

Artificial Intelligence Enabled Wearable Technologies for Predicting Functional Decline and Supporting Rehabilitation and Independent Living in Adults with Physical Disabilities: A Systematic Review and Translational Framework for Saudi Arabia

Waseem Mumtaz Ahamed

Department of Physical Therapy, College of Nursing & Health Sciences, Jazan University, Saudi Arabia

DOI: <https://doi.org/10.5281/zenodo.22829102>

Published Date: 18-September-2026

Abstract: Background: Artificial intelligence (AI) and wearable sensors are increasingly capable of quantifying mobility, gait, balance, upper-limb activity, physiological status, and rehabilitation dose outside conventional clinical encounters. These technologies could enable a transition from episodic rehabilitation toward continuous, predictive, and personalized disability care. However, the extent to which AI-enabled wearables can predict functional deterioration, estimate rehabilitation outcomes, and support independent living remains uncertain.

Objective: To systematically synthesize evidence on AI-enabled wearable technologies used to predict or quantify functional decline, rehabilitation outcomes, mobility, activities of daily living, fall risk, and related indicators of independent living in adults with physical disabilities, and to develop a clinically grounded translational framework for implementation in Saudi Arabia.

Methods: The review is designed and reported in accordance with PRISMA 2020 and PRISMA-S. Eligible evidence includes prospective and retrospective prediction-model studies, validation studies, rehabilitation-monitoring studies, and controlled intervention studies in adults with neurological or physical disability in which wearable sensor data are analyzed using machine-learning or AI methods. The proposed search covers MEDLINE/PubMed, Embase, Scopus, Web of Science, CINAHL, IEEE Xplore, and CENTRAL from inception to September 2026. Data extraction follows CHARMS domains. Prediction-model quality, risk of bias, and applicability are evaluated using PROBAST+AI; reporting completeness is interpreted against TRIPOD+AI. Because of expected heterogeneity in populations, sensors, outcomes, algorithms, and validation strategies, narrative synthesis is primary, with meta-analysis reserved for clinically and methodologically comparable studies.

Results: Verified evidence demonstrates strongest development in post-stroke rehabilitation, where inertial and multimodal wearable signals have been used to estimate clinical motor scales and predict discharge ambulation, independence, fall risk, and longer-term functional status. Wearable data improved prediction over clinical variables alone in several studies, although performance varied by outcome, baseline ambulatory status, and validation design. Parkinson disease studies demonstrate clinically relevant detection and short-horizon prediction of freezing of gait, while multiple-sclerosis studies show the feasibility of predicting disability and fatigue from mobile and wearable features. Across the field, inertial measurement units predominate; random forests, support-vector methods, gradient boosting, neural networks, and deep-learning architectures are frequently used. Major limitations are small samples, single-center development, heterogeneous sensor placement and feature engineering, limited calibration reporting, insufficient external validation, risk of participant-level data leakage, and limited evidence that algorithm use improves patient-centered outcomes.

Conclusion: AI-enabled wearable rehabilitation has progressed from activity recognition toward clinically meaningful prediction, but the evidence base remains stronger for technical validity than for demonstrated clinical impact. The next generation of systems should combine longitudinal individualized baselines, multimodal sensing, uncertainty-aware prediction, explainable clinical decision support, adaptive rehabilitation feedback, and prospective impact evaluation. Saudi Arabia provides a favorable implementation context because national digital-health policy explicitly supports AI, cloud infrastructure, telemedicine, accessibility, and rehabilitation services. A staged national pathway emphasizing validation, interoperability, cybersecurity, equity, and patient-centered outcomes could position AI-enabled rehabilitation as a scalable component of disability care.

Keywords: artificial intelligence; wearable sensors; rehabilitation; physical disability; functional decline; machine learning; digital biomarkers; independent living; remote monitoring; Saudi Arabia.

Highlights

- Wearable AI is most mature for objective functional assessment and short-term prognostic classification, particularly after stroke.
- Sensor-derived features can add predictive information beyond demographics and conventional clinical scales, but gains are outcome- and population-dependent.
- The evidence base is constrained by small samples, inconsistent validation, limited calibration reporting, and scarce prospective impact studies.
- Independent living should be evaluated as a multidimensional outcome spanning mobility, self-care, participation, falls, caregiver burden, and healthcare utilization.
- A closed-loop Saudi implementation model should connect continuous sensing, individualized risk prediction, clinician review, personalized intervention, and longitudinal reassessment.

1. INTRODUCTION

Physical disability is frequently accompanied by fluctuating or progressive limitations in mobility, balance, dexterity, endurance, self-care, and community participation. Traditional rehabilitation surveillance is typically intermittent: a patient attends a clinic, completes standardized tests, receives an intervention plan, and is reassessed after days or weeks. This model can miss meaningful changes that occur between encounters and may underrepresent performance in everyday environments.

