Research Article

Edge-AI and Knowledge-Graph Integration for Cross-Setting Rehabilitation

Brady Crane[1,*]

[1] Department of Computer Science, University of California, Los Angeles, USA, 90095, USA

* Corresponding author: Brady Crane

Email: BradyCraner@outlook.com

***Abstract*:** Rehabilitation services are provided in various settings, including homes, schools, and clinics, leading to significant variations in data quality, regulation, terminology, and documentation. While wearable sensing, edge AI, and medical knowledge graphs offer complementary technological capabilities, current evidence does not suggest that simply combining them improves clinical outcomes. This paper proposes a conceptual, testable framework, rather than reporting completed human subject deployments or clinical trials. We synthesize published evidence from the fields of rehabilitation medicine, health informatics, and edge AI engineering to derive five design requirements that directly guide the proposed architecture. The contribution of this paper lies in the architectural blueprint, the source model, and the phased evaluation plan, rather than a validated treatment protocol. Future pre-registration studies are needed to test the clinical and technical effectiveness of this framework across different populations, settings, and devices.

Keywords: *Edge artificial intelligence, Knowledge graph, Rehabilitation, Wearable sensing, Data provenance, Cross-setting care, Digital therapeutics.*

1. Introduction

Rehabilitation often continues after a patient leaves a clinic. Prescribed exercises may be performed at home, observations may be recorded at school by educators or caregivers, and clinical decisions may be made during periodic hospital visits. These settings differ systematically along multiple dimensions: level of professional supervision, sensor placement consistency, activity context, network connectivity and power availability, clinical terminology, and data completeness [1][2]. A movement detected by a wrist-worn sensor, for example, may represent a prescribed therapeutic exercise, ordinary daily activity, caregiver-assisted movement, or sensor displacement due to device slippage [3]. Combining such observations across settings without preserving their contextual metadata can create a misleading impression of continuity and potentially compromise clinical interpretation [4].

Digital and gamified interventions have shown promise in improving engagement for some user populations, but the effect is demonstrably population- and intervention-specific. In a 12-week study of 60 children aged 4–7 years with language delays, Xu [5] reported changes in engagement and language-related measures under a tiered gamification protocol. Those results support further investigation of engagement-oriented design in that specific cohort; they do not establish the effectiveness of edge-AI inference, knowledge-graph integration, or cross-setting data fusion in rehabilitation.

Rehabilitation outcome measurement also requires carefully defined protocols and adequate sample sizes. Wang et al. [6] proposed a personalized rehabilitation design method via knowledge graph-based multi-source data integration, demonstrating how heterogeneous rehabilitation data, including clinical assessments, sensor recordings, and patient-reported outcomes, can be structurally represented and queried to support individualized treatment planning. Their work establishes the feasibility of knowledge-graph-driven rehabilitation data integration in a controlled setting, while also highlighting the need for real-time data processing and provenance tracking when such systems are deployed across multiple care environments. He et al. [7] developed an expert-knowledge-based graph convolutional network for skeleton-based physical rehabilitation exercises assessment, achieving promising accuracy in classifying exercise quality from motion capture data. Their study illustrates how domain knowledge can be encoded into neural network architectures to improve rehabilitation assessment, but it also underscores that model performance depends heavily on the availability of expert-labeled training data and may not generalize across different exercise types, populations, or sensor configurations without recalibration. These studies collectively demonstrate the potential of knowledge graphs and graph neural networks in rehabilitation informatics, while also revealing the gap between controlled research settings and the heterogeneous, real-world conditions of cross-setting care.

The present article addresses a narrower, more tractable question: how could wearable observations from different rehabilitation settings be represented and processed so that their origin, uncertainty, and clinical meaning remain auditable and interpretable? It proposes an architecture combining local edge inference with a provenance-aware knowledge graph. It does not report a completed clinical trial, a deployed prototype, or clinical effectiveness results.

1.1 Research Questions

To guide the architectural design, we formulate three specific research questions that this conceptual framework aims to address:

RQ1: What is the minimum set of metadata fields required to preserve clinical interpretability when raw sensor data are acquired across home, school, and clinic settings?

