Journal of Biomechanics 206 (2026) 113526

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech

Short communication

Viability of automating the landing error scoring system using inertial measurement units

Zane Hamilton [a,*], Jessica Turner [a], Jemma König [a], Kim Hébert-Losier [b]

[a] School of Computing and Mathematical Sciences, University of Waikato, Tauranga, New Zealand
[b] School of Sport and Human Movement, Te Wānanga Waiora Division of Health, University of Waikato, Tauranga, New Zealand

ARTICLE INFO

Keywords:
Anterior cruciate ligament
Movement screen
Injury risk
Machine learning

ABSTRACT

The Landing Error Scoring System (LESS) is an assessment tool used for identifying movement patterns linked with non-contact anterior cruciate ligament injuries during a double-leg jump-landing; however, the LESS is scored by experts using 2D video recordings, which limits large-scale screening. This study explores the viability of using inertial measurement unit (IMU) data to automate scoring of 17 LESS items, using conventional machine learning model architecture. Forty healthy participants completed six jumps each, and raw movement data from three IMU sensors placed on the sacrum and medial-inferior aspects of the tibias were processed and segmented into the key phases of the double-leg jump-landing. A total of 218 jumps were used to train supervised machine learning models using various subsets of the processed dataset, including extracted temporal and statistical features. Performance metrics were extracted, and the best performing models for each scoring item were evaluated against a majority-class baseline classifier (ZeroR). Results indicated moderate improvements over ZeroR, ranging from 0.3 % to 22.5 %, with some models demonstrating limited recognition of the minority class. These findings provide a foundation for future research in IMU-based LESS automation, and improving the accessibility of movement screening tools.

1. Introduction

Lower extremity injuries are the most common injury site across sports, with injury to the anterior cruciate ligament (ACL) being among the leading knee-related injuries worldwide (Moses et al., 2012). The Landing Error Scoring System (LESS) is a movement screening tool made up of 17 total scoring items (Appendix A), aimed at identifying movement patterns from a double-leg jump-landing (DLJL) task linked with non-contact ACL injury (Padua et al., 2009). The first 15 items are classified as either a “1″ when an error is present, or a “0” if an error is absent. Both items 16 and 17 use a three-point scale that is more subjective, where “0” indicates an excellent landing, “1” an average landing, and “2” a poor landing.

The LESS has been used to inform targeted interventions by identifying specific movements associated with high-risk landing mechanics (Padua et al., 2015). Athletes demonstrating poor landing technique can be directed toward neuromuscular training programmes that emphasise landing technique retraining such as plyometric exercises, movement-control drills, balance activities and other neuromuscular exercises (Hanzlíková et al., 2021). The LESS has also been used as an outcome measure to evaluate improvements in landing mechanics following intervention (Hanzlíková et al., 2021). For example, decreases in LESS scores have been reported following a 6-week neuromuscular training programme in collegiate basketball athletes (Pfile et al., 2016), an 8-week coach-led injury-prevention programme in youth soccer players (Pryor et al., 2017), and the SportsMetrics programme in both soccer and volleyball athletes (Mohammadi & Khosravani, 2024; Saki et al., 2021). These findings demonstrate the utility of the LESS for monitoring changes in movement quality and assessing intervention effectiveness over time.

A lower total LESS score indicates fewer movement errors and patterns linked to ACL injury; therefore, a lower risk of non-contact ACL injury and a better jump landing technique. For individuals who scored 5 or more in their overall LESS score, the risk ratio for non-contact ACL injury was 10.7 times greater than individuals with scores less than 5 (Padua et al., 2015). Scoring using the LESS requires a trained rater to score every participant from 2D video recordings, limiting the throughput possible in large-scale screening contexts due to the time and expertise required (Hébert-Losier et al., 2020). By automating the LESS scoring, there is greater accessibility for field or training environment

* Corresponding author.
E-mail address: zane.hamilton@waikato.ac.nz (Z. Hamilton).