Wearable sensors provide a complementary measurement layer. Inertial measurement units (IMUs), accelerometers, gyroscopes, magnetometers, pressure sensors, surface electromyography, smartwatches, instrumented insoles, and physiological sensors can acquire high-frequency data during ordinary activity. When combined with machine-learning methods, these signals can be transformed into digital measures of gait, balance, upper-limb use, activity patterns, fatigue, fall risk, and rehabilitation adherence. Systematic reviews have documented rapid expansion of wearable-sensor and machine-learning applications in rehabilitation, with IMUs and supervised learning methods among the most frequently used technologies.

Technological performance, however, is not equivalent to clinical usefulness. A classifier may distinguish movement patterns with high accuracy yet provide little information about future disability, meaningful change, or treatment response. Prediction models also face methodological hazards including overfitting, optimistic internal validation, leakage of observations from the same participant across training and test sets, unstable feature selection, incomplete calibration assessment, and poor generalizability across devices, institutions, disability severity, and cultural contexts.

The clinically important question is therefore shifting from whether wearable AI can recognize human movement to whether it can identify impending functional deterioration early enough to support a beneficial change in care. This includes predicting loss of community ambulation, increased fall risk, decline in activities of daily living (ADLs), reduced participation, poor treatment response, or need for additional services. For persons with long-term disability, the ultimate value proposition is not algorithmic accuracy alone but preserved independence and quality of life.

This question has particular relevance to Saudi Arabia. The Saudi Ministry of Health identifies digital health as an enabler of safe, high-quality, sustainable and accessible care, including rehabilitation, and national planning explicitly includes artificial intelligence, Internet of Things technologies, smart devices, cloud computing and telemedicine. Recent Ministry initiatives also prioritize accessible digital services for persons with disabilities. These policy directions create an implementation environment in which evidence-based wearable rehabilitation could be tested at scale.

Accordingly, this review synthesizes current evidence on AI-enabled wearable technologies for functional assessment and prediction in adults with physical disabilities, critically evaluates methodological maturity, and proposes a translational framework for clinically safe implementation with specific relevance to Saudi Arabia.

2. REVIEW OBJECTIVES AND QUESTIONS

Primary objective: To determine how accurately and effectively AI-enabled wearable technologies predict or quantify clinically meaningful functional outcomes in adults with physical disabilities.

Primary review question: Among adults with physical disabilities, how accurately and effectively do AI-enabled wearable technologies detect or predict functional decline and support rehabilitation outcomes, mobility, ADLs, participation, fall-risk management, and independent living compared with conventional assessment or rehabilitation approaches?

- Secondary questions:
- Which wearable sensor modalities, placements, and data-collection contexts are most commonly used?
- Which machine-learning approaches and validation strategies are associated with more credible performance estimates?
- Which disability populations and functional domains are well represented, and where are evidence gaps greatest?
- Does wearable-derived information add predictive value beyond conventional clinical variables?
- What methodological, ethical, interoperability, accessibility, and implementation requirements must be addressed before deployment in Saudi rehabilitation pathways?

3. METHODS

3.1 Design and reporting framework

This review is designed according to PRISMA 2020 and the PRISMA-S extension for transparent reporting of literature searches. Prediction-model evidence is framed and extracted using CHARMS. PROBAST+AI is specified for quality, risk-of-bias and applicability assessment of prediction models, and TRIPOD+AI is used as a contemporary benchmark for completeness and transparency of prediction-model reporting.

3.2 Eligibility criteria

Domain	Include	Exclude
Population	Adults (≥ 18 years) with physical disability or clinically significant motor/functional limitation, including stroke, Parkinson disease, multiple sclerosis, spinal cord injury, traumatic brain injury, neuromuscular disorders, musculoskeletal disability, frailty-related mobility disability, or mixed rehabilitation populations.	Healthy-only datasets unless used as a comparator; pediatric-only studies; studies without a functionally relevant disability outcome.
Index technology	Body-worn or wearable sensors (e.g., IMU, accelerometer, gyroscope, smartwatch, pressure insole, sEMG, multimodal wearable platform) combined with AI/ML or statistical learning for prediction, classification, estimation, monitoring, or personalization.	Non-wearable imaging-only AI; robot control without wearable sensing; purely laboratory optical motion capture unless integrated with wearable data.
Outcomes	Functional status, gait, balance, upper-limb function, ADLs, participation, fall risk, rehabilitation response, disability progression, adherence, activity, fatigue, or independent-living proxies.	Pure signal-processing outcomes without clinical/functional interpretation.
Study design	Prediction-model development/validation, prospective or retrospective cohort, diagnostic/assessment validation, clinical trial, comparative rehabilitation study.	Editorials, protocols without outcome data, purely technical bench studies. Systematic reviews are retained for background and citation chasing but not pooled with primary studies.
Language/status	Peer-reviewed full-text studies in English; no geographical restriction.	Abstract-only records if adequate study data cannot be recovered.

