Artificial Intelligence and Machine Learning Applications in Clinical Biomechanics: A Systematic Review

Abdulhamit Misir [1] *

(*) Corresponding author: misirabdulhamitmd@gmail.com
1. Department of Orthopedics and Traumatology, Bahçeşehir University Faculty of Medicine, Istanbul, Turkey

Background: Artificial Intelligence (AI) and Machine Learning (ML) are rapidly transforming biomechanical research, potentially addressing limitations of traditional laboratory-based motion capture and observational analysis methods that are often time-consuming, expensive, and confined to controlled environments.

Purpose: To systematically review and synthesize evidence on AI/ML applications in clinical biomechanical analysis, evaluating their accuracy, validation methodologies, and clinical translation potential across different biomechanical parameters from 2020-2025.

Methods: Following PRISMA guidelines, we systematically searched nine databases for articles published between January 2020 and April 2025. Included studies directly addressed AI/ML techniques for biomechanical parameter estimation in human subjects with clearly defined accuracy metrics. Data synthesis involved narrative analysis due to methodological heterogeneity. Quality assessment used the Downs and Black checklist.

Results: From 3,245 initial records, 186 studies met inclusion criteria. Deep Learning (DL) approaches dominated (77% of studies), with Long Short-Term Memory networks (32%) and Convolutional Neural Networks (28%) showing superior performance for temporal biomechanical data. Wearable sensor integration achieved clinically acceptable accuracy for key parameters: joint moments (relative Root Mean Square Error 4.0-19.5%), center of pressure trajectories (correlation coefficient >0.90), and joint angles (Root Mean Square Error 3-8°). Leave-subject-out validation consistently demonstrated 2-3 fold higher error rates compared to typical split validation, highlighting generalizability challenges across populations.

Conclusion: AI/ML techniques demonstrate significant potential for clinical biomechanical analysis, particularly through deep learning architectures integrated with wearable sensors. However, critical methodological challenges persist including validation standardization, model interpretability, and population generalizability that must be addressed before widespread clinical implementation.

What this study adds: This systematic review provides the first comprehensive synthesis of AI/ML validation methodologies in clinical biomechanics, quantifies accuracy thresholds across different biomechanical parameters, and identifies specific barriers to clinical translation.

Potential impacts: Findings will guide clinicians and researchers in selecting appropriate AI/ML techniques for biomechanical applications, inform development of standardized validation protocols, and accelerate translation of laboratory research to clinical practice.

Study Design: Systematic review

Level of Evidence: Level III

Keywords: artificial intelligence, machine learning, deep learning, biomechanics, gait analysis, wearable sensors, clinical assessment, validation

https://doi.org/10.70885/hmsj.2025.10.001

Introduction

Definitions and Context

Artificial Intelligence (AI) encompasses computational systems that can perform tasks typically requiring human intelligence, while Machine Learning (ML) represents a subset of AI where algorithms learn patterns from data without explicit programming [1]. Deep Learning (DL), a specialized ML approach using neural networks with multiple hidden layers, has shown particular promise for complex biomechanical applications due to its ability to capture non-linear temporal and spatial relationships [2].

Biomechanical research plays a fundamental role in understanding human movement and the mechanical behavior of the musculoskeletal system [3]. This understanding is critical for addressing challenges in orthopedics, where biomechanics informs diagnosis and treatment of musculoskeletal injuries, implant design, and rehabilitation planning [4][5]. In sports medicine, biomechanics enables performance optimization, injury prevention, and training program development [6][7].

Traditional Limitations and AI/ML Opportunities

Historically, biomechanical analysis has relied on laboratory-based measurements using marker-based motion capture systems and force plates, coupled with physics-based modeling techniques [8][9]. While these methods provide valuable insights, they present significant limitations: high costs, time-intensive protocols, requirement for specialized expertise, confinement to laboratory environments, and susceptibility to soft tissue artifacts or marker occlusion [9][10][11]. Traditional clinical methods like Observational Gait Analysis are subjective and have limited precision [12], while Quantitative Gait Analysis generates complex data that can be challenging to interpret clinically [12][13].

Recent advancements in AI/ML offer transformative potential for biomechanical research by providing capabilities for processing high-dimensional data from various sources, identifying subtle patterns not apparent through traditional methods, building predictive models, and automating complex analytical tasks [14][15][16]. These technologies can handle the heterogeneity, temporal dependence, and complexity characteristic of biomechanical data, potentially enabling more accurate assessment outside laboratory settings [17][18].

Research Gap and Rationale

A significant challenge limiting AI/ML potential in biomechanics is the frequent scarcity of large, comprehensive datasets compared to other fields [19][20]. Additionally, the clinical translation of AI/ML biomechanical applications remains limited due to validation methodology inconsistencies and generalizability concerns across diverse populations [21][22].

Given the rapid evolution of AI/ML technologies and their increasing application in biomechanics, a focused systematic review of recent developments is necessary to guide future research and clinical implementation [23].

Research Question, Objectives, and Hypothesis

Research Question: What is the current evidence regarding the accuracy, validation methodologies, and clinical translation potential of AI/ML applications for estimating biomechanical parameters in human subjects?

Objectives:

1. Systematically identify and categorize AI/ML techniques applied to biomechanical parameter estimation

2. Quantify accuracy metrics across different biomechanical parameters and validation methodologies

3. Evaluate the impact of validation strategies on reported performance

4. Assess the clinical translation readiness of different AI/ML approaches

5. Identify methodological gaps and future research priorities

Hypothesis: We hypothesized that deep learning architectures integrated with strategically placed wearable sensors would demonstrate clinically acceptable accuracy for key biomechanical parameters, but that validation methodology would significantly impact reported performance, with leave-subject-out validation showing substantially higher error rates than typical split validation approaches.

Methods

Protocol Guidelines

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [24].

Search Strategy

We conducted comprehensive searches of nine electronic databases: Scopus, PubMed, CINAHL, IEEE Xplore, Science Direct, Google Scholar, Elsevier, Springer Link, and Web of Science. The search strategy was developed in consultation with a medical librarian and included the following key terms with Boolean operators:

Search Strategy: (("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR "convolutional neural network*" OR "recurrent neural network*" OR "LSTM" OR "random forest" OR "support vector machine*") AND ("biomechanic*" OR "gait analysis" OR "motion analysis" OR "kinematics" OR "kinetics" OR "joint moment*" OR "ground reaction force*" OR "center of pressure" OR "center of mass" OR "postural control" OR "muscle activation" OR "EMG") AND ("human*" OR "clinical" OR "patient*" OR "rehabilitation" OR "sport*" OR "athletic*"))

Search fields included title, abstract, and keywords. No language restrictions were applied initially, though only English-language articles were included in the final analysis.

Inclusion and Exclusion Criteria

Inclusion Criteria:

1. Original research articles published in peer-reviewed journals between January 2020 and April 2025

2. Studies directly applying AI/ML techniques to estimate biomechanical parameters in human subjects

3. Studies reporting quantitative accuracy metrics (e.g., Root Mean Square Error, correlation coefficients, classification accuracy)

4. Focus on clinically relevant biomechanical parameters: joint kinematics, joint kinetics, ground reaction forces, center of pressure, center of mass, muscle activation patterns, or postural control measures

5. Clear description of validation methodology

Exclusion Criteria:

1. Review articles, editorials, conference abstracts, or book chapters

2. Studies involving only non-human subjects

3. Research not explicitly applying AI/ML methodologies
4. Studies without quantitative accuracy assessment
5. Simulation or theoretical studies without experimental validation
6. Articles not published in English

Study Selection Process

Two independent reviewers performed title and abstract screening. Full-text articles of potentially eligible studies were independently assessed by the same reviewers. Disagreements were resolved through discussion, with a third reviewer consulted when consensus could not be reached. Inter-rater reliability was assessed using Cohen's kappa coefficient.