https://doi.org/10.1016/j.jbiomech.2026.113526
Accepted 13 August 2026
Available online 14 August 2026

testing and potential for immediate feedback. Inertial measurement units (IMUs) are commonly used to capture human movement data (Wang et al., 2023), and various machine learning models have used movement data provided by IMUs as input to classify jump-related performance outcomes in research, such as identifying key phases of a jump (Chiang et al., 2020; Hendry et al., 2020), jump height (Mascia et al., 2023), or the estimation of ground reaction forces (Chaaban et al., 2021; Cerfoglio et al., 2021; Hendry et al., 2020). Given the success of wearable sensors in jump-landing analysis, we propose scoring the LESS may no longer require cameras. We investigate the applicability of IMU sensor data for scoring the LESS from a DLJL task; aimed at providing greater accessibility to LESS scores in field and training environments. To our knowledge, no prior study has evaluated whether IMU-derived features can classify the LESS scoring process using conventional machine learning models.

2. Methods

This study used historical IMU data from a previous collection involving the LESS to focus on the accuracy of the algorithms without interfering with hyperparameter tuning or data collection techniques. Forty participants (21 men, 17 women, 2 others, 20.3 ± 3.9 years, 176.1 ± 9.3 cm, 75.8 ± 12.8 kg) (Appendix B) performed 218 total jump occurrences (five to six jumps per participant). Each participant jumped horizontally from a 30 cm high box to 50 % of their body height, followed by an immediate vertical jump for maximal height. Three IMeasureU Blue Trident IMU sensors (Vicon Motion Systems Ltd., Oxford, UK) recorded accelerometer and gyroscope data at 1600 Hz and 1125 Hz, respectively. A minimal set of sensors was used with two sensors placed bilaterally on the medial tibia, just above the medial malleolus, and a sensor placed on the posterior aspect of the pelvis/sacrum, between the posterior superior iliac spines (Hébert-Losier et al., 2024).

The customised algorithms created as part of this paper are publicly available.[1] A low-pass fourth-order Butterworth filter was applied to the aligned dataset, selected based on its common application in related work (Cerfoglio et al., 2021; Chiang et al., 2020; Turner et al., 2025). Within the DLJL, the LESS items are scored between initial contact and maximum knee flexion (Padua et al., 2009). Two key phases were segmented: take-off to initial contact (S1), and initial contact to maximum knee flexion (S2). Segmentation of these key events was performed algorithmically using IMU-derived acceleration thresholds, with further validation against video-based timestamps. Movement data from S1 and S2 were merged into one dataset (S1A2) to evaluate whether pre-landing movement indicated tendencies toward high-risk patterns, compared to S2 alone. Sensor subsets were also investigated using solely ankle, pelvis or all available data.

The types of features extracted from the IMU data (Table 1) included statistical features based on prior jump-related work. Specifically, the mean and standard deviation were selected given their influence on the classification performance of jump types in trampoline-based gymnastics (Woltmann et al., 2023), and minimum and maximum values highlight differences in acceleration and rotational velocity peaks between participants in countermovement and jump-landing tasks (Chaaban et al., 2021; Chiang et al., 2020). Root Mean Square (RMS) was included to complement the mean, while variance characterised the signal spread. While these features are not specific to the LESS, they provide movement summaries that reduce sensitivity to the variability in time taken to perform the movement. Statistical measures were extracted from three axes across all three sensors, producing 54 measures per subset (6 features x 9 total axes of data). Features were then merged with participant characteristics for model integration, with additional subsets created using raw data, statistical features and all features combined.

Table 1
Selected Features to be Evaluated/Derived from Sensor Data.