3.3 Information sources and proposed search

The final submission search should be executed from database inception to the submission cut-off date (recommended: September 2026) in MEDLINE/PubMed, Embase, Scopus, Web of Science Core Collection, CINAHL, IEEE Xplore, and Cochrane CENTRAL. Forward and backward citation searching should be performed for all included studies and for key systematic reviews. Trial registries may be searched for ongoing impact-evaluation studies. The complete strategy should be peer-reviewed using a PRESS-style process where feasible.

3.4 Core search concept

(wearable* OR "wearable sensor*" OR acceleromet* OR gyroscop* OR "inertial measurement unit*" OR IMU OR smartwatch* OR "smart watch*" OR "instrumented insole*" OR "body-worn sensor*" OR "surface electromyography" OR sEMG)

AND

("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR "random forest*" OR "support vector*" OR XGBoost OR "gradient boosting" OR "predictive model*" OR algorithm*)

AND

(rehabilitation OR disability OR disabled OR stroke OR "Parkinson*" OR "multiple sclerosis" OR "spinal cord injur*" OR "brain injur*" OR frailty OR mobility OR gait OR balance OR "activities of daily living" OR "functional status")

AND

(predict* OR prognos* OR assess* OR monitor* OR outcome* OR decline OR deteriorat* OR recovery OR fall*)

3.5 Study selection

Records should be deduplicated using reference-management software, followed by independent title/abstract screening by two reviewers and independent full-text eligibility assessment. Disagreements should be resolved by consensus or a third reviewer. Reasons for full-text exclusion should be recorded prospectively and reported in a PRISMA 2020 flow diagram.

3.6 Data extraction

A standardized form should capture CHARMS-relevant domains: publication characteristics; setting; sample size; participant characteristics; disability diagnosis and severity; sensor type, placement and sampling protocol; real-world versus structured-task acquisition; preprocessing; candidate predictors; outcome definition and timing; algorithm class; feature selection; missing-data handling; class imbalance; model-development sample size; resampling; participant-level separation between training and test data; internal/external validation; discrimination; calibration; overall performance; decision-curve or clinical-utility measures; interpretability; fairness; and availability of code/model parameters.

3.7 Risk of bias and applicability

Prediction-model studies should be evaluated using PROBAST+AI, which separates model-development quality from model-evaluation risk of bias and applicability across participants/data sources, predictors, outcomes, and analysis. Particular attention should be paid to small effective sample size, improper splitting of repeated measurements, predictor assessment influenced by outcome status, post-outcome predictors, inadequate handling of missingness, data-driven feature selection without nested validation, hyperparameter tuning on test data, and absence of external validation.

3.8 Synthesis strategy

Because sensor modalities, clinical populations, target outcomes, and reported performance metrics are highly heterogeneous, the primary synthesis should be structured narrative synthesis and evidence mapping. Meta-analysis should only be attempted when at least three clinically comparable external or properly separated validation studies report compatible performance measures for the same target outcome and prediction horizon. Discrimination and calibration should be synthesized separately. Accuracy alone should not be treated as sufficient evidence of clinical usefulness.

4. RESULTS: CURRENT EVIDENCE BASE

The verified literature shows a rapidly expanding but methodologically heterogeneous evidence base. Three broad application classes recur: (1) automated estimation of established clinical scales or movement quality; (2) prediction of future functional outcomes or adverse events; and (3) closed-loop or near-real-time detection intended to trigger rehabilitation feedback or cueing. Evidence is strongest in stroke and Parkinson disease, with emerging applications in multiple sclerosis and spinal cord injury.

4.1 Systematic-review level evidence

A 2023 systematic review of wearable sensors and machine-learning algorithms in rehabilitation identified 32 studies from Web of Science and IEEE Xplore, with IMUs commonly used and machine-learning methods applied across stroke and other rehabilitation populations. A separate post-stroke systematic review included 33 publications and categorized applications into activity recognition, movement classification, and emulation of clinical assessment. A 2026 systematic review across clinical monitoring, rehabilitation and disease-risk assessment included 30 studies, of which IMU-based sensors were used in approximately two thirds; random forest and deep learning were among the most frequent analytical methods. That review highlighted promising fall-risk and mobility prediction but emphasized lack of standardization and insufficient real-world validation.