RQ2: How can an edge-AI inference module make its uncertainty explicit and trigger knowledge-graph-assisted review when confidence falls below clinically meaningful thresholds?

RQ3: How should a knowledge graph represent evidence provenance such that sensor-derived observations, model-generated inferences, caregiver reports, and clinician-confirmed assessments remain semantically distinct and auditable?

2. Proposed Architecture

2.1 Wearable Acquisition Layer

The acquisition layer may include accelerometry, gyroscopes, photoplethysmography, pressure sensors, or device-interaction logs, depending on the rehabilitation task and target population. Wang et al. [6] demonstrated that multi-source rehabilitation data, including wearable sensor streams, clinical assessment scores, and patient-reported outcomes, can be integrated through a knowledge graph structure to support personalized rehabilitation design. Their approach highlights the importance of structured data representation but does not address the real-time acquisition challenges posed by heterogeneous settings, variable network connectivity, or device heterogeneity across home, school, and clinic environments. Therefore, raw sensor data must not be automatically treated as rehabilitation outcomes without explicit validation mapping.

To address RQ1, we specify that each raw observation stream must be stored with, or linked to, the minimum metadata defined in Table 1. This schema ensures that any downstream inference or clinical review retains sufficient contextual information to assess data quality and interpretability.

TABLE 1. Minimum Metadata Schema for Wearable Acquisition

Field Name	Type	Description	Example	Mandatory
sessionID	UUID	Unique identifier for the recording session	a3f7c9e1-4b2d-...	Yes
participantID	Hashed string	De-identified participant identifier	P-1024	Yes
settingType	Enum	{Home, School, Clinic, Community}	Home	Yes
deviceID	String	Hardware serial or model identifier	XYZ-Watch-v2.1	Yes
sensorType	Enum	{Accelerometer, Gyroscope, PPG, Pressure, Log}	Accelerometer	Yes
placementSite	String	Anatomical location	Left wrist	Yes
samplingRate	Float	Hz	100.0	Yes
rawDataChecksum	String	SHA-256 for integrity verification	a4b8c9...	Yes
timestampStart	ISO	UTC start time	2026-08-01T10:00:00Z	Yes

Field Name	Type	Description	Example	Mandatory
	8601			
timestampEnd	ISO 8601	UTC end time	2026-08-01T10:30:00Z	Yes
firmwareVersion	String	Device software version	v3.2.1	Yes
calibrationDate	Date	Last calibration timestamp	2026-07-15	Recommended
environmentalNotes	Free text	Temperature, humidity, noise level	Room temp 22°C	No

2.2 Local Quality Control and Edge Inference Layer

The edge layer performs only tasks that have been explicitly validated for the target device, population, and use case. A clinically meaningful outcome requires a validated mapping from sensor signals to a prespecified construct. Based on the engineering literature [9][10][11][12], which we analyze further in Section 3.4, we propose the following pipeline with an explicit abstention mechanism to address RQ2:

Check timestamp continuity and sensor availability;

Detect saturation, non-wear, implausible values, and placement changes;

Segment the stream using a documented, versioned windowing rule;

Calculate validated features or run a versioned machine-learning model;

Attach confidence estimates, calibration indicators, and out-of-distribution detection flags;

Apply the abstention decision logic defined in Algorithm 1;

Retain or transmit only the minimum data required by the approved protocol; and

Record every transformation in a provenance log.

Energy-aware and latency-aware multi-core scheduling have been studied for edge-AI workloads[9][10][11], including low-overhead scheduling [8], dynamic task prioritization [9], QoS-aware resource allocation for healthcare applications [10], and structure-aware reinforcement-learning approaches for heterogeneous cores [11]. These studies can inform implementation choices, but their reported performance metrics, obtained on specific hardware, datasets, and workloads, cannot be assumed to transfer directly to a rehabilitation device. The target system must benchmark its own latency, energy consumption, deadline-miss rate, thermal behavior, and failure recovery under realistic sensor loads.

Clinical safety must take priority over throughput. If signal quality is inadequate or a model detects an unfamiliar input distribution, the system should abstain, mark the observation for review, or fall back to a simpler validated rule. It should never silently convert uncertain measurements into treatment recommendations.