Type	Feature
Temporal	Flight Time from the Box to Initial Contact (S1)
	Time from Initial Contact to Maximum Knee Flexion (S2)
	Total Movement Time (S1A2)
Statistical-Based	Maximum
	Mean
	Minimum
	Root Mean Square
	Standard Deviation
	Variance
Characteristic	Age (y)
	Body Mass Index (kg/m^2)
	Height (cm)
	Mass (kg)
	Gender
	Past Injury

Different classification tasks and datasets favour different algorithms (Gómez and Rojas, 2016), therefore seven common supervised machine learning algorithms were implemented in their standard configuration using the Scikit-Learn library for Python.[2] The models selected included K-nearest neighbours (KNN), Gaussian Naive Bayes (GNB), support-vector (SVC), gradient boosting (GBC), stochastic gradient descent (SGD), deep feedforward neural network (DFF), and random forest (RFC) classifiers. Five-fold cross validation was used, where individual jumps were randomly distributed across folds. Precision, recall and weighted F1-score were extracted from each model. Weighted F1-score acts as the harmonic mean of precision and recall (Naidu et al., 2023), useful for imbalanced datasets as it reflects performance across all classes rather than favouring the majority.

The overall scoring item counts in the data available were scored by a trained clinician (Table 2). With an imbalanced dataset, the classifiers will demonstrate bias toward the most frequently observed class; which leads to accuracy being a representation of the dataset imbalance, rather than the accuracy of the classifier. To mitigate this, the results from the best performing model were compared to a ZeroR approach for every scoring item. ZeroR predicts the majority class for all instances, ignoring all input features.

3. Results

Performance metrics were extracted from all the best performing models in each scoring item (Table 3). Models trained using statistical features generally outperformed those using raw data in overall F1-score, with 10 of the 14 best performing models relying on summary statistical descriptors, rather than the raw signals. The F1-score observed from classification of the overall LESS score was comparatively low (26.3 %), therefore results for classifying the overall score were not included in further analysis.

Precision and recall were varied across all the models, with some models indicating that scoring-item specific features had been recognised. For example, the GNB model used for classifying ankle plantarflexion (scoring item 4) was able to correctly classify 54 % of the less-dominant class. In contrast, classifying lateral trunk flexion at initial contact (scoring item 6) had a similar distribution of classes, however the RFC model was only able to correctly classify 3 % of the less dominant class. The performance in many models demonstrates an underlying link with class distribution, rather than individual model performance. For items with highly imbalanced classes (such as items 7, 12, and 14), the highest overall F1-scores are observed (89.8–92.9 %), with markedly low recall for the minority class (i.e., 17 % recall for scoring

[1] https://github.com/zanephamilton/Landing-Error-Scoring-System-Automation-Algorithms.

[2] version 1.4.1, https://scikit-learn.org/stable/supervised_learning.html.

Table 2
Distribution of Classes in the Landing Error Scoring System From 218 Jumps.

(a) Overall Landing Error Scoring System Score Counts

	1	2	3	4	5	6	7	8	9	10	11+
Total	2	2	4	22	52	54	47	17	14	4	0
Distribution (%)	0.9	0.9	1.8	10.1	23.9	24.8	21.6	7.8	6.4	1.8	0

(b) *Landing Error Scoring System Item Counts*

Scoring Item	Class Target 0	1	2	Distribution (%)
1	39	179	–	82
2	218	0	–	100
3	218	0	–	100
4	180	38	–	83
5	62	156	–	72
6	186	32	–	85
7	205	13	–	94
8	106	112	–	51
9	94	124	–	57
10	217	1	–	100
11	136	82	–	62
12	202	16	–	93
13	217	1	–	100
14	194	24	–	89
15	56	162	–	74
16	69	144	5	66
17	6	197	15	90

Table 3
Best Performing Model and Performance for all LESS Scoring Items.