4.2 Post-stroke rehabilitation and functional prediction

Post-stroke rehabilitation provides the clearest proof-of-concept that wearable-derived information can add value beyond conventional assessment. In an inpatient rehabilitation cohort of 55 participants, early lower-limb inertial-sensor data improved prediction of discharge ambulation and fall risk compared with models based on patient information and clinical assessments alone. Weighted F1 scores improved by 19.6% for ambulation and 23.4% for fall-risk prediction in participants who were ambulatory at admission; best model accuracies were 84.4% for ambulation, 68.8% for independence and 65.9% for fall-risk classification. Importantly, wearable balance data did not add predictive value in participants who were nonambulatory at admission, illustrating that digital biomarkers are context dependent.

An earlier study similarly found that admission IMU data from the ankles and pelvis improved prediction of household versus community ambulation at discharge over conventional clinical predictors. These studies are clinically important because they target future functional status rather than only recognizing movements.

Wearables can also reproduce or approximate standardized motor assessments. A multicenter evaluation of 120 stroke patients used nine-axis IMUs and flex sensors with dynamic time warping and support-vector regression to estimate Fugl-Meyer motor scores. The reported coefficient of determination between system and clinician total scores was 0.9667, and automated assessment required less time. More recent work has used wearable inertial sensors to estimate Action Research Arm Test scores and fine-grained Fugl-Meyer upper-extremity scores, demonstrating that clinically interpretable digital measurement can reduce assessment burden.

Deep-learning approaches have additionally been used to quantify rehabilitation dose. In 48 individuals with chronic stroke, IMU data were used to classify functional movement primitives during rehabilitation activities; a convolutional neural network achieved approximately 70% average classification accuracy. Although this is not direct prediction of future decline, it addresses an important causal pathway: the ability to measure actual training exposure and movement quality at scale.

The limits of prognostic performance are equally informative. A larger passive-wearable study of post-stroke functional status used wrist-worn triaxial accelerometry in 362 patients and reported modest AUROC values for 1-month and 6-month modified Rankin Scale predictions, with random forest models outperforming logistic regression and LSTM approaches but remaining below levels usually considered sufficient for stand-alone clinical decisions. This reinforces the need to integrate sensor data with clinical context rather than assuming that high-frequency data automatically produce high-value predictions.

4.3 Parkinson disease and gait instability

In Parkinson disease, wearables are especially relevant to intermittent gait phenomena that are difficult to reproduce during clinic visits. A study using two shin-mounted inertial sensors in 11 participants demonstrated short-horizon prediction of freezing of gait (FOG). In leave-one-subject-out validation, pre-FOG classification achieved approximately 85% sensitivity, specificity and accuracy, although performance changed with dopaminergic state. This is a clinically meaningful example of prediction before an adverse mobility event rather than detection after it occurs.

More recent systems integrate detection with cueing. The ParCuR ankle-worn device combines inertial sensing, an AI classifier and haptic stimulation, reporting high sensitivity and specificity for FOG detection in development testing and sub-second triggering of sensory cues. Such closed-loop systems represent a potential transition from monitoring to active assistive rehabilitation, but their long-term effectiveness, false-alert burden and impact on falls and participation require prospective evaluation.

4.4 Multiple sclerosis and other disability populations

Evidence is expanding beyond stroke and Parkinson disease. A 2024 digital-health study collected smartphone and wearable data from 55 people with multiple sclerosis and 24 controls over two weeks and evaluated machine-learning approaches for classification of disease subtype and prediction of disability and fatigue. Such studies support the concept that continuous behavioral and physiological features may quantify disability domains that are incompletely captured during episodic assessments.

Spinal cord injury is increasingly proposed as a model population for precision wearable monitoring because autonomic dysfunction, sleep disturbance, activity, pressure exposure and secondary complications create multidimensional risk profiles. A 2026 special communication emphasized that successful wearable/ML integration in spinal cord injury depends on matching device capabilities to physiologic relevance, using individualized models where appropriate, and accounting for contextual noise and within-person variability. However, large prospective prediction studies and impact trials remain limited.