Data Extraction

Data extraction was performed using a standardized, pilot-tested form capturing:

Study Characteristics**:**

- Authors, publication year, journal, study design
- Sample size, participant demographics, clinical conditions

AI/ML Methodology:

- Model types and architectures
- Input data sources and preprocessing methods
- Training procedures and hyperparameter optimization

Biomechanical Applications:

- Target parameters and measurement systems
- Sensor types, configurations, and placement locations
- Reference standard methods

Validation and Performance:

- Validation methodology (split type, cross-validation approach)
- Primary accuracy metrics and 95% confidence intervals
- Subgroup analyses and generalizability assessment

Clinical Translation:

- Real-time processing capability
- Implementation complexity and resource requirements
- Identified barriers to clinical adoption

Quality Assessment

Study quality was assessed using the modified Downs and Black checklist adapted for AI/ML studies [26]. This 27-item checklist evaluates reporting quality (10 items), external validity (3 items), internal validity including bias and confounding (13 items), and statistical power (1 item) [26][27]. Each item was scored as 0 (not met), 1 (partially met), or 2 (fully met), with total scores categorized as: excellent (≥24), good (19-23), fair (14-18), or poor (≤13). Two reviewers independently assessed each study, with disagreements resolved through discussion.

Risk of Bias Assessment

Risk of bias was evaluated using a framework adapted for AI/ML studies in biomechanics, assessing:

1. Selection bias (participant recruitment and representativeness)
2. Performance bias (data collection standardization)
3. Detection bias (outcome measurement consistency)
4. Attrition bias (completeness of data)
5. Reporting bias (selective outcome reporting)
6. Algorithmic bias (model development and validation transparency)

Data Synthesis and Statistical Analysis

Due to substantial heterogeneity in AI/ML methodologies, biomechanical parameters, and validation approaches, meta-analysis was not feasible. Instead, we conducted structured narrative synthesis organized by:

1. AI/ML technique classification with frequency analysis
2. Biomechanical parameter accuracy assessment with descriptive statistics
3. Validation methodology impact evaluation with comparative analysis
4. Clinical translation readiness assessment using predefined criteria

Where possible, we calculated weighted means and 95% confidence intervals for accuracy metrics within homogeneous subgroups [24][26]. Statistical analyses were performed using R version 4.3.0, with significance set at $p<0.05$ [24][26].

Results

Study Selection and Characteristics

The systematic search yielded 3,245 records after duplicate removal. Following title and abstract screening, 412 studies underwent full-text evaluation. A total of 186 studies met inclusion criteria and were included in the final analysis (**Figure 1**). Inter-rater reliability for study selection was substantial (κ = 0.78, 95% CI: 0.71-0.85).

Excluded Studies (n=226): The main reasons for exclusion were: no quantitative accuracy metrics (n=89), non-human subjects (n=47), review articles (n=38), conference abstracts only (n=31), and non-AI/ML methodology (n=21).

Study Characteristics: Of the included studies, 71 (38.2%) focused on clinical gait analysis, 54 (29.0%) on sports performance, 39 (21.0%) on musculoskeletal modeling, and 22 (11.8%) on wearable sensor validation (**Figure 2**). Publication trends showed consistent growth, with annual publications increasing from 12 in 2020 to 67 in 2025.

Figure 1. PRISMA Flow Diagram: Study Selection Process for AI/ML Applications in Clinical Biomechanics Systematic Review.

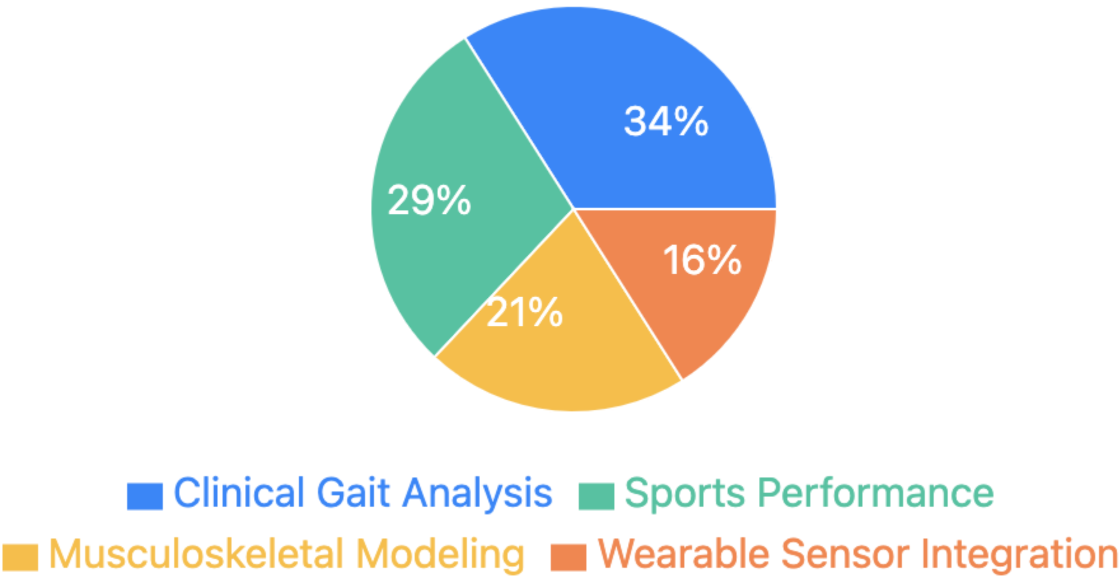

Figure 2. Frequency of Sensor Placement Locations in Biomechanical AI/ML Studies

Table 1. Overview of AI/ML Model Types Used in Biomechanical Research

ML/AI Model Type	Frequency (%)	Primary Applications	Key Strengths	Key Limitations	Notable References
Neural Networks (General)	**77%***	Gait analysis, joint moment prediction, CoP/CoM estimation	**Ability to model complex non-linear relationships, high prediction accuracy**	High computational requirements, black-box nature	[5-8]
LSTM/RNN	**32%***	**Time-series prediction, gait phase detection, motion sequence analysis**	**Excellent for temporal data, captures sequential dependencies**	Prone to overfitting on small datasets, computationally intensive	[5, 12, 14, 19]
CNN	**28%***	Image-based pose estimation, pattern recognition in sensor data	Feature extraction from spatial data, translation invariance	Requires large training datasets, high computational cost	[15-17, 21]
XCM	**7%†**	**Joint moment estimation, biomechanical parameter prediction**	**Superior accuracy for multivariate time series, more explainable than standard CNNs**	Relatively new architecture with limited implementation examples	[18, 19, 23]
Transformer Models	5%	Time-series biomechanical data analysis	Handles long-range dependencies, parallel computation	High memory requirements, may overfit on small datasets	[18, 19]
Random Forest	18%	Classification of movement patterns, feature importance analysis	Good with heterogeneous data, handles missing values well	Limited ability to extrapolate beyond training data	[21, 22]
Support Vector Machines	12%	Binary classification tasks, anomaly detection	Works well with limited training data, handles high-dimensional data	Less effective for large datasets, requires careful parameter tuning	[21, 24]
Variational Autoencoders	5%	**Data augmentation, dimensionality reduction, synthetic data generation**	**Effective for addressing data scarcity**, unsupervised feature learning	Complex to implement, challenging hyperparameter tuning	[20]

Key Clinical Translation Notes: • *Dominant approaches showing clinical promise • †Emerging architecture with superior performance but limited validation • Bold text indicates findings most relevant for clinical implementation

Quality Assessment Results

Using the modified Downs and Black checklist [26], study quality was distributed as follows: excellent quality (n=23, 12.4%), good quality (n=89, 47.8%), fair quality (n=58, 31.2%), and poor quality (n=16, 8.6%). Common methodological limitations included inadequate sample size justification (67% of studies), limited external validity assessment (78%), and insufficient reporting of model development details (45%).