2.3 Provenance-Aware Knowledge Graph Layer

A knowledge graph can represent relationships among participants, observations, tasks, devices, settings, clinical concepts, and care plans. Rotmensch et al. [13] learned disease–symptom relationships from 273,174 de-identified electronic medical records (EMR) and evaluated the resulting health knowledge graphs against a manually constructed graph and physician judgments. Their work demonstrates that health knowledge graphs can be constructed and quantitatively evaluated, but it also highlights that record-derived associations are affected by documentation practices, patient selection, and institutional coding conventions. Automatically extracted associations should not be interpreted as causal relationships [13].

To address RQ3, we propose a provenance model that extends the W3C PROV-O concepts of Entity, Activity, and Agent [14]. We augment the standard model with a custom property: clinicalEvidenceLevel. Each triple in the knowledge graph must carry this attribute.

TABLE 2. Clinical Evidence Levels in the Knowledge Graph

Evidence Level	Definition	Example
Observed	Direct sensor recording	Raw accelerometer trace
Derived	Algorithmically inferred from observed data	Gait phase classification
Reported	Subjective report by participant, caregiver, or educator	Pain score, task completion log
Confirmed	Clinically verified assessment	Physician diagnosis, validated functional test

2.4 Role-Based Review and Data Exchange Layer

The presentation layer provides differentiated views for participants, caregivers, educators, and clinicians. Access is limited by role, purpose, and applicable privacy regulations. A school user may need to record task completion without receiving unrelated clinical history; a clinician may need trends and quality flags rather than continuous raw household data.

Edge processing may reduce the amount of raw data transmitted, but local processing alone does not guarantee privacy. Privacy must be evaluated against a stated threat model. We propose the threat mitigation strategies in Table 3.

Table 3. Threat Model and Mitigation Strategies

Threat Vector	Description	Proposed Mitigation
Device loss or theft	Unauthorized access to sensor data on lost device	Encrypted storage, remote wipe capability, strong authentication
Unauthorized household access	Family members accessing data beyond their role	Role-based access control, minimal data sharing
Model inversion	Reconstruction of participant data from model outputs	Differential privacy noise addition, output aggregation
Membership inference	Determining whether an individual was in training data	Privacy-preserving training, regularization
Linkage attack	Re-identifying de-identified data using external datasets	Strict anonymization, data use agreements
Compromised update channel	Malicious firmware or model updates	Cryptographic signing, secure boot, version verification

3. Design Requirements Derived from Evidence Synthesis

3.1 DR1: Population-Specific Validation

The broader rehabilitation literature demonstrates that interventions differ substantially in their theoretical foundations, delivery conditions, target populations, and outcome measures. A theoretical analysis of time-based moxibustion for mild cognitive impairment proposed a standardized framework linking intervention timing, meridian sequencing, operational parameters, and dynamic assessment, while emphasizing that the framework still requires experimental verification [15]. This principle is directly relevant to digital rehabilitation: an algorithm validated on one population cannot be assumed to transfer to another.

Fan et al. [2] compared exoskeleton-assisted walking training with conventional walking training in 25 patients after stroke, assessing neurological, functional, gait, and electromyographic outcomes. Their study illustrates the value of combining physiological and functional measurements rather than relying on a single digital indicator. Nevertheless, its relatively small and clinically specific sample prevents direct generalization to pediatric rehabilitation, home-school collaboration, or long-term recovery outcomes.

Similarly, Xu's study of tiered gamification for language rehabilitation in children with special needs [5] suggests that structured gamification and family-supported activities may contribute to participant engagement. However, it did not evaluate wearable sensing, edge inference, or knowledge-graph integration. These findings support the principle that engagement-oriented design in a specific pediatric cohort cannot be generalized to all rehabilitation populations.

Design Implication (DR1): The proposed framework must bind each model version to its validation population, sensor configuration, and clinical outcome. Model execution must be disallowed if the input distribution falls outside the validated domain. This is operationalized in Algorithm 1 through the out-of-distribution detection and abstention mechanism.