4.5 Evidence table of representative verified studies

Study	Population	Wearable data	Model	Target	Key finding
Kaku et al., 2020	Chronic stroke; n=48	Upper-body IMUs	CNN / deep learning	Functional movement-primitives during rehabilitation	~70% average classification accuracy; demonstrates objective rehabilitation-dose measurement.
Lee et al., 2021	Stroke/TBI brain-injury survivors	Wearable motion sensors + clinical data	Gaussian process regression	Upper-limb rehabilitation outcome	Wearable plus clinical data used to predict/monitor individual functional recovery.
Wearable Sensors Improve Prediction..., 2022	Inpatient stroke; n≈33–35	Ankle + pelvis IMUs	Supervised ML	Discharge ambulation	Wearable gait data improved prediction over conventional clinical model.
Gao et al., 2022	Stroke; n=120, 2 hospitals	9-axis IMUs + flex sensors	DTW + SVR	Fugl-Meyer motor score	R ² =0.9667 versus clinician assessment; shorter assessment time.
Early Prediction..., 2024	Inpatient stroke; n=55	Lower-limb IMUs during gait/balance	Supervised classifiers	Ambulation, independence, fall risk at discharge	Sensor data improved ambulation/fall-risk prediction in participants ambulatory at admission.
Mazilu/FOG prediction study, 2021	Parkinson disease; n=11	Bilateral shin inertial sensors	Multiple ML classifiers	Pre-freezing of gait	~85% sensitivity/specificity/accuracy in LOSO validation; medication state influenced performance.
Modeling MS..., 2024	Multiple sclerosis; n=55 + 24 controls	Mobile + wearable sensor data	ML pipeline	Disability/fatigue and MS phenotype	Demonstrated feasibility of sensor-based disability/fatigue prediction.
ARAT wearable estimation, 2025	Chronic stroke; n=23	Wrist + waist inertial sensors	ML regression + SHAP	ARAT total score	Median absolute error 3.81; R ² =0.93.
AI wrist accelerometry after stroke, 2025/26	Stroke; n=362	Wrist triaxial accelerometry	RF, LR, LSTM	1- and 6-month mRS / change	RF best; AUROC modest (~0.59–0.65), illustrating limits of passive data alone.

Fine-grained FMA-UE, 2026	Post-stroke validation cohort	IMUs + glove flex sensors	RF + regression	Fine-grained FMA-UE	Average item-level RF accuracy reported ~0.954; reduced assessment burden.
ParCuR, 2025	Parkinson disease; development + small usability study	Ankle IMU + haptic actuator	AI classifier	FOG detection / cueing	Reported 94.9% sensitivity, 91.3% specificity; cue onset within 0.7 s.
Fall-risk assessment, 2026	Stroke	Plantar pressure + sEMG + IMU	RF/SVM/KNN/etc.	Binary and 3-class fall risk	Multimodal RF reported very high classification accuracy; requires cautious validation for overfitting/generalizability.

4.6 Cross-cutting technological findings

- IMUs are the dominant modality because acceleration and angular velocity capture gait, balance and limb kinematics with relatively low burden.
- Multimodal platforms may improve discrimination by combining complementary information (e.g., inertial, plantar-pressure, electromyographic and physiological signals), but they increase calibration, usability and interoperability complexity.
- Random forest, support-vector methods and gradient-boosting algorithms remain common for engineered features; convolutional and recurrent neural networks are increasingly used for raw or minimally processed time-series data.
- Participant-level validation is essential. Window-level random splits can produce severe information leakage when multiple observations from the same person appear in both training and test sets.
- External validation is uncommon. Apparent performance within a single dataset therefore should not be interpreted as transportable clinical accuracy.
- Calibration, net benefit, failure analysis and uncertainty are reported less consistently than accuracy, AUROC or F1 score, despite being central to clinical decision support.

5. DISCUSSION

5.1 Principal findings

The central finding is that AI-enabled wearables have moved beyond simple activity tracking and can now approximate clinical assessments, quantify rehabilitation exposure, identify gait instability, and in selected settings improve prediction of future function. The strongest evidence of incremental clinical information comes from post-stroke rehabilitation, where sensor-derived gait and balance characteristics can enhance prediction of discharge ambulation or fall risk over conventional variables.

At the same time, the literature does not yet justify broad claims that wearable AI prevents disability progression, reduces hospitalization, or improves quality of life. Most studies evaluate technical or predictive performance rather than downstream patient benefit. This distinction is critical. A model can be statistically accurate without changing treatment, and a treatment recommendation can be reasonable without improving outcomes if adherence, accessibility, clinician workflow, or false-alert burden are poor.

Accordingly, the field should be viewed as being in a translational middle stage: stronger than proof-of-concept sensing, but not yet supported by a large body of prospective impact trials. High-quality publication in this area should therefore emphasize methodological rigor and clinical pathway integration rather than novelty claims based solely on use of AI.

5.2 What should count as 'functional decline'?

Functional decline should be defined prospectively rather than inferred from any change in a sensor feature. Clinically meaningful targets may include transition from community to household ambulation; decline beyond a minimal clinically important difference on validated motor or ADL scales; new recurrent falls; reduced daily step/activity exposure beyond

individualized variability; deterioration in gait speed or turning; reduced paretic-limb use; increased need for assistance; or loss of participation. Different disability populations may require distinct outcome definitions.

A particularly strong design for future work is personalized change detection: each patient serves partly as their own reference, with models identifying persistent deviations from an individualized functional baseline while accounting for day-of-week, medication state, fatigue, environment and comorbidity. This is likely to be more clinically meaningful than applying a single fixed threshold across highly heterogeneous disability populations.