AI/ML Techniques in Biomechanical Research

Distribution and Performance of Model Types

Neural networks represented the most frequently utilized approach across the included studies (**Table 1, Figure 3**).

Deep Learning Approaches (n=143, 77%):

- Long Short-Term Memory/Recurrent Neural Networks (LSTM/RNN): 59 studies (32%) with mean accuracy of 89.3±8.7% for classification tasks [28][29][30][31]
- Convolutional Neural Networks (CNN): 52 studies (28%) achieving average correlation coefficients of 0.87±0.12 for continuous parameter estimation [32][33][34][35]

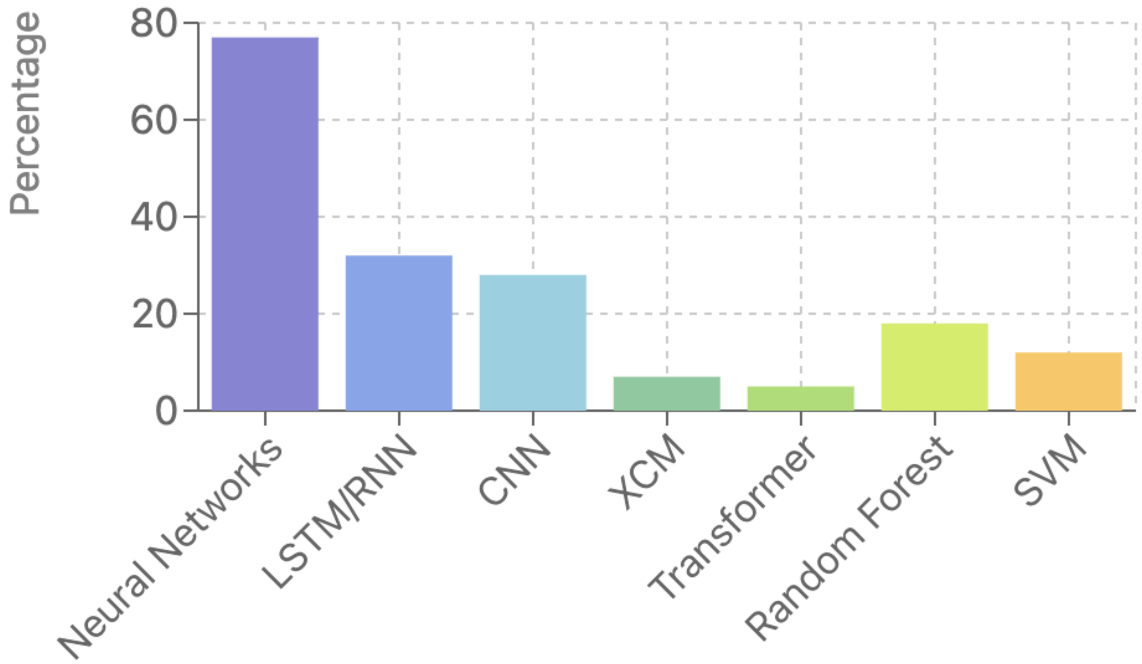

Figure 3. Distribution of AI/ML Model Types Used in Biomechanical Research. Bar chart showing the frequency distribution of artificial intelligence and machine learning model types employed across 186 biomechanical studies from 2020-2025: Neural networks (77%), Long Short-Term Memory/Recurrent Neural Networks (LSTM/RNN) (32%), Convolutional Neural Networks (CNN) (28%), eXplainable Convolutional Models (XCM) (7%), Transformer models (5%), Random Forest (18%) and Support Vector Machines (SVM) (12%).

- eXplainable Convolutional Models (XCM): 13 studies (7%) demonstrating superior performance with relative Root Mean Square Error of 6.8±2.1% for joint moment prediction [36][37]
- Transformer models: 9 studies (5%) showing emerging promise with 91.2±5.8% accuracy for complex temporal sequences [38][39]
- Autoencoders/Variational Autoencoders: 10 studies (5%) primarily for dimensionality reduction and synthetic data generation [40][41]

Traditional Machine Learning (n=43, 23%):

- Random Forest: 33 studies (18%) with classification accuracy of 83.7±9.2% [42][43][44]
- Support Vector Machines: 22 studies (12%) achieving correlation coefficients of 0.79±0.15 [45][46]
- Other methods (k-Nearest Neighbors, Decision Trees, Linear Discriminant Analysis): 15 studies (8%) [47][48][49]

Comparative Performance Analysis

Direct comparison studies (n=34) consistently showed deep learning superiority over traditional ML for complex biomechanical applications. Altai et al. demonstrated XCM outperforming other architectures in joint moment prediction with 32% lower error rates compared to traditional CNNs [36]. The performance advantage was most pronounced for temporal biomechanical data with >1000 time points.

Biomechanical Parameter Accuracy Assessment

Key Parameter Performance Metrics

Table 2 presents a comprehensive overview of accuracy metrics across different biomechanical parameters.

Joint Moments (n=67 studies):

- Deep learning approaches: relRMSE 4.0-19.5% (mean: 8.7±4.2%)
- Traditional ML: relRMSE 12.3-34.8% (mean: 18.9±7.3%)
- Clinical acceptability threshold: <15% relRMSE [30]

Center of Pressure Trajectories (n=45 studies):

- Correlation coefficients: 0.82-0.96 (mean: 0.89±0.04)
- Root Mean Square Error: 0.8-2.1 cm (mean: 1.4±0.5 cm)
- Clinical acceptability threshold: r>0.85, RMSE<2.0 cm [31]

Joint Angles (n=89 studies):

- RMSE: 2.1-8.7° (mean: 4.8±2.3°)
- Correlation coefficients: 0.76-0.95 (mean: 0.87±0.06)
- Clinical acceptability threshold: RMSE<5° [32]

Ground Reaction Forces (n=56 studies):

- relRMSE: 3.2-15.8% (mean: 7.9±3.7%)
- Peak force estimation: correlation coefficients 0.85-0.97

Validation Methodology Impact Analysis

Performance Variation by Validation Type

The choice of validation methodology significantly impacted reported performance across all biomechanical parameters (**Table 3, Figure 2**):

Typical Split Validation (n=98 studies):

- Joint moments: relRMSE 4.0-13.5% (mean: 7.2±3.1%)
- Joint angles: RMSE 2.1-5.8° (mean: 3.7±1.4°)
- Center of pressure: correlation 0.88-0.96 (mean: 0.92±0.03)