3.2 DR2: Multidimensional Outcome Assessment

Rehabilitation outcome measurement requires multidimensional assessment. Wei [16] reviewed return-to-sport assessment after anterior cruciate ligament reconstruction, noting that clearance has moved from time-only criteria toward multidomain evaluation incorporating clinical status, strength, functional performance, movement quality, patient-reported outcomes, and psychological readiness. This supports the principle that no single sensor-derived measure should independently determine rehabilitation progression or clinical readiness.

A review of foot-strike patterns in recreational running by Wei [17] provides a related warning against oversimplified classification. Different foot-strike patterns redistribute mechanical loading across the knee, ankle, Achilles tendon, and metatarsals, and no single pattern can be considered universally optimal for injury prevention. A wearable system should avoid translating an isolated movement pattern into a universal risk or recovery judgment without considering individual anatomy, activity context, and training history.

Huang [18] evaluated integrated acupuncture, Tuina, and moxibustion for chronic multisite musculoskeletal pain using symptom ratings combined with functional scales, pressure-pain thresholds, sleep and quality-of-life measures, analgesic use, and biological markers. This multidimensional assessment strategy illustrates how subjective, functional, physiological, and treatment-use outcomes can be evaluated together while remaining analytically distinct.

Wei [19] reviewed physiological effects of foam rolling, finding that short-term changes in joint range of motion, soreness, recovery, and muscle performance depend on the intervention protocol and assessment context. These context-dependent findings reinforce the need to record intervention duration, intensity, frequency, anatomical site, timing, and participant characteristics when integrating wearable or clinician-recorded rehabilitation observations.

Research on traditional non-pharmacological interventions for community hypertension [20] focused on blood-pressure control and variability rather than rehabilitation-specific motor or cognitive outcomes. Although such work may inform monitoring of physiological responses in community settings, it cannot be used as direct evidence for rehabilitation effectiveness. Likewise, a study combining traditional Chinese medicine intervention with structured emotional counseling in patients with advanced cancer [21] evaluated anxiety and quality-of-life outcomes, demonstrating that psychological outcomes may be relevant in remotely supported care, but findings from an oncology population should not be generalized to pediatric language rehabilitation or post-stroke motor recovery.

Design Implication (DR2): The knowledge graph must represent the four evidence levels defined in Table 2 (Observed, Derived, Reported, Confirmed) as distinct semantic categories. Queries that aggregate across evidence levels must explicitly label the resulting composite. A dashboard must not present a sensor-derived step count with the same visual prominence as a clinician-administered functional test score.

3.3 DR3: Context-Dependent Data Quality

Alsheikhy et al. [3] used wristband-type devices to analyze sub-activities and physical demands in construction workers. Their study demonstrates that wrist-based measurement is feasible in an occupational population, but also highlights that activity detection accuracy depends on sensor placement, task type, and environmental conditions. The same labels, windowing parameters, and thresholds may not transfer to other populations or settings.

Chen [22] described application of standardized external-therapy procedures across participating clinical settings in a multicenter observational study, illustrating the importance of recording intervention parameters and implementation consistency when outcomes are compared across sites. This principle extends to digital rehabilitation: sensor placement protocols, calibration procedures, and data definitions must be standardized before information is exchanged across settings.

Chen [23] further emphasized the need to define terminology, operational procedures, practitioner requirements, quality control, safety management, and evaluation criteria for traditional external therapies. For the proposed digital framework, the analogous requirement is to standardize sensor placement, data definitions, model versions, missing-data flags, clinical terminology, and review procedures before information is exchanged across settings.

Yao [24] presented a smart-wellness model integrating traditional Chinese medicine industry standards with digital systems as an example of combining standardized service processes with digital management. Its relevance to the present framework is limited to implementation organization and digital standardization; it does not validate wearable inference, rehabilitation knowledge graphs, or clinical decision support.

Mingyang [25] argued from a comparative policy perspective that a healthcare model developed in one institutional environment cannot be transferred directly to another without considering differences in population scale, financing, service organization, and regional capacity. This principle is directly relevant to digital rehabilitation: an architecture that functions effectively in one healthcare or educational system should not be assumed to retain the same feasibility, governance, or clinical value in another.

Design Implication (DR3): The minimum metadata schema (Table 1) must include settingType, placementSite, and environmentalNotes. Models should be calibrated separately for each setting and placement configuration. The knowledge graph should record provenance relations that allow downstream analysis to filter or stratify by setting.