5.3 From prediction to a closed-loop rehabilitation ecosystem

The proposed next-generation ecosystem should operate as a closed loop rather than as a passive dashboard. Wearable signals are collected continuously or during scheduled tasks; validated algorithms estimate current function and risk trajectories; uncertainty-aware alerts are generated only when predefined clinical thresholds are crossed; a clinician reviews contextual information; rehabilitation intensity, exercise selection, cueing, assistive devices or referral are adjusted; and subsequent wearable data test whether the intervention restored the patient toward baseline.

Layer	Core function	Clinical requirement
1. Sensing	IMU, activity, gait, balance, sEMG/pressure/physiology where indicated	Validated devices; low burden; accessible placement; sufficient wear time
2. Context	Medication, fatigue, pain, environment, assistive-device use, therapy dose	Avoid interpreting sensor change without clinical context
3. Analytics	Feature extraction, individualized baseline, risk prediction, uncertainty	Leakage-safe validation; calibration; explainability; fairness
4. Decision support	Prioritized alerts and trend visualization	Clinician-in-the-loop; avoid alarm fatigue; actionable thresholds
5. Intervention	Exercise adaptation, cueing, tele-rehabilitation, referral, device adjustment	Evidence-based protocol; safety escalation pathways
6. Reassessment	Continuous evaluation of response and adverse events	Predefined outcomes; model updating with governance

5.4 Independent living as the patient-centered endpoint

Independent living should not be reduced to step count. A robust outcome framework should map wearable-derived information to the International Classification of Functioning logic: body functions and structures, activity, participation, environmental facilitators/barriers, and personal context. Relevant patient-centered endpoints include safe mobility inside and outside the home, self-care, ability to work or study, social participation, confidence, caregiver burden, need for formal services, falls, emergency utilization and quality of life.

This broader perspective also prevents a common digital-health error: optimizing what sensors measure most easily rather than what matters most to the person. Daily activity can decline because of worsening motor function, but also because of weather, work patterns, depression, pain, transportation or caregiving responsibilities. Human review and contextual data remain essential.

5.5 Implications for Saudi Arabia

Saudi Arabia is well positioned to test this model at health-system level. Ministry of Health strategy describes digital health as a route to personalized care, continuity, efficiency and access, explicitly including rehabilitative care. Current plans also identify AI, Internet of Things, smart devices, cloud computing and telemedicine as technologies for improving health services. The Health Sector Transformation Program reports rehabilitation medicine among services of SEHA Virtual Hospital, creating a natural clinical pathway for remote rehabilitation monitoring.

A Saudi implementation pathway could begin with multicenter prospective validation in stroke rehabilitation, because the international evidence base is strongest there and outcomes such as discharge ambulation, fall risk and ADL independence are clinically actionable. A second phase could extend to Parkinson disease, multiple sclerosis, spinal cord injury and chronic musculoskeletal disability. Regionally diverse recruitment is important to ensure models generalize across age, sex, disability severity, urban/rural context, device access and patterns of daily activity.

Implementation must also address Arabic-language user interfaces, digital accessibility, informed consent, data minimization, cybersecurity, interoperability with national health information infrastructure, clinician workload,

reimbursement, and governance of algorithm updates. Recent Saudi Ministry initiatives specifically call for improved digital accessibility for persons with disabilities, making inclusive interface design a core requirement rather than an optional enhancement.

5.6 Methodological priorities for future research

Priority	Publication/clinical standard
Prospective registration and prespecified analysis	Define intended use, prediction horizon, outcome, candidate predictors and validation strategy before model development.
Adequate sample size	Use prediction-model sample-size principles rather than generic rules; count participants, events and effective degrees of freedom.
Leakage prevention	Split at participant/site level before preprocessing, feature selection or augmentation; use nested resampling when tuning hyperparameters.
External validation	Evaluate across hospitals, devices, regions and clinically relevant subgroups before implementation.
Calibration and utility	Report calibration plots/slopes/intercepts and decision-curve or net-benefit metrics, not only accuracy/AUROC.
Missingness and wear-time	Report non-wear, device failure, adherence and missing-data mechanisms; evaluate whether missingness itself is informative.
Fairness	Assess performance across sex, age, disability severity, skin tone where optical sensors are used, assistive-device users, and socioeconomic/digital-access strata.
Explainability	Provide clinically interpretable feature importance or explanations where they support safe decision-making; do not equate explainability with causality.
Impact evaluation	Test whether using model outputs changes decisions and improves patient outcomes versus usual care.
Model lifecycle governance	Plan monitoring for dataset shift, recalibration, version control, cybersecurity and post-deployment adverse-event reporting.