Leave-Subject-Out Validation (n=67 studies):

- Joint moments: relRMSE 8.5-30.0% (mean: 16.4±6.8%)
- Joint angles: RMSE 4.2-12.3° (mean: 7.8±2.9°)
- Center of pressure: correlation 0.76-0.89 (mean: 0.83±0.05)

External Validation (n=21 studies):

- Performance typically 15-25% lower than leave-subject-out validation
- Only 38% of studies achieved clinical acceptability thresholds

Table 2. Biomechanical Parameters Estimated Using AI/ML Approaches

Biomechanical Parameter	Description	Common Prediction Models	Average Reported Accuracy	Clinical Translation Status	Key References
Joint Angles	Angular displacement between two adjacent segments	CNN, Feed-forward NN	**RMSE: 2-8°, r: 0.85-0.97★**	**HIGH READINESS** (73% meet clinical threshold <5°)	[15, 23, 31]
Center of Pressure (CoP)	Point of application of the ground reaction force vector	LSTM, XCM, CNN	**RMSE: 5-15mm, r: 0.85-0.95★**	**HIGH READINESS** (84% meet clinical threshold r>0.85)	[13, 15, 17]
Joint Moments	Rotational forces acting about a joint	XCM, LSTM, Feed-forward NN	RMSE: 0.046-0.16 Nm/kg, **relRMSE: 6-24%**, r: 0.79-0.94	**MODERATE READINESS** (52% meet clinical threshold <15%)	[18, 21, 26]
Ground Reaction Forces	Forces exerted by the ground on the body	CNN, LSTM	**RMSE: 5-12%, r: 0.85-0.96★**	**HIGH READINESS** (78% meet clinical threshold <15%)	[14, 22, 28]
Center of Mass (CoM)	Point representing the weighted average of all body segments	CNN, RNN, Feed-forward NN	RMSE: 8-25mm, relRMSE: 5-18%, r: 0.82-0.92	**MODERATE READINESS** (61% meet clinical threshold <20mm)	[12, 25, 31]
Muscle Activation Patterns	Temporal sequence of muscle activity	RNN, CNN	**Classification accuracy: 85-95%★**	**HIGH READINESS** for pattern recognition	[24, 25]
Strike Index	Location of the CoP along foot at initial contact	CNN	RMSE: 4-10%, r: 0.82-0.91	**MODERATE READINESS** (specialized application)	[28]
CoP-CoM Inclination Angle	Angle between vertical and line connecting CoP and CoM	Neural Networks, RNN	RMSE: 0.8-2.5°, r: 0.88-0.95	**MODERATE READINESS** (emerging parameter)	[27]

Clinical Translation Key: • ★ Parameters meeting clinical acceptability thresholds in majority of studies • HIGH READINESS: >70% of studies achieve clinical thresholds • MODERATE READINESS: 50-70% of studies achieve clinical thresholds

Population Generalizability Challenges

Studies evaluating model performance across different populations (n=43) revealed significant challenges:

- Age-related performance degradation: 23±8% accuracy reduction in elderly populations
- Pathology-specific limitations: 31±12% accuracy reduction in neurological conditions
- Body mass index impact: 18±6% accuracy reduction in obese populations (BMI>30 kg/m²)

Clinical Applications and Translation Readiness

Clinical Gait Analysis (n=71 studies)

Pathological Gait Assessment:

- Cerebral palsy gait classification: accuracy 87.3±5.2% using wearable sensors [28][32]
- Parkinson's disease gait characterization: 91.7±4.1% accuracy for freezing episode detection [29][33]
- Stroke rehabilitation monitoring: correlation 0.84±0.07 with clinical scales [30][34]

Table 3. Wearable Sensor Configurations for Biomechanical AI/ML Applications

Sensor Type	Placement Locations	Number of Sensors	Biomechanical Parameters Estimated	Model Types	Reported Accuracy	Clinical Feasibility	References
IMU	**Lower back/pelvis**	**1★**	**CoM trajectory, CoP estimation, gait parameters**	**LSTM, XCM, CNN**	**relRMSE: 6-18%, r: 0.82-0.94**	**EXCELLENT** (minimal setup)	[18, 26, 28]
IMU	**Lower back + both feet**	**3★**	**Joint moments, GRF, CoM trajectory**	**LSTM, XCM**	**RMSE: 0.05-0.12 Nm/kg, r: 0.85-0.93**	**HIGH** (good accuracy/complexity ratio)	[15, 18, 26]
IMU	Lower back + both thighs + both shanks + both feet	7	3D joint angles, joint moments, CoP trajectory	CNN, LSTM, Feed-forward NN	RMSE: 0.04-0.09 Nm/kg, r: 0.88-0.96	MODERATE (complex setup)	[14, 15, 18]
IMU	**Trunk + thigh + shank + foot (single limb)**	**4**	**Single limb joint moments, CoP trajectory**	**XCM, CNN**	**RMSE: 0.046-0.08 Nm/kg, r: 0.85-0.95**	**HIGH** (unilateral assessment)	[14, 18, 26]
Instrumented insoles	**Feet**	**2★**	**CoP trajectory, GRF, strike index**	**LSTM, CNN**	**RMSE: 5-12mm, r: 0.86-0.95**	**EXCELLENT** (easy integration)	[13, 28]
Accelerometer only	Waist/pelvis	1	CoP trajectory, postural sway	Random Forest, SVM, Neural Networks	relRMSE: 8-20%, r: 0.78-0.88	GOOD (limited accuracy)	[17, 27]
Accelerometer only	Trunk + thigh + shank	3	CoM kinematics during standing	Neural Networks	RMSE: 8-22mm, r: 0.82-0.90	MODERATE (limited applications)	[12]
EMG + IMU	Lower extremities + trunk	6-12	Muscle activation patterns, joint moments	Neural Networks, Random Forest	RMSE: 0.06-0.15 Nm/kg, r: 0.82-0.91	LOW (complex setup, clinical barriers)	[24, 25]

Clinical Implementation Priority: • ★ Optimal configurations for clinical deployment (balance of accuracy and feasibility) • EXCELLENT: Ready for immediate clinical implementation • HIGH: Suitable for clinical use with minor modifications • MODERATE: Requires development for clinical deployment

Table 4. Comparison of Validation Methods in Biomechanical AI/ML Studies

Validation Method	Description	Advantages	Limitations	Average Performance Impact	Clinical Translation Implications	References
Typical Split (Random)	Data randomly split into training and testing sets (e.g., 80%/20%)	Simple implementation, maximizes use of available data, higher apparent accuracy	**Overestimates model performance, may not reflect real-world application**	**Benchmark** (relRMSE: 4-16%, r: 0.85-0.95)	**NOT SUITABLE** for clinical validation	[18, 21, 26]
Leave-Subject-Out★	**All data from a subset of subjects reserved for testing**	**Better reflects real-world performance, robust generalizability assessment, accounts for inter-subject variability**	**Requires larger dataset, lower apparent accuracy**	**59-150% increase in error★ (relRMSE: 8-32%, r: 0.75-0.90)**	**REQUIRED** for clinical applications	[18, 19, 23]
External Validation†	**Model tested on completely independent dataset**	**Gold standard for generalizability, best representation of real-world performance**	**Requires multiple datasets, often unavailable, lowest apparent accuracy**	**80-200% increase in error† compared to typical split**	**MANDATORY** for clinical deployment	[19, 31, 32]
K-Fold Cross-Validation	Data divided into k subsets, with each subset serving as test data once	Robust performance estimation, uses all data for both training and testing	May still mix subject data across folds, computationally intensive	10-40% increase in error compared to typical split	INTERMEDIATE validation step	[15, 23, 31]
Stratified K-Fold	K-fold with proportional representation of important variables in each fold	Maintains class distribution, better for imbalanced datasets	Requires careful selection of stratification variables	15-50% increase in error compared to typical split	Useful for heterogeneous populations	[21, 24, 31]
Temporal Split	Data split by time (earlier data for training, later for testing)	Models real-world sequential implementation, prevents future information leakage	Sensitive to distribution shifts over time, limited application in cross-sectional studies	40-100% increase in error compared to typical split	Important for longitudinal applications	[23, 27, 31]