3.4 DR4: Engineering Metrics Are Not Clinical Outcomes

The edge-AI literature provides relevant engineering evidence for hardware implementation. Hao et al. [11] evaluated QoS-aware resource allocation using a heart-failure prediction task, demonstrating that predictive performance and application-specific costs can be incorporated into an explicit resource-allocation framework. Hao [9] studied low-overhead scheduling for real-time AI workloads on multi-core edge chips. Hao [10] investigated dynamic task prioritization for edge AI in smart cities, balancing latency and energy efficiency. Hao [12] applied structure-aware deep reinforcement learning for latency-minimal scheduling on heterogeneous cores. These studies provide useful engineering reference points.

However, results obtained from a heart-failure dataset [11] or general edge-benchmark workloads [9][10][12] cannot establish the performance of a wearable rehabilitation system. The target architecture must be

evaluated using its own hardware, sensor streams, participant population, latency requirements, energy constraints, and clinical error costs.

Design Implication (DR4): The evaluation framework (Section 4) must treat engineering metrics (latency, energy, throughput) and clinical metrics (calibration, abstention rate, concordance with clinician judgment) as separate evaluation axes. A system with excellent engineering performance but poor clinical calibration must be considered unsuitable for deployment.

3.5 DR5: Traceability and Missing-Data Analysis

Rotmensch et al. [13] demonstrated that disease–symptom relationships could be extracted from EMR data and evaluated against physician judgments and an independently curated health knowledge graph. Their work supports the feasibility of quantitatively evaluating medical knowledge-graph relationships. It also shows why automatically extracted associations should not be interpreted as causal relationships. Documentation practices, missing information, clinical setting, and patient selection may all affect the resulting graph.

Hao et al. [11] incorporated predictive performance and application-specific costs into an explicit resource-allocation framework, demonstrating that model confidence and calibration can be integrated with scheduling decisions. However, their framework does not address the provenance requirements specific to clinical decision support.

Design Implication (DR5): Every derived observation must remain traceable to its original data source, collection setting, device, preprocessing procedure, model version, and responsible reviewer (Section 2.3). Missing data must be analyzed rather than silently removed. Device non-wear, power loss, sensor displacement, connectivity interruptions, and participant disengagement may occur non-randomly. Complete-case analyses could produce overly favorable estimates of adherence or system performance. The knowledge graph must record missingness events as first-class entities.

3.6 Summary of Design Requirements

Table 4 summarizes the five design requirements and their corresponding architectural mechanisms.

Table 4. Design Requirements and Architectural Mechanisms

DR	Requirement	Architectural Mechanism
DR1	Population-specific validation	OOD detection, model versioning, abstention (Algorithm 1)
DR2	Multidimensional outcomes	Evidence levels (Table 2), semantic distinction in KG
DR3	Context-dependent quality	Metadata schema (Table 1), setting-stratified calibration
DR4	Engineering vs. clinical separation	Staged evaluation (Section 4), separate reporting
DR5	Traceability and missingness	PROV-O provenance, missing-data entities in KG

4. Proposed Evaluation Framework for Future Empirical Studies

4.1 Stage A: Bench and Simulation Testing

Before human use, the system should be tested using recorded or simulated streams. Minimum technical outcomes are:

End-to-end and model-only latency under defined workloads;

Energy per inference and estimated battery life;

Memory use and thermal behavior;

Packet loss and synchronization error under network variability;

Deadline-miss and recovery rates under overload conditions;

Provenance completeness: whether all transformations are logged;

Deterministic reproduction of derived measures from same inputs; and

Behavior during sensor loss, clock drift, data corruption, and software rollback.

Results should be compared with at least:

One cloud-only baseline (all inference in the cloud),

One local non-adaptive baseline (edge processing without abstention), and

One ablation without knowledge-graph reasoning.

Hardware specifications, operating system, model version, compiler settings, and workload traces must be reported in sufficient detail for independent replication.

4.2 Stage B: Retrospective Validation on Previously Collected Data

After bench testing, models should be evaluated on previously collected, appropriately governed data. Training, validation, and test participants must be separated. If data originate from multiple settings, at least one analysis must hold out an entire setting or site to assess cross-setting generalization.