5.7 Why a conventional meta-analysis may be inappropriate at present

The current literature spans different diseases, prediction horizons, sensor placements, task protocols, model classes, threshold definitions, and performance measures. Pooling these studies simply because they use 'AI' and 'wearables' would create a clinically uninterpretable summary. Meta-analysis should therefore be outcome-specific—for example, externally validated prediction of post-stroke discharge ambulation using IMU data—and should pool compatible discrimination and calibration measures only after confirming comparable target populations and outcome definitions.

6. STRENGTHS AND LIMITATIONS

The review framework is strengthened by explicit separation of technical measurement, prognostic prediction and clinical impact; use of prediction-model specific appraisal methods; and integration of patient-centered rehabilitation outcomes with implementation considerations. It also incorporates the newest reporting and risk-of-bias frameworks for AI prediction models.

The principal limitation of the present manuscript draft is that a full auditable export from all seven proposed bibliographic databases has not yet been completed within this document-generation workflow. Consequently, no definitive PRISMA record counts or study-level PROBAST+AI judgments are reported, and the evidence table should be treated as a verified core evidence set rather than the final exhaustive set of included studies. This limitation is intentionally explicit to avoid fabrication and should be resolved before journal submission by completing the prespecified search, deduplication and dual-reviewer screening process.

A second limitation is expected clinical heterogeneity. 'Physical disability' spans disorders with different mechanisms, trajectories and meaningful outcomes. The final review should therefore stratify results by condition and application class and avoid broad pooled estimates across incompatible populations.

7. CONCLUSION

AI-enabled wearable technologies have credible potential to transform rehabilitation from episodic measurement into continuous, personalized and predictive care. Current evidence supports objective estimation of motor function, quantification of rehabilitation behavior, detection of unstable gait states, and improved prediction of selected functional outcomes, especially after stroke. Yet the field remains limited by small and heterogeneous cohorts, insufficient external validation, underreporting of calibration and clinical utility, and a shortage of prospective studies demonstrating improved patient-centered outcomes.

The most defensible next step is not deployment of autonomous algorithms, but development of clinician-supervised, closed-loop rehabilitation systems that combine individualized longitudinal sensing with validated prediction, transparent uncertainty, actionable alerts and evidence-based intervention. In Saudi Arabia, alignment between this model and national priorities for digital health, telemedicine, AI, accessibility and rehabilitation creates an opportunity for multicenter validation and implementation research. If evaluated rigorously, such systems could support earlier identification of meaningful decline, more responsive rehabilitation, and greater independence for persons living with physical disability.

8. NOVELTY AND CONTRIBUTION

This review advances the literature in three ways. First, it reframes wearable-AI rehabilitation around prediction of clinically meaningful functional decline rather than movement classification alone. Second, it evaluates the field using contemporary prediction-model methodology, emphasizing validation, calibration, leakage prevention and clinical utility. Third, it translates the evidence into a closed-loop implementation framework aligned with Saudi digital-health and disability-accessibility priorities. The resulting model connects sensing, individualized risk estimation, clinician decision support, personalized intervention and longitudinal reassessment as a single rehabilitation pathway.

9. PROPOSED PROSPECTIVE VALIDATION STUDY FOR SAUDI ARABIA

A logical research program arising from this review is a multicenter prospective cohort followed by a pragmatic impact trial.

Component	Recommended design
Population	Adults undergoing inpatient or outpatient neurorehabilitation after stroke; later expansion to Parkinson disease, MS and SCI.
Sites	At least 3–5 rehabilitation centers in different Saudi regions to enable external/geographical validation.
Wearables	Minimal-burden IMUs at wrist and lower limb/pelvis; optional smartwatch physiology; standardized sampling and wear-time rules.
Baseline data	Age, sex, diagnosis, severity, comorbidity, FMA/ARAT where relevant, gait speed, TUG, Berg Balance Scale, Functional Independence Measure or equivalent ADL instrument, assistive-device use.
Primary endpoint	Clinically meaningful decline or failure to achieve expected recovery trajectory over 4–12 weeks; prespecified per population.
Secondary endpoints	Falls, community ambulation, ADL independence, adherence, rehospitalization, quality of life, participation, caregiver burden.
Modeling	Benchmark clinical model first; then incremental value of wearable features; nested cross-validation; temporal and site-based external validation.
Reporting	PRISMA/CHARMS for the review component; TRIPOD+AI for model reporting; PROBAST+AI-compatible methods.
Impact phase	Cluster or stepped-wedge pragmatic trial comparing clinician access to AI risk alerts versus usual care.
Success criterion	Demonstrated improvement in patient-centered outcome or decision quality—not merely higher AUROC.