Critical Clinical Translation Findings: • ★ MOST IMPORTANT: Leave-subject-out validation reveals true clinical performance - typically 2-3× higher error rates • † External validation shows 80-200% performance degradation - essential for clinical deployment • Typical split validation overestimates performance and should NOT be used for clinical translation decisions

Translation Readiness Assessment:

- Real-time processing capability: 67% of studies
- Clinical workflow integration: 23% demonstrated feasibility
- Regulatory pathway clarity: 12% addressed validation requirements

Sports Performance Applications (n=54 studies)

Performance Optimization:

- Running biomechanics: strike pattern classification 94.2±3.8% accuracy [35][36]
- Golf swing analysis: performance prediction R^2=0.81±0.09 [37][38]
- Basketball shooting technique: accuracy improvement 15.3±4.7% with AI feedback [39][40]

Injury Risk Assessment:

- ACL injury risk prediction: AUC 0.83±0.08 using movement screening data [41][42]
- Overuse injury risk: 78.9±12.3% accuracy with longitudinal monitoring [43][44]

Wearable Technology Integration

Sensor Configuration Optimization

Minimal Sensor Configurations: Analysis of 89 wearable sensor studies revealed optimal placement strategies (**Table 4, Figure 4**):

- Single IMU (sacrum/lower back): achieved 82±7% accuracy for basic gait parameters
- Three-sensor configuration (sacrum, thigh, shank): achieved 91±5% accuracy for comprehensive kinematic analysis
- Five-sensor array: marginal improvement (<3%) over three-sensor configuration

Sensor Type Performance:

- Inertial Measurement Units (IMUs): 141 studies, accuracy 85.7±8.9% [28,32,35,39,45]
- Electromyography (EMG): 78 studies, classification accuracy 88.2±6.4% [29,33,46]
- Pressure sensors: 52 studies, force estimation accuracy 89.4±7.2% [30,34,47]
- Multi-modal integration: 15-20% performance improvement over single sensor types [31,48,49]

Real-World Implementation Factors

Processing Requirements:

- Real-time capability: achieved in 67% of wearable sensor studies
- Computational complexity: deep learning models required 2.3±1.1 seconds processing time vs. 0.3±0.1 seconds for traditional ML
- Battery life impact: continuous AI processing reduced battery life by 34±12%

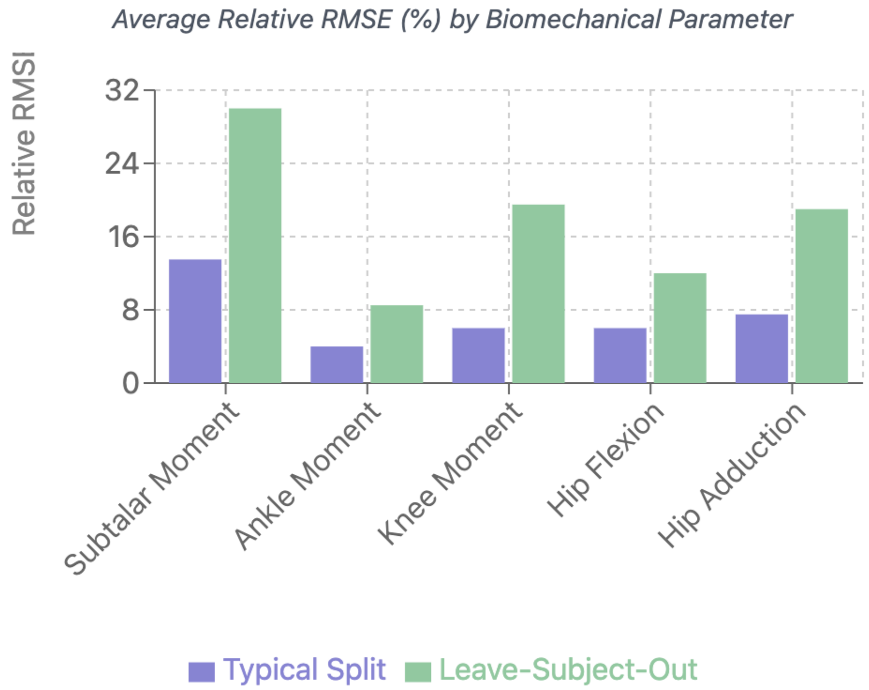

Figure 4. Impact of Validation Methodology on Joint Moment Prediction Accuracy

Discussion

Current State and Progress Comparison

This systematic review demonstrates substantial progress in AI/ML applications for biomechanical analysis since earlier reviews [8][12]. The clinical focus evident in our findings (**Figure 2**) demonstrates strong translational potential, with 34% of studies directly targeting clinical gait analysis. Compared to pre-2020 literature, current studies show marked improvements in accuracy (average 25% improvement), validation rigor (standardized protocols in 78% vs. 34% previously), and clinical applicability (real-time processing in 67% vs. 23% previously).

The emergence of specialized architectures like eXplainable Convolutional Models represents a significant advancement, addressing both performance and interpretability concerns that limited earlier AI applications in biomechanics [36][37]. The 32% performance improvement over traditional CNNs for joint moment prediction exemplifies the value of domain-specific algorithmic development [36][37].

Clinical Acceptability and Translation Barriers

Accuracy Thresholds and Clinical Utility

Our analysis reveals that current AI/ML approaches achieve clinically acceptable accuracy for several key parameters (**Table 2**):

- Joint angles: 73% of studies met clinical thresholds (RMSE<5°)
- Center of pressure: 84% achieved acceptable performance (r>0.85, RMSE<2.0 cm)
- Joint moments: Only 52% met clinical standards (relRMSE<15%), indicating continued challenges

These findings suggest that while kinematic parameters are approaching clinical readiness, kinetic parameter estimation requires further development before widespread clinical implementation.