LESS Scoring Item	Best Performing Model	Class Targets 0 Precision (%)	0 Recall (%)	1 Precision (%)	1 Recall (%)	2 Precision (%)	2 Recall (%)	Accuracy (%)	F1-score (%)
Total Score	GBC	–	–	–	–	–	–	27.9	26.3
1	DFF	42	26	85	93	–	–	80.7	77.9
4	GNB	91	96	82	54	–	–	89.0	87.8
5	RFC	73	31	78	95	–	–	77.1	73.3
6	RFC	86	98	20	3	–	–	84.4	78.9
7	RFC	95	100	40	17	–	–	95.0	92.9
8	KNN	57	63	61	54	–	–	58.3	57.4
9	GNB	55	34	62	80	–	–	60.0	57.0
11	SGD	70	73	54	50	–	–	64.3	63.7
12	SGD	93	99	27	13	–	–	92.7	90.2
14	RFC	92	99	63	32	–	–	91.7	89.8
15	RFC	62	35	81	93	–	–	78.4	75.4
16	GBC	59	43	75	84	30	40	70.2	68.8
17	GBC	83	100	100	92	80	100	93.2	94.9

item 7). In contrast, items with more balanced distributions (items 8 and 9) demonstrated lower performance, with weighted F1-scores around 57 %.

In multi-class items (16 and 17), the underlying link between distribution and performance was similar. Item 17 achieved the greatest performance across all scoring items (94.9 %), with a recall of 92 % for the dominant class (average landing condition) indicating only 8 % of the average classes were misclassified as either excellent or poor. The recall rate for both the less-dominant classes (poor and excellent) were 100 %, meaning they were all correctly classified. Item 16 did not perform similarly (68.8 %), with much lower recall in the minority classes (roughly 40 % for both stiff and soft landing classes).

All models demonstrated improvement over ZeroR (Table 4), however performance varied (0.3 % to 22.5 %). Improvement over ZeroR indicates that there was at least some learning of the minority classes. Although there was an underlying link with distribution across the LESS, these results indicate that for some of the scoring items, IMU-derived features were able to distinguish between classes beyond the distribution alone.

Table 4
Comparison between ZeroR and Best Performing Model F1-Score Performance.

Scoring Item	Weighted F1-score (%) ZeroR	Best Performing Model	Difference (%)
1	74.1	77.9	+3.8
4	74.7	87.8	+13.1
5	59.7	73.3	+13.6
6	78.6	78.9	+0.3
7	91.2	92.9	+1.7
8	34.9	57.4	+22.5
9	41.3	57.0	+15.7
11	47.9	63.7	+15.8
12	89.1	90.2	+1.1
14	83.8	89.8	+6.0
15	63.4	75.4	+12.0
16	52.6	68.8	+16.2
17	85.8	94.9	+9.1

4. Discussion

This study investigated the feasibility of automating individual LESS scoring items using IMU data and conventional machine learning models. Classifying the total LESS score alone performed poorly. The total LESS score aggregates flexion angles across multiple joints, foot contact asymmetry and joint displacement, making it a more complex classification than individual errors, reflected in the comparatively low F1-score (26.3 %). Future investigation should target individual scoring items to reduce inaccuracy and increase performance.

Class imbalance was a defining factor influencing the performance of each model. Several scoring items presented an extreme imbalance in distribution, with error-present classes occurring in fewer than 10 % for seven scoring items. Models achieved high F1-scores for these scoring items, but offered minimal improvement over ZeroR, indicating that they were largely predicting the majority (error-absent) class, rather than recognising the class itself. Items 7, 12 and 14 highlighted this pattern, where despite the largest class imbalances (90–94 %) producing the highest weighted F1-scores (89.8–94.9 %), ZeroR improvements were minimal (+1.7 % to 6.0 %), indicating majority-class prediction rather than genuine learning.

In comparison, items 8, 9, 11 and 16 showed more noticeable improvement over ZeroR (greater than 15.7 %), suggesting the data available contained identifiable movement information for partial recognition of stance width, asymmetric contact, and overall joint displacement. Classifying a narrow stance width (item 8) showed the largest improvement over ZeroR (+22.5 %), in contrast to identifying a wide stance width (item 7) which had one of the lower improvements (+1.7 %). These two scoring items are detecting the same stance positioning, however item 8 provided greater recognition of the minority class due to its balanced class distribution, allowing the classifiers the ability to learn from more error-present instances. Additionally, the model used for identifying a foot position with the toe facing inwards (item 9) demonstrated an improvement over ZeroR (+15.7 %), however there were no instances of a toe facing outward landing stance (item 10), meaning a direct comparison between foot positioning errors could not be made.

