Mineta, S., Koyama, T., Yamaguchi, S., Inagaki, H. & Sekine, Y. Epidemiology of lateral ankle sprain focusing on indirect contact mechanism in male and female soccer players: an 18.5-month cohort study. *Injury* **55,** 111699 (2024). https://doi.org/10.1016/j.injury.2024.111699

3. Tayfur, A. et al. Are landing patterns in jumping athletes associated with patellar tendinopathy? A systematic review with evidence gap map and meta-analysis. *Sports Med.* **52,** 123–137 (2022). https://doi.org/10.1007/s40279-021-01550-6

4. Kim, C. et al. Effects of patellofemoral pain syndrome on changes in dynamic postural stability during landing in adult women. *Appl. Bionics Biomech.* **2022,** 7452229 (2022). https://doi.org/10.1155/2022/7452229

5. Ardern, C. L., Taylor, N. F., Feller, J. A. & Webster, K. E. Fifty-five per cent return to competitive sport following anterior cruciate ligament reconstruction surgery: an updated systematic review and meta-analysis including aspects of physical functioning and contextual factors. *Br. J. Sports Med.* **48,** 1543–1552 (2014). https://doi.org/10.1136/bjsports-2013-093398

6. Koga, H. et al. Mechanisms for noncontact anterior cruciate ligament injuries: knee joint kinematics in 10 injury situations from female team handball and basketball. *Am. J. Sports Med.* **38,** 2218–2225 (2010). https://doi.org/10.1177/0363546510373570

7. Wang, L. I. The lower extremity biomechanics of single- and double-leg stop-jump tasks. *J. Sports Sci. Med.* **10,** 151–156 (2011).

8. Wikstrom, E. A., Tillman, M. D., Smith, A. N. & Borsa, P. A. A new force-plate technology measure of dynamic postural stability: the dynamic postural stability index. *J. Athl. Train.* **40,** 305–309 (2005).

9. Kawaguchi, K. et al. Dynamic postural stability is decreased during the single-leg drop landing task in male collegiate soccer players with chronic ankle instability. *Orthop. J. Sports Med.* **10,** 23259671221107343 (2022). https://doi.org/10.1177/23259671221107343

10. Tan, T. et al. A scoping review of portable sensing for out-of-lab anterior cruciate ligament injury prevention and rehabilitation. *npj Digit. Med.* **6,** 46 (2023). https://doi.org/10.1038/s41746-023-00782-2

11. Fransz, D. P., Huurnink, A., de Boode, V. A., Kingma, I. & van Dieën, J. H. Time to stabilization in single leg drop jump landings: an examination of calculation methods and assessment of differences in sample rate, filter settings and trial length on outcome values. *Gait Posture* **41,** 63–69 (2015). https://doi.org/10.1016/j.gaitpost.2014.08.018

12. Withrow, T. J., Huston, L. J., Wojtys, E. M. & Ashton-Miller, J. A. The relationship between quadriceps muscle force, knee flexion, and anterior cruciate ligament strain in an in vitro simulated jump landing. *Am. J. Sports Med.* **34,** 269–274 (2006). https://doi.org/10.1177/0363546505280906

13. Quatman, C. E. & Hewett, T. E. The anterior cruciate ligament injury controversy: is "valgus collapse" a sex-specific mechanism? *Br. J. Sports Med.* **43,** 328–335 (2009). https://doi.org/10.1136/bjsm.2009.059139

14. Matijevich, E. S., Branscombe, L. M., Scott, L. R. & Zelik, K. E. Ground reaction force metrics are not strongly correlated with tibial bone load when running across speeds and slopes: implications for science, sport and wearable tech. *PLoS One* **14,** e0210000 (2019). https://doi.org/10.1371/journal.pone.0210000

15. Camomilla, V., Bergamini, E., Fantozzi, S. & Vannozzi, G. Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: a systematic review. *Sensors* **18,** 873 (2018). https://doi.org/10.3390/s18030873

16. Hernandez, V., Dadkhah, D., Babakeshizadeh, V. & Kulić, D. Lower body kinematics estimation from wearable sensors for walking and running: a deep learning approach. *Gait Posture* **83,** 185–193 (2021). https://doi.org/10.1016/j.gaitpost.2020.10.026

17. Fathian, R., Khandan, A., Chiu, L. Z. F. & Rouhani, H. Assessment of countermovement jump with and without arm swing using a single inertial measurement unit. *Sports Biomech.* **23,** 2390–2407 (2024). https://doi.org/10.1080/14763141.2022.2032296

18. Sheerin, K. R., Reid, D. & Besier, T. F. The measurement of tibial acceleration in runners—a review of the factors that can affect tibial acceleration during running and evidence-based guidelines for its use. *Gait Posture* **67,** 12–24 (2019). https://doi.org/10.1016/j.gaitpost.2018.09.017

19. Alcantara, R. S., Edwards, W. B., Millet, G. Y. & Grabowski, A. M. Predicting continuous ground reaction forces from accelerometers during uphill and downhill running: a recurrent neural network solution. *PeerJ* **10,** e12752 (2022). https://doi.org/10.7717/peerj.12752