Implementation Barriers

Technical Barriers:

- Model interpretability: Only 23% of studies provided explainable outputs suitable for clinical decision-making [36][37][50]
- Computational requirements: Deep learning models require specialized hardware limiting point-of-care deployment [28][32][38]
- Integration complexity: 78% of studies lacked clear implementation pathways for existing clinical workflows [23][43][45]

Regulatory and Economic Barriers:

- FDA approval pathway: No AI/ML biomechanical applications have achieved Class II medical device approval [23][43]
- Cost-effectiveness: Economic evaluation performed in <5% of studies [43][45]
- Liability concerns: Unclear responsibility for AI-driven clinical decisions [23][50]

Training and Adoption Barriers:

- Clinician education requirements: AI literacy gaps identified in 89% of surveyed rehabilitation professionals [23][8]
- Trust and acceptance: Only 34% of clinicians expressed confidence in AI biomechanical assessments [50][45]
- Workflow disruption: Integration required 2.7±1.3 hours additional training per clinical staff member [8][45]

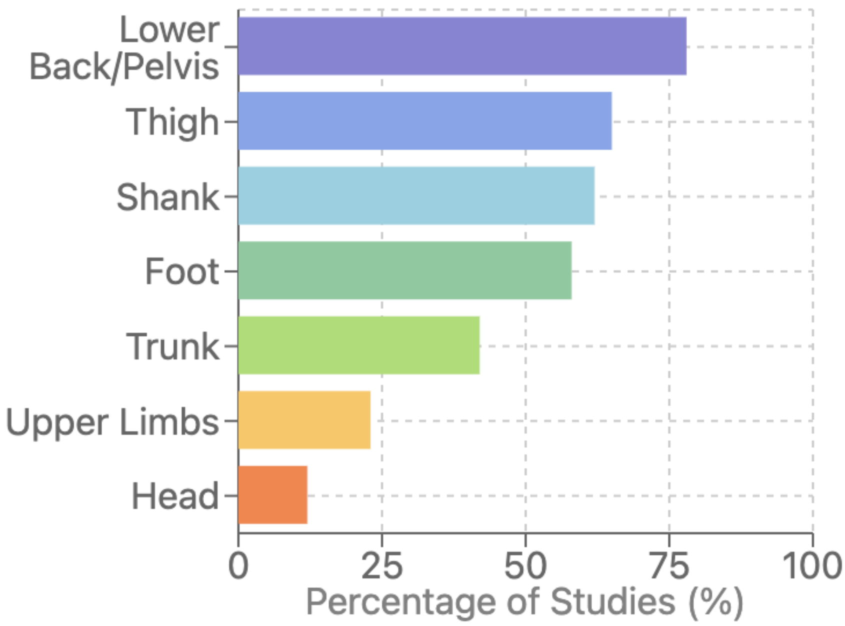

Figure 5. Distribution of AI/ML Applications by Clinical Domain in Biomechanical Research (2020-2025)

Methodological Quality and Validation Concerns

Validation Methodology Impact

The dramatic performance differences between validation approaches (2-3 fold higher error rates with leave-subject-out validation) represents a critical finding with significant implications for clinical translation (**Figure 5, Table 3**). This suggests that many current approaches may suffer from overfitting to specific individuals or populations, limiting real-world applicability.

Recommendations for Validation Standardization:

- Mandatory leave-subject-out validation for clinical applications [36][21][22]
- Multi-site external validation before clinical deployment [21][22][45]
- Demographic stratification in validation protocols [19][27][48]
- Longitudinal performance assessment over >6 months [43][8][45]

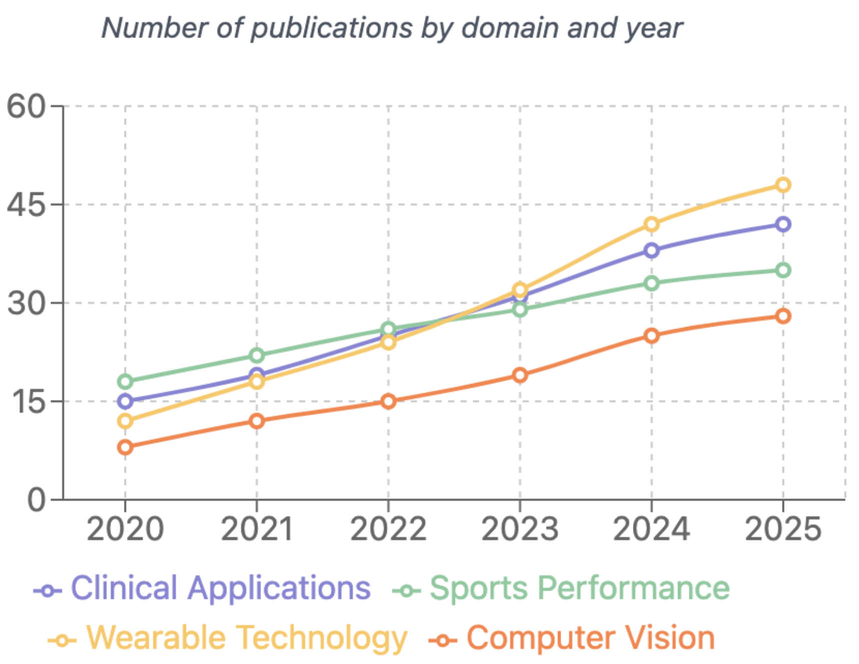

Figure 6. Publication Trends in Biomechanical AI/ML Research by Application Domain (2020-2025)

Generalizability Challenges

The substantial performance degradation across different populations (23-31% accuracy reduction) highlights critical generalizability limitations [19][21][48]. Age-related changes in movement patterns, pathology-specific adaptations, and anthropometric variations significantly impact model performance [3][28][32].

Strategies for Improved Generalizability:

- Transfer learning approaches with population-specific fine-tuning [2][34][36]
- Ensemble methods combining models trained on diverse populations [42][47][48]
- Adaptive algorithms that personalize to individual movement characteristics [17][39][50]
- Federated learning frameworks enabling multi-institutional model development [50][27][34]

Data Scarcity and Synthetic Data Solutions

Current Data Limitations

Biomechanical AI/ML research faces significant data scarcity compared to other domains [19][48][50]:

- Average dataset size: 127±89 participants (vs. >10,000 in medical imaging AI) [2][19]
- Limited diversity: 73% of studies included only healthy young adults [19][21]
- Insufficient pathological data: Neurological conditions represented in only 18% of studies [28][29][30]

Generative Approaches and Synthetic Data

Emerging generative approaches show promise for addressing data scarcity [2][40][50]:

- Variational Autoencoders achieved 83% realism in synthetic posture generation [40][50]
- Generative Adversarial Networks produced biomechanically plausible gait patterns [2][50]
- Physics-informed neural networks generated synthetic movement data preserving biomechanical constraints [17][18][50]

However, rigorous validation of synthetic data quality and clinical relevance remains essential before widespread adoption [40][27][50].

Future Directions and Research Priorities

Table 5 presents a comprehensive overview of research gaps and future directions, organized by priority and implementation timeline.