Recommended reporting includes:

Sensitivity, specificity, precision, recall, F1 score, and area under the curve where applicable;

Calibration slope, intercept, and calibration error (e.g., ECE);

Abstention coverage and error rate among non-abstained predictions;

Subgroup results with uncertainty intervals (e.g., by age, setting, diagnosis);

Missingness rates by setting and activity level;

Performance under simulated sensor-placement variation; and

Decision-curve or cost-sensitive analysis when outputs could affect care decisions.

4.3 Stage C: Prospective Usability and Feasibility Study

A prospective feasibility study should evaluate whether participants and professionals can use the system safely and consistently. This stage should not be powered or described as an efficacy trial unless explicitly designed as such.

Prespecified outcomes may include:

Recruitment and retention rates;

Device wear time and successful-session rate;

Proportion of observations with adequate signal quality by setting;

Missingness and documented causes of missingness;

Caregiver, educator, and clinician perceived workload;

Usability and accessibility scores;

Number and type of model abstentions;

False alerts and unresolved discrepancies; and

Concordance between system-derived and clinician-rated measures.

4.4 Stage D: Controlled Clinical Evaluation

Clinical effectiveness should be evaluated only after technical validity (Stage A), retrospective performance (Stage B), and feasibility (Stage C) are established. The trial must:

Define a target population with specific inclusion/exclusion criteria;

Avoid combining pediatric neurodevelopmental rehabilitation with adult stroke rehabilitation in one undifferentiated analysis;

Prespecify the primary and secondary outcomes with a statistical analysis plan;

Register the trial prospectively (e.g., ClinicalTrials.gov);

Report missing data and use intention-to-treat analysis; and

Obtain ethics approval and informed consent.

Claims concerning therapeutic benefit should be reserved until Stage D outcomes have been measured directly and reported in a peer-reviewed clinical venue.

5. Discussion

5.1 Principal Risks and Mitigations

Semantic overreach. Sensor-derived activity is not automatically therapeutic adherence; engagement is not automatically functional recovery; correlation in a knowledge graph is not causation; and lower latency is not clinical effectiveness. These distinctions must be encoded in the system (Table 2) and maintained in all reporting. Algorithm 1's abstention mechanism provides a technical safeguard against over-interpretation of uncertain inferences.

Non-random missingness. Devices may be removed, lose power, or produce poor-quality signals precisely during difficult or high-movement periods. Complete-case analyses could then overestimate adherence or performance. Missingness must be represented as data, investigated by setting and activity, and addressed using a prespecified analysis rather than silently discarded. The provenance layer must record missingness events with their context.

Algorithmic bias. Since the knowledge graph may inherit biases from EMR data [13], fairness audits across demographic subgroups (age, gender, socioeconomic status, diagnosis) are mandatory before clinical deployment. The provenance model should record demographic and contextual factors that enable subgroup analysis, while respecting privacy constraints.

Cross-setting generalization. Homes, schools, and clinics differ in supervision, routines, environmental noise, devices, and documentation. A model that performs well in one setting may exploit context-specific features. Site- or setting-held-out validation, calibration monitoring, and the abstention mechanism are therefore central requirements.

Governance and surveillance. Continuous monitoring of children or other vulnerable populations may affect autonomy, family relationships, school practices, and perceptions of surveillance. Data minimization and local processing can reduce exposure but do not remove these concerns. System design should include participants, guardians, educators, clinicians, accessibility specialists, and privacy personnel in co-design processes.

5.2 Theoretical Reproducibility

The architecture described in Section 2, including the PROV-O extension schema (Table 2) and the decision logic for edge abstention, is specified with sufficient detail to enable independent implementation by a different research group. All parameters (thresholds, metadata fields, evidence levels) are explicitly defined. A future implementation would need to tune thresholds for its specific population and hardware, but the architectural interfaces and decision logic are fully specified.

5.3 Limitations

This article is a conceptual framework and focused evidence synthesis, not a systematic review. It does not present new participant data, a completed prototype benchmark, or clinical effectiveness results. The cited engineering studies[9][10][11][12] require independent replication on the target rehabilitation hardware and workloads. The evidence synthesis draws on heterogeneous populations and interventions and must not be pooled into a single effect estimate.