The primary limitations of this paper stem from the restricted sensor configuration and class imbalance. Many of the scoring items of the LESS involve the foot, ankle, knee, hip, and trunk; meaning the movement data captured at the tibia and sacrum are only in proximity to these movement patterns, not directly extracted from segments of interest. For example, items 1 and 12 (knee flexion) require a thigh-shank angle that cannot be resolved from the sensor placements available, consistent with their low ZeroR improvement (+3.8 %, +1.1 %). Additionally, observing the subsets of data used in each best-performing model (Appendix C), stance width (item 8) used ankle data alone and showed the largest improvement over ZeroR (+22.5 %), consistent with this construct being measurable directly at the instrumented segment. However, joint displacement (item 16) also used ankle data alone and showed a large improvement (+16.2 %), but as this item reflects a subjective judgement, meaning this alignment is not clearly explained by sensor placement or item definition. Foot position (item 9) also improved noticeably (+15.7 %), but similarly this model was trained only using the pelvis data and not the data provided from the ankle sensors in closer proximity.

The subsets used by the best performing models did not always leverage features relevant to the specific joint movement targeted being evaluated in a scoring item, such as the best performing model for symmetric foot contact (item 11) using the raw movement data (Appendix C), bypassing the temporal features dedicated to identifying initial contact from the left and right ankle sensors respectively. This finding raises concerns for real-world applications, as a model that overlooks relevant features is unlikely to generalise effectively to the variability inherent in real-world DLJL performance. Future work should leverage feature importance analysis to evaluate the most influential features in each model, to better validate real-world applicability and generalisation.

Additionally, a jump-level cross-validation split was chosen to maximise the training data available, given the modest sample size. As jumps from the same participant could appear in both training and test sets, these results may overestimate generalisation performance relative to a participant-level split. The participant group also involved healthy young adults, meaning generalisation to other populations, including those of different age groups or with previous injuries, remains an important consideration for future work.

This study demonstrates that conventional machine learning models trained on IMU-derived data from the pelvis and tibias provide only limited improvements over a ZeroR approach for automated LESS scoring, indicating that such an approach is not yet robust enough for real-world deployment. However, performance metrics in some models indicated modest recognition of movement error patterns in the IMU-derived features, highlighting potential pathways for partial automation of the LESS. These findings have implications for how LESS-guided interventions are delivered in practice. With manual scoring, assessment is typically restricted to limited testing windows, such as pre-season screening or scheduled rehabilitation check-ins. An automated, field-deployable scoring approach would allow movement quality to be assessed more frequently and at a lower cost, supporting real-time feedback within a training cycle rather than isolated pre/post assessments, particularly relevant for monitoring whether targeted training interventions are producing the intended changes as they occur.

Future work should consider the influence of sensor placement for specific scoring items, expanding the dataset for better distribution, more advanced model architecture, and further hyperparameter tuning, in addition to exploring more complex features which may capture aspects of landing dynamics not represented by the current feature set, such as signal shape descriptors, frequency-domain descriptors or quantitative jump performance metrics. Alternatively, separately evaluating model performance for scoring items with balanced distributions may better isolate true classification ability. Collectively, these steps may improve classification performance and move IMU-based LESS automation closer to reliable use in field and training environments.

CRediT authorship contribution statement

Zane Hamilton: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Jessica Turner:** Writing – review & editing, Supervision, Conceptualization. **Jemma König:** Writing – review & editing, Supervision, Conceptualization. **Kim Hébert-Losier:** Writing – review & editing, Supervision, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Appendix A: Landing Error Scoring System Items, Adapted From Padua et al. (2009).