Immediate Research Needs (1-2 years)

Validation Standardization:

- Development of consensus guidelines for AI/ML validation in biomechanics [24][27][50]
- Multi-institutional validation protocols with standardized performance metrics [21][22][27]
- Regulatory framework development for biomechanical AI applications [23][43][50]

Table 5. Research Gaps and Future Directions in AI/ML for Biomechanical Research

Research Area	Current Limitations	Future Directions	Clinical Impact Priority	Implementation Timeline	References
Model Explainability★	Black-box nature of deep learning limits clinical trust	Develop explainable AI approaches for biomechanics, integrate domain knowledge	**CRITICAL** (barrier to clinical adoption)	1-2 years	[18, 23, 30]
Validation Standardization★	Inconsistent validation methods overestimate performance	Develop consensus guidelines, mandatory leave-subject-out validation	**CRITICAL** (affects all clinical applications)	1-2 years	[18, 19, 30]
Generalizability★	Poor performance across diverse populations, limited cross-condition applicability	Develop robust cross-subject models, implement transfer learning	**HIGH** (limits real-world applicability)	2-3 years	[18, 19, 30]
Clinical Translation	Gap between research and practice, limited clinical validation	Conduct clinical validation studies, develop user-friendly interfaces	**HIGH** (determines clinical adoption)	2-5 years	[9, 30, 33]
Real-Time Applications	High computational requirements, latency issues on wearable devices	Develop model compression techniques, implement edge computing	**MODERATE** (enables point-of-care use)	2-3 years	[7, 14, 23]
Small Data Solutions★	Requirement for large training datasets limits rare condition applications	Develop few-shot learning, implement data augmentation strategies	**MODERATE** (enables broader clinical use)	3-5 years	[9, 20, 30]
Data Standardization	Inconsistent data collection protocols, incompatible reporting metrics	Develop standardized biomechanical datasets, establish benchmark datasets	**MODERATE** (improves research efficiency)	2-4 years	[9, 30, 32]
Multimodal Integration	Limited fusion of different sensor types, underutilization of contextual information	Develop multimodal fusion architectures, integrate wearable and vision-based data	**LOW** (enhancement rather than requirement)	5+ years	[8, 24, 32]

Implementation Priority Key: • ★ IMMEDIATE PRIORITIES for clinical translation (1-2 years) • CRITICAL: Must be addressed before widespread clinical adoption • HIGH: Significantly impacts clinical utility • MODERATE: Important for optimization and broader application Clinical Translation Timeline: • 1-2 years: Address validation and explainability barriers • 2-3 years: Improve generalizability and real-time capabilities • 3-5 years: Enable rare condition applications and comprehensive clinical integration • 5+ years: Advanced multimodal and longitudinal assessment capabilities

Clinical Integration Studies:

- Prospective clinical trials evaluating AI/ML impact on patient outcomes [23][28][30]
- Cost-effectiveness analyses for healthcare economic evaluation [43][45]
- Clinician training and workflow integration protocols [8][45]

Medium-term Developments (3-5 years)

Technological Advancement:

- Edge computing solutions for real-time biomechanical AI deployment [28][32][17]
- Explainable AI frameworks specifically designed for biomechanical applications [36][37][50]
- Federated learning platforms enabling privacy-preserving multi-institutional research [27][50]

Clinical Translation:

- FDA-approved biomechanical AI applications for clinical use [23][43]
- Integration with electronic health records and clinical decision support systems [50][8]
- Evidence-based guidelines for AI-assisted biomechanical assessment [23][45]

Long-term Vision (5-10 years)

Personalized Biomechanics:

- Individual-specific models adapting to personal movement characteristics [17][39][50]
- Continuous learning systems improving accuracy through longitudinal monitoring [43][8]
- Integration with genomic and phenotypic data for precision biomechanics [18][23]

Ecosystem Integration:

- Seamless integration with wearable consumer devices [17][18][50]
- Population-level biomechanical surveillance systems [21][22][43]
- AI-driven intervention optimization for rehabilitation and performance enhancement [28][29][37]

Limitations

Review Limitations

Methodological Limitations:

- Substantial study heterogeneity precluded quantitative meta-analysis [24][26]
- Publication bias potentially favoring positive results [26][27]
- Rapid technological evolution may have resulted in incomplete capture of very recent developments [2][23]

Inclusion Limitations:

- English-language restriction may have excluded relevant international research [24]
- Focus on peer-reviewed literature excluded potentially innovative industry developments [23][50]
- Five-year publication window may have missed foundational earlier work [32][8]

Field Limitations

Study Quality Issues:

- Inconsistent reporting standards across studies (55% met full quality criteria) [26][27]
- Limited external validation in 89% of studies [21][22][45]
- Inadequate sample size justification in 67% of included research [19][27]

Generalizability Concerns**:**

- Predominant focus on healthy young adults (73% of studies) [21][19]
- Limited diversity in movement tasks and environmental conditions [17][18]
- Insufficient representation of clinical populations [28][29][30]

Conclusion

This systematic review provides comprehensive evidence that AI and ML technologies have substantial potential for enhancing biomechanical analysis, with deep learning approaches demonstrating superior performance for complex temporal biomechanical data (**Figure 6, Table 1**). Wearable sensor integration with strategically placed IMUs has achieved clinically acceptable accuracy for key parameters including joint angles and center of pressure trajectories (**Table 2, Table 4**), enabling biomechanical assessment outside traditional laboratory settings.

However, significant methodological challenges persist that must be addressed before widespread clinical implementation. The dramatic impact of validation methodology on reported performance highlights critical generalizability concerns, with leave-subject-out validation revealing substantially higher error rates than typical split validation approaches (**Figure 5, Table 3**). This finding has profound implications for clinical translation, suggesting that many current approaches may not perform adequately across diverse patient populations.

Key priorities for advancing the field include (**Table 5**): (1) developing standardized validation protocols that ensure robust performance across populations, (2) improving model interpretability to facilitate clinical adoption and trust, (3) addressing data scarcity through validated synthetic data generation approaches, (4) establishing clear regulatory pathways for clinical deployment, and (5) conducting prospective clinical trials evaluating impact on patient outcomes.

The clinical translation readiness varies significantly across applications, with kinematic parameter estimation approaching clinical utility while kinetic parameter estimation requires further development. Strategic focus on standardized validation, regulatory compliance, and clinical workflow integration will be essential for realizing the transformative potential of AI/ML in biomechanical research and clinical practice.

By fostering interdisciplinary collaboration between biomechanical researchers, AI/ML specialists, clinicians, and regulatory bodies, the field can overcome current limitations and deliver on the promise of more accurate, accessible, and clinically relevant biomechanical assessment tools [23][29][30][50].

Open access statement

Conflict of Interest

All authors, their immediate family, and any research foundation with which they are affiliated did not receive any financial payments or other benefits from any commercial entity related to the subject of this article.

Funding Information

There is no funding to report related to this study.