20. Sun, T., Li, D., Fan, B., Tan, T. & Shull, P. B. Real-time ground reaction force and knee extension moment estimation during drop landings via modular LSTM modeling and wearable IMUs. *IEEE J. Biomed. Health Inform.* **27,** 3222–3233 (2023). https://doi.org/10.1109/JBHI.2023.3268239

21. Bradach, M. M. et al. The effects of a simple sensor reorientation procedure on peak tibial accelerations during running and correlations with ground reaction forces. *Sensors* **23,** 6048 (2023). https://doi.org/10.3390/s23136048

22. Derie, R. et al. Tibial acceleration-based prediction of maximal vertical loading rate during overground running: a machine learning approach. *Front. Bioeng. Biotechnol.* **8,** 33 (2020). https://doi.org/10.3389/fbioe.2020.00033

23. Halilaj, E. et al. Machine learning in human movement biomechanics: best practices, common pitfalls and new opportunities. *J. Biomech.* **81,** 1–11 (2018). https://doi.org/10.1016/j.jbiomech.2018.09.009

24. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* **30** (eds Guyon, I. et al.) (Curran Associates, 2017). https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

25. Song, B. et al. Estimating ground reaction forces from inertial sensors. *IEEE Trans. Biomed. Eng.* **72,** 595–608 (2025). https://doi.org/10.1109/TBME.2024.3465373

26. Xiang, L. et al. Recent machine learning progress in lower limb running biomechanics with wearable technology: a systematic review. *Front. Neurorobot.* **16,** 913052 (2022). https://doi.org/10.3389/fnbot.2022.913052

27. Nardon, M. et al. Kinematics but not kinetics alterations to single-leg drop jump movements following a subject-tailored fatiguing protocol suggest an increased risk of ACL injury. *Front. Sports Act. Living* **6,** 1418598 (2024). https://doi.org/10.3389/fspor.2024.1418598

28. Sonsukong, A. et al. Ankle biomechanics during multidirectional landings in athletes with chronic ankle instability. *Foot Ankle Spec.* **17,** 249–258 (2024). https://doi.org/10.1177/19386400231208522

29. Ross, S. E., Guskiewicz, K. M. & Yu, B. Single-leg jump-landing stabilization times in subjects with functionally unstable ankles. *J. Athl. Train.* **40,** 298–304 (2005).

30. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2,** 56–67 (2020). https://doi.org/10.1038/s42256-019-0138-9

31. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning* (Proceedings of Machine Learning Research **70**), 3319–3328 (2017). https://proceedings.mlr.press/v70/sundararajan17a.html

32. Grinsztajn, L., Oyallon, E. & Varoquaux, G. Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems* **35** (eds Koyejo, S. et al.), 507–520 (Curran Associates, 2022). https://doi.org/10.52202/068431-0037

33. Bogaert, S., Davis, J. & Vanwanseele, B. Predicting vertical ground reaction force characteristics during running with machine learning. *Front. Bioeng. Biotechnol.* **12,** 1440033 (2024). https://doi.org/10.3389/fbioe.2024.1440033

34. Decker, M. J., Torry, M. R., Wyland, D. J., Sterett, W. I. & Steadman, J. R. Gender differences in lower extremity kinematics, kinetics and energy absorption during landing. *Clin. Biomech. (Bristol, Avon)* **18,** 662–669 (2003). https://doi.org/10.1016/S0268-0033(03)00090-1

35. Cerfoglio, S. et al. Machine learning-based estimation of ground reaction forces and knee joint kinetics from inertial sensors while performing a vertical drop jump. *Sensors* **21,** 7709 (2021). https://doi.org/10.3390/s21227709

36. Horak, F. B. Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls? *Age Ageing* **35,** ii7–ii11 (2006). https://doi.org/10.1093/ageing/afl077

37. Ghislieri, M., Gastaldi, L., Pastorelli, S., Tadano, S. & Agostini, V. Wearable inertial sensors to assess standing balance: a systematic review. *Sensors* **19,** 4075 (2019). https://doi.org/10.3390/s19194075

38. Akiyama, Y., Kazumura, K., Okamoto, S. & Yamada, Y. Utilizing inertial measurement units for detecting dynamic stability variations in a multi-condition gait experiment. *Sensors* **24,** 7044 (2024). https://doi.org/10.3390/s24217044

39. Xiang, L. et al. Rethinking running biomechanics: a critical review of ground reaction forces, tibial bone loading, and the role of wearable sensors. *Front. Bioeng. Biotechnol.* **12,** 1377383 (2024). https://doi.org/10.3389/fbioe.2024.1377383

40. Chaaban, C. R. et al. Combining inertial sensors and machine learning to predict vGRF and knee biomechanics during a double limb jump landing task. *Sensors* **21,** 4383 (2021). https://doi.org/10.3390/s21134383

41. Janzing, D., Minorics, L. & Bloebaum, P. Feature relevance quantification in explainable AI: a causal problem. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (Proceedings of Machine Learning Research **108**), 2907–2916 (2020). https://proceedings.mlr.press/v108/janzing20a.html

42. Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C. & Kording, K. P. The need to approximate the use-case in clinical machine learning. *GigaScience* **6,** 1–9 (2017). https://doi.org/10.1093/gigascience/gix019