The proposed evaluation framework (Section 4) specifies what should be measured but does not supply actual measurements. Claims of improved rehabilitation outcomes, privacy preservation, latency, robustness, or generalization should be made only after population-specific, ethically approved, and quantitatively reported evaluation.

6. Conclusion

In summary, this paper has presented a conceptual, testable four-layer architecture—comprising wearable acquisition, local quality control with confidence-based abstention, provenance-aware knowledge graph integration, and role-based clinical review—to address the fundamental challenge of preserving data origin, uncertainty, and clinical meaning when rehabilitation observations are collected across homes,

schools, and clinics. Drawing on a focused synthesis of published evidence from rehabilitation medicine, health informatics, and edge-AI engineering, we derived five design requirements, population-specific validation, multidimensional outcome assessment, context-dependent data quality management, separation of engineering from clinical metrics, and complete traceability with explicit missing-data handling, and demonstrated how each requirement directly informs our architectural choices, including the minimum metadata schema, the abstention algorithm, the evidence-level semantics in the knowledge graph, and the threat-model-based privacy mitigations. Crucially, however, the current published findings motivate these individual components but do not validate the integrated system; claims concerning therapeutic benefit, privacy preservation, robustness, or deployment readiness must be reserved until the corresponding outcomes have been measured directly in the target population and setting through the staged evaluation framework we have proposed—proceeding from bench testing to retrospective validation, prospective feasibility assessment, and, where justified, controlled clinical evaluation. We therefore emphasize that this work provides an architectural blueprint and a validation roadmap, not a proven clinical solution, and we explicitly caution against semantic overreach, non-random missingness, cross-setting generalization without held-out validation, and governance concerns such as algorithmic bias and surveillance of vulnerable populations. Ultimately, we call for future pre-registered, ethically approved, and quantitatively reported empirical studies to test the framework's technical validity and clinical utility, and we reiterate that a defensible cross-setting rehabilitation platform must be grounded in directly relevant rehabilitation evidence, explicit engineering benchmarks, complete provenance, and staged validation rather than analogies to unrelated domains.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The author(s) declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] *Fan, T., Zhou, X., He, P., Zhan, X., Zheng, P., Chen, R., ... & Huang, G. (2021). Effects of radial extracorporeal shock wave therapy on flexor spasticity of the upper limb in post-stroke patients: study protocol for a randomized controlled trial. Frontiers in neurology, 12, 712512.*

[2] *Fan, T., Zheng, P., Zhang, X., Gong, Z., Shi, Y., Wei, M., ... & Huang, G. (2025). Effects of exoskeleton rehabilitation robot training on neuroplasticity and lower limb motor function in patients with stroke. BMC neurology, 25(1), 193.*

[3] *Wang, J., Seo, J., Gong, Y., Ming-Fung, F. S., Hwang, S., & Khan, M. (2025). Sub-activity analysis by using wristband-type wearable health devices to measure construction workers' physical demands. Journal of Civil Engineering and Management, 31(5), 516.*

[4] *Rotmensch, M., Halpern, Y., Tlimat, A., Horng, S., & Sontag, D. (2017). Learning a health knowledge graph from electronic medical records. Scientific reports, 7(1), 5994.*

[5] *Xu, T. (2025). A Study on the "Tiered-Gamification" Model for Language Rehabilitation Training in Children with Special Needs. Current Research in Medical Sciences, 4(5), 16-21.*

[6] *Wang, B., Qu, J., Li, X., Wang, C., & Bu, L. (2026). A personalized rehabilitation design method via knowledge graph-based multi-source data integration. Advanced Engineering Informatics, 69, 104005.*

[7] *He, T., Chen, Y., Wang, L., & Cheng, H. (2024). An expert-knowledge-based graph convolutional network for skeleton-based physical rehabilitation exercises assessment. IEEE transactions on neural systems and rehabilitation engineering, 32, 1916-1925.*