	Scoring Item	Definition of Error
1	Knee flexion at IC	Knee flexion less than 30°
2	Hip flexion at IC	Thigh is in line with the trunk (hips not flexed)
3	Trunk flexion at IC	Trunk is vertical or extended at the hips (i.e., not flexed)
4	Ankle plantar flexion at IC	Heel-to-toe or flat foot landing at IC
5	Knee valgus at IC	Centre of the patella is medial to the midfoot at IC
6	Lateral trunk flexion at IC	Midline of the trunk is flexed to the left or the right side of the body at IC
7	Stance width (wide)	Feet are positioned greater than shoulder width apart at IC
8	Stance width (narrow)	Feet are positioned less than shoulder width apart at IC
9	Foot position (toe-in)	Foot is externally rotated more than 30° between IC and MKF
10	Foot position (toe-out)	Foot is internally rotated more than 30° between IC and MKF
11	Symmetric footcontact at IC	One foot lands before the other foot or one foot lands heel to toe and the other foot lands toe to heel
12	Knee flexiondisplacement	Knee flexes less than 45° between IC and MKF
13	Hip flexion at MKF	Knee flexes less than 45° between IC and MKF
14	Trunk flexion at MKF	Trunk does not flex more between IC and MKF
15	Knee valgusdisplacement	At the point of maximum medial knee position,the centre of the patella is medial to themidfoot
16	Joint displacement	Soft (0), Average (1), Stiff (2)
17	Overall impression	Excellent (0), Average (1), Poor (2)

** Note: IC: Initial Contact; MKF: Maximum Knee Flexion*

Appendix B:. Participant characteristics

Characteristics	Men(n = 21)	Women(n = 17)	Other(n = 2)	All(n = 40)
Age (y)	20.0 ± 3.9	20.8 ± 4.1	18.4 ± 0.03	20.3 ± 3.9
Height (cm)	183.1 ± 5.8	168.6 ± 6.0	166.5 ± 0.7	176.1 ± 9.3
Mass (kg)	81.6 ± 12.7	70.0 ± 10.1	64.1 ± 2.1	75.8 ± 12.8
Previous Injury	14	10	2	26

Appendix C:. Processing subset used by the best performing models

LESS Scoring Item	Best Performing Model	Processing Subset		
		Data	Segment	Sensors
Total Score	GBC	Raw	S1A2	Ankles
1	DFF	Raw	S2	All
4	GNB	Stats	S1A2	All
5	RFC	Stats	S2	All
6	RFC	Both	S2	All
7	RFC	Stats	S1A2	All
8	KNN	Both	S1A2	Ankles
9	GNB	Stats	S2	Pelvis
11	SGD	Raw	S1A2	All
12	SGD	Stats	S2	Pelvis
14	RFC	Stats	S2	Pelvis
15	RFC	Stats	S2	All
16	GBC	Raw	S2	Ankles
17	GBC	Stats	S2	Pelvis

Data availability

Data will be made available on request.

References

Cerfoglio, S., Galli, M., Tarabini, M., Bertozzi, F., Sforza, C., Zago, M., 2021. Machine learning-based estimation of ground reaction forces and knee joint kinetics from inertial sensors while performing a vertical drop jump. Sensors 21 (22), 7709.

Chaaban, C.R., Berry, N.T., Armitano-Lago, C., Kiefer, A.W., Mazzoleni, M.J., Padua, D. A., 2021. Combining inertial sensors and machine learning to predict vGRF and knee biomechanics during a double-limb jump-landing task. Sensors 21 (13), 4383. https://doi.org/10.3390/s21134383.

Chiang, Y.-Y., Shih, W.-Y., Chen, W.-H., Huang, J.-L., Shiang, T.-Y., 2020. In: A Machine Learning-based Countermovement Performance Measurement Method Using a Wearable IMU. IEEE, pp. 79–85.

Gómez, D., Rojas, A., 2016. An empirical overview of the no free lunch theorem and its effect on real-world machine learning classification. Neural Comput. 28 (1), 216–228.

Hanzlíková, I., Athens, J., Hébert-Losier, K., 2021. Factors influencing the Landing Error Scoring System: Systematic review with meta-analysis. J. Sci. Med. Sport 24 (3), 269–280.