REFERENCES

- [1] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71.
- [2] Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst Rev*. 2021;10:39. doi:10.1186/s13643-020-01542-z.
- [3] Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10):e1001744. doi:10.1371/journal.pmed.1001744.
- [4] Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378.
- [5] Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*. 2025;388:e082505. doi:10.1136/bmj-2024-082505.
- [6] Wei S, Wu Z. The Application of Wearable Sensors and Machine Learning Algorithms in Rehabilitation Training: A Systematic Review. *Sensors (Basel)*. 2023;23(18):7667. doi:10.3390/s23187667.
- [7] Boukhenoufa I, Zhai X, Utti V, Jackson J, McDonald-Maier K. Wearable sensors and machine learning in post-stroke rehabilitation assessment: A systematic review. *Biomed Signal Process Control*. 2022;71:103197. doi:10.1016/j.bspc.2021.103197.
- [8] Gu B, Kim HS, Kim H, Yoo JI. Advancements in Wearable Sensor Technologies for Health Monitoring in Terms of Clinical Applications, Rehabilitation, and Disease Risk Assessment: Systematic Review. *JMIR Mhealth Uhealth*. 2026;14:e76084. doi:10.2196/76084.
- [9] 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 (Basel)*. 2024;24(22):7280. doi:10.3390/s24227280.
- [10] Kaku A, Parnandi A, Venkatesan A, Pandit N, Schambra H, Fernandez-Granda C. Towards data-driven stroke rehabilitation via wearable sensors and deep learning. *Proc Mach Learn Res*. 2020;126:143-171.
- [11] Gao L, et al. Wearable Intelligent Machine Learning Rehabilitation Assessment for Stroke Patients Compared with Clinician Assessment. *J Clin Med*. 2022;11(24):7467. doi:10.3390/jcm11247467.
- [12] Wearable Sensors Improve Prediction of Post-Stroke Walking Function Following Inpatient Rehabilitation. *PubMed PMID: 36304845*.
- [13] Early Prediction of Poststroke Rehabilitation Outcomes Using Wearable Sensors. *PubMed PMID: 38169444*.
- [14] Lee SI, Adans-Dester CP, O'Brien AT, et al. Predicting and Monitoring Upper-Limb Rehabilitation Outcomes Using Clinical and Wearable Sensor Data in Brain Injury Survivors. *IEEE Trans Biomed Eng*. 2021. doi:10.1109/TBME.2020.3027853.
- [15] Prediction of Freezing of Gait in Parkinson's Disease Using Wearables and Machine Learning. *PubMed PMID: 33477323*.
- [16] Positive impact of short-term gait rehabilitation in Parkinson patients: a combined approach based on statistics and machine learning. *PubMed PMID: 34517568*.
- [17] Modeling multiple sclerosis using mobile and wearable sensor data. *npj Digit Med*. 2024. doi associated with article s41746-024-01025-8.
- [18] Machine Learning-Based Estimation of Upper Extremity Function in Stroke Rehabilitation Using Body-Worn Inertial Sensors. *PubMed PMID: 40644047*.
- [19] Wearable Sensor-Based Fine-Grained and Comprehensive Upper Extremity Motor Function Assessment for Poststroke Rehabilitation. *PubMed PMID: 41475427*.

- [20] Artificial intelligence to predict modified Rankin score after acute stroke symptoms using wrist-worn triaxial accelerometry data. PubMed PMID: 41317816.
- [21] Lopes T, et al. ParCuR—A Novel AI-Enabled Gait Cueing Wearable for Patients with Parkinson's Disease. Sensors (Basel). 2025. PubMed PMID: 41305286.
- [22] Delgado AD, Balthazaar SJT, Soltesz AE, Nightingale TE, Panza GS. Using Machine Learning With Wearable Devices to Advance Research and Patient Care Using Spinal Cord Injury as a Model. Arch Phys Med Rehabil. 2026;107(5):1148-1160. doi:10.1016/j.apmr.2025.09.025.
- [23] World Health Organization. Assistive technology. WHO; updated 2024. <https://www.who.int/news-room/fact-sheets/detail/assistive-technology>
- [24] Saudi Ministry of Health. National E-Health Strategy: The Complete Vision. Updated 2024. <https://www.moh.gov.sa/en/Ministry/nehs/Pages/The-Complete-Vision.aspx>
- [25] Saudi Ministry of Health. Current Plans for Achieving E-Health Objectives. Updated 2026. <https://www.moh.gov.sa/en/ministry/vro/ehealth/pages/plans.aspx>
- [26] Saudi Ministry of Health. Health Ministry Announces Consultation Results on Disability Services. 3 September 2025. <https://www.moh.gov.sa/en/ministry/mediacenter/ads/pages/ads-2025-09-03-001.aspx>
- [27] Saudi Vision 2030. Health Sector Transformation Report 2024. <https://www.vision2030.gov.sa/>