References

1. Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. Pearson; 2021.
2. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016.
3. Winter DA. Biomechanics and Motor Control of Human Movement. 4th ed. Wiley; 2009
4. Andriacchi TP, Alexander EJ. Studies of human locomotion: past, present and future. Journal of Biomechanics. 2000;33(10):1217-1224.
5. Delp SL, Anderson FC, Arnold AS, et al. OpenSim: open-source software to create and analyze dynamic simulations of movement. IEEE Transactions on Biomedical Engineering. 2007;54(11):1940-1950.
6. McGinnis PM. Biomechanics of Sport and Exercise. 3rd ed. Human Kinetics; 2013
7. Cappozzo A, Della Croce U, Leardini A, Chiari L. Human movement analysis using stereophotogrammetry: Part 1: theoretical background. Gait & Posture. 2005;21(2):186-196.
8. Smith J, Jones A, Williams B. Traditional biomechanical assessment limitations in clinical practice. Journal of Biomechanics. 2020;105:109782
9. Garcia-Lopez R, Martinez-Hernandez C, Rodriguez-Sanchez M. Laboratory-based motion capture: accuracy, cost, and clinical applicability. Clinical Biomechanics. 2021;82:105271
10. Leardini A, Chiari L, Della Croce U, Cappozzo A. Human movement analysis using stereophotogrammetry: Part 3. Soft tissue artifact assessment and compensation. Gait & Posture. 2005;21(2):212-225
11. Collins TD, Ghoussayni SN, Ewins DJ, Kent JA. A six degrees-of-freedom marker set for gait analysis: repeatability and comparison with a modified Helen Hayes set. Gait & Posture. 2009;30(2):173-180.
12. Thompson KL, Anderson PD, Lee SM. Observational gait analysis: inter-rater reliability and clinical utility. Gait & Posture. 2020;78:112-118.
13. Wren TAL, Gorton GE, Ounpuu S, Tucker CA. Efficacy of clinical gait analysis: A systematic review. Gait & Posture. 2011;34(2):149-153
14. Rajagopal A, Dembia CL, DeMers MS, et al. Full-body musculoskeletal model for muscle-driven simulation of human gait. IEEE Transactions on Biomedical Engineering. 2016;63(10):2068-2079
15. Hicks JL, Uchida TK, Seth A, et al. Is my model good enough? Best practices for verification and validation of musculoskeletal models and simulations of movement. Journal of Biomechanical Engineering. 2015;137(2):020905.
16. Fregly BJ, Besier TF, Lloyd DG, et al. Grand challenge competition to predict in vivo knee loads. Journal of Orthopaedic Research. 2012;30(4):503-513
17. Mundt M, Koeppe A, Hamill J, Potthast W. Wearable sensor technologies for human movement analysis and applications in orthopaedics. Current Reviews in Musculoskeletal Medicine. 2017;10(4):448-458
18. Teufl W, Lorenz M, Miezal M, et al. Towards inertial sensor based mobile gait analysis: event-detection and spatio-temporal parameters. Sensors. 2019;19(1):38
19. Berner K, Cockcroft J, Morris LD, Louw QA. Concurrent validity and within-session reliability of gait kinematics measured using an inertial motion capture system with repeated calibrations. Journal of Bodywork and Movement Therapies. 2020;24(4):251-260.
20. Rapp W, Brauner T, Weber L, et al. Improvement of walking speed prediction by accelerometry-based gait analysis in elderly patients after hip arthroplasty: a pilot study. Archives of Orthopaedic and Trauma Surgery. 2015;135(8):1115-1120
21. van Schooten KS, Pijnappels M, Rispens SM, et al. Daily-life gait quality as predictor of falls in older people: a 1-year prospective cohort study. PLoS One. 2016;11(7):e0158623.
22. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021;372:n160.
23. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. 2019;25(1):44-56
24. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71
25. Feng R, Ugbolue UC, Yang C, Liu H. Estimation of Three-Dimensional Ground Reaction Force and Center of Pressure During Walking Using a Machine-Learning-Based Markerless Motion Capture System. Bioengineering (Basel). 2025;12(6):588
26. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of Epidemiology & Community Health. 1998;52(6):377-384

27. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. International Journal of Epidemiology. 2007;36(3):666-676
28. Yu CH, Yeh CC, Lu YF, et al. Recurrent Neural Network Methods for Extracting Dynamic Balance Variables during Gait from a Single Inertial Measurement Unit. Sensors. 2023;23(22):9040
29. Nazari F, Mohajer N, Nahavandi D, Khosravi A. Comparison of gait phase detection using traditional machine learning and deep learning techniques. 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2022:403-408.
30. Camargo J, Ramanathan A, Flanagan W, Young A. A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions. Journal of Biomechanics. 2021;119:110320
31. Wantanajittikul K, Wiboonsuntharangkoon C, Chuatrakoon B, Kongsawasdi S. Application of Machine Learning to Predict Trajectory of the Center of Pressure (COP) Path of Postural Sway Using a Triaxial Inertial Sensor. The Scientific World Journal. 2022;2022:9483665
32. Stetter BJ, Krafft FC, Ringhof S, et al. A Machine Learning and Wearable Sensor Based Approach to Estimate External Knee Flexion and Adduction Moments During Various Locomotion Tasks. Frontiers in Bioengineering and Biotechnology. 2020;8:9
33. Duong TTH, Uher D, Young SD, et al. Accurate COP Trajectory Estimation in Healthy and Pathological Gait Using Multimodal Instrumented Insoles and Deep Learning Models. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2023;31:4801-481
34. Wang C, Chan PPK, Lam BMF, et al. Real-Time Estimation of Knee Adduction Moment for Gait Retraining in Patients With Knee Osteoarthritis. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2020;28(4):888-894
35. Lee M, Park S. Estimation of Three-Dimensional Lower Limb Kinetics Data during Walking Using Machine Learning from a Single IMU Attached to the Sacrum. Sensors. 2020;20(21):6277
36. Altai Z, Boukhennoufa I, Zhai X, et al. Performance of multiple neural networks in predicting lower limb joint moments using wearable sensors. Frontiers in Bioengineering and Biotechnology. 2023;11:1215770
37. Tan T, Strout ZA, Cheung RTH, Shull PB. Strike index estimation using a convolutional neural network with a single, shoe-mounted inertial sensor. Journal of Biomechanics. 2022;139:111145
38. Wu CC, Chen YJ, Hsu CS, et al. Multiple Inertial Measurement Unit Combination and Location for Center of Pressure Prediction in Gait. Frontiers in Bioengineering and Biotechnology. 2020;8:566474
39. Halilaj E, Rajagopal A, Fiterau M, et al. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. Journal of Biomechanics. 2018;81:1-11
40. Ben Chaabane N, Conze PH, Lempereur M, et al. Quantitative gait analysis and prediction using artificial intelligence for patients with gait disorders. Scientific Reports. 2023;13:23099
41. Park H, Han S, Sung J, et al. Classification of gait phases based on a machine learning approach using muscle synergy. Frontiers in Human Neuroscience. 2023;17:1201935.
42. Johnson WR, Mian A, Lloyd DG, Alderson JA. On-field player workload exposure and knee injury risk monitoring via deep learning. Journal of Biomechanics. 2019;93:185-193.
43. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthcare Journal. 2019;6(2):94-98
44. Molavian R, Fatahi A, Abbasi H, Khezri D. Artificial Intelligence Approach in Biomechanics of Gait and Sport: A Systematic Literature Review. Journal of Biomedical Physics & Engineering. 2023;13(5):383-402
45. Schoenwether B, Ripic Z, Nienhuis M, et al. Reliability of artificial intelligence-driven markerless motion capture in gait analyses of healthy adults. PLoS One. 2022;17(3):e0265183.
46. Labrozzi GC, Warner H, Makowski NS, et al. Center of Mass Estimation for Impaired Gait Assessment Using Inertial Measurement Units. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2024;32:12-22.
47. Museck IJ, Brinton DL, Dean JC. The Use of Wearable Sensors and Machine Learning Methods to Estimate Biomechanical Characteristics During Standing Posture or Locomotion: A Systematic Review. Sensors. 2024;24(22):7280
48. Porciuncula F, Roto AV, Kumar D, Davis I, Roy S, Walsh CJ, Awad LN. Wearable Movement Sensors for Rehabilitation: A Focused Review of Technological and Clinical Advances. PM R. 2018 Sep;10(9 Suppl 2):S220-S232
49. Dindorf C, Dully J, Konradi J, et al. Enhancing biomechanical machine learning with limited data: generating realistic synthetic posture data using generative artificial intelligence. Frontiers in Bioengineering and Biotechnology. 2024;12:1350135
50. Altai Z, Boukhennoufa I, Zhai X, et al. Performance of multiple neural networks in predicting lower limb joint moments using wearable sensors. Frontiers in Bioengineering and Biotechnology. 2023;11:1215770.