Hébert-Losier, K., Hanzlíková, I., Zheng, C., Streeter, L., Mayo, M., 2020. The 'DEEP' landing error scoring system. Appl. Sci. 10 (3), 892.

Hébert-Losier, K., Knighton, H., Finlayson, S.J., Dubois, B., Esculier, J.-F., Beaven, C.M., 2024. Biomechanics and subjective measures of recreational male runners in three shoes running outdoors: a randomised crossover study. Footwear Sci. 16 (1), 13–23. https://doi.org/10.3390/app10030892.

Hendry, D., Leadbetter, R., McKee, K., Hopper, L., Wild, C., O'sullivan, P., Straker, L., Campbell, A., 2020. An exploration of machine-learning estimation of ground reaction force from wearable sensor data. Sensors 20 (3), 740.

Mascia, G., De Lazzari, B., Camomilla, V., 2023. Machine learning-aided jump height estimate democratization through smartphone measures. Front. Sports Active Living 5, 1112739. https://doi.org/10.3389/fspor.2023.1112739.

Mohammadi, H., Khosravani, M., 2024. The effect of SportsMetrics training on landing technique in female volleyball players with dynamic knee valgus. The Scientific Journal of Rehabilitation Medicine 13 (3), 626–639.

Moses, B., Orchard, J., Orchard, J., 2012. Systematic review: annual incidence of ACL injury and surgery in various populations. Res. Sports Med. 20 (3–4), 157–179.

Naidu, G., Zuva, T., Sibanda, E.M., 2023. A review of evaluation metrics in machine learning algorithms. Computer Science on-Line Conference 15–25.

Padua, D.A., DiStefano, L.J., Beutler, A.I., De La Motte, S.J., DiStefano, M.J., Marshall, S. W., 2015. The landing error scoring system as a screening tool for an anterior cruciate ligament injury–prevention program in elite-youth soccer athletes. J. Athl. Train. 50 (6), 589–595.

Padua, D.A., Marshall, S.W., Boling, M.C., Thigpen, C.A., Garrett Jr, W.E., Beutler, A.I., 2009. The Landing Error Scoring System (LESS) is a valid and reliable clinical assessment tool of jump-landing biomechanics: the JUMP-ACL study. Am. J. Sports Med. 37 (10), 1996–2002.

Pfile, K.R., Gribble, P.A., Buskirk, G.E., Meserth, S.M., Pietrosimone, B.G., 2016. Sustained improvements in dynamic balance and landing mechanics after a 6-week neuromuscular training program in college women's basketball players. J. Sport Rehabil. 25 (3), 233–240.

Pryor, J.L., Root, H.J., Vandermark, L.W., Pryor, R.R., Martinez, J.C., Trojian, T.H., DiStefano, L.J., 2017. Coach-led preventive training program in youth soccer players improves movement technique. J. Sci. Med. Sport 20 (9), 861–866.

Saki, F., Mohammadi, H., Shakiba, E., Ramezani, F., 2021. Does SportsMetrics soccer training improve LESS and dynamic balance in soccer players? a randomized controlled trial. Physical Treatments-Specific Physical Therapy Journal 11 (4), 269–278.

Turner, J.A., Reiche, E.T., Hartshorne, M.T., Lee, C.C., Blodgett, J.M., Padua, D.A., 2025. Open source, open science: Development of the open landing error scoring system as the automated landing error scoring system. J. Athl. Train. 60 (8), 590–599.

Wang, X., Yu, H., Kold, S., Rahbek, O., Bai, S., 2023. Wearable sensors for activity monitoring and motion control: a review. Biomimetic Intell. Rob. 3 (1), 100089.

Woltmann, L., Hartmann, C., Lehner, W., Rausch, P., Ferger, K., 2023. Sensor-based jump detection and classification with machine learning in trampoline gymnastics. German Journal of Exercise and Sport Research 53 (2), 187–195.