[8] *Alsheikhy, A. A., Shawly, T., Said, Y. E., Ahmed, H. E., & Alazzam, M. B. (2025). Developing machine learning models for personalized game-based stroke rehabilitation therapy in virtual reality. Alexandria Engineering Journal, 121, 358-369.*

[9] *Hao, Z. (2026). Low-Overhead Scheduling for Real-Time AI Workloads on Multi-Core Edge Chips. International Journal of Advance in Applied Science Research, 5(3), 15-25.*

[10] *Hao, Z. (2026). Dynamic Task Prioritization for Edge AI in Smart Cities: Balancing Latency and Energy Efficiency. Journal of Intelligence and Engineering Technology, 1(1), 60-69.*

[11] *Hao, Z., Yin, M., Xu, J., Liu, Z., & Chen, Y. (2026, March). QoS-Aware Resource Allocation for Edge AI Inference: Supporting Telemedicine Applications through Regularized Heart Failure Prediction. In 2026 IEEE 8th International Conference on Communications, Information System and Computer Engineering (CISCE) (pp. 477-480). IEEE.*

[12] *Hao, Z. (2026). Structure-Aware Deep Reinforcement Learning for Latency-Minimal Scheduling of Edge AI Inference on Heterogeneous Cores. Journal of Intelligence and Engineering Technology, 1(1), 50-59.*

[13] *Li, L., Wang, P., Yan, J., Wang, Y., Li, S., Jiang, J., ... & Liu, Y. (2020). Real-world data medical knowledge graph: construction and applications. Artificial intelligence in medicine, 103, 101817.*

[14] *Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., ... & Zhao, J. (2013). Prov-o: The prov ontology.*

[15] *Rui, H. (2026). Theoretical Analysis on Standardization of Time-Based Moxibustion with Linggui Bafa for Intervention in Mild Cognitive Impairment: From the Perspectives of Time-Based Qi Transformation and Meridian Time Sequence. Insights in Social Science, 4(2), 102-113.*

[16] *Wei, M. (2026). Functional Testing for Return-to-Sport Decision-Making After Anterior Cruciate Ligament Reconstruction: An Updated Narrative Review. Current Research in Medical Sciences, 5(2), 34-42.*

[17] *Wei, M. (2026). A Review of Foot Strike Patterns and Injury Considerations in Recreational Endurance Runners. Studies in Sports Science and Physical Education, 4(1), 8-18.*

[18] *Huang, R. (2026). Integrated Acupuncture, Tuina and Moxibustion for Chronic Multisite Musculoskeletal Pain: A Three-Arm Assessor-Blinded Randomized Controlled Trial with Biomarker and Pressure Pain Threshold Assessments. Journal of Innovations in Medical Research, 5(2), 16-22.*

[19] *Wei, M. (2026). A Review of the Physiological Effects of Foam Rolling on Skeletal Muscle. Journal of Innovations in Medical Research, 5(2), 9-15.*

[20] *Yuan, D. (2025). The Impact of Traditional Chinese Non-Pharmacological Therapies on Blood Pressure Control in Community Patients with Hypertension. Current Research in Medical Sciences, 4(5), 8-15.*

[21] *Yuan, D. (2025). Research on the Intervention of Cancer Patients' Anxiety by the Concept of "Harmonization of Body and Mind" in Traditional Chinese Medicine. Journal of Innovations in Medical Research, 4(5), 19-25.*

[22] *Chen, Y. (2026). Standardized Traditional Chinese Medicine External Therapy for Chronic Soft Tissue Pain: A Multicenter Observational Study. Current Research in Medical Sciences, 5(3), 36-45.*

[23] *Chen, Y. (2026). Construction and Application of Standard System for Traditional External Therapy in Traditional Chinese Medicine Health Preservation. Journal of Innovations in Medical Research, 5(2), 23-31.*

[24] *Yao, C. (2026). Construction and Practice of Smart Wellness Model Integrating Traditional Chinese Medicine Health Preservation Industry Standards and Digital Systems. Insights in Social Science, 4(2), 76-86.*

[25] *Mingyang, W. (2026). Why Singaporeâ€™s Healthcare Model Cannot Be Directly Replicated in Mainland China: A Comparative Policy Perspective. Insights in Social Science, 4(2), 24-31.*