ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

Review Article

Cross-Setting Rehabilitation Data Integration: An Edge AI and Knowledge Graph Framework

Daniel R. Mercer[1,*]

[1] Department of Computer Science, StanfordStanford University, 94305, USA

* Corresponding author: Daniel R. Mercer

Email: DanielRMercer@outlook.com

Abstract: Rehabilitation services are delivered across home, school, and clinical settings, creating fundamental barriers to achieving personalized, continuous care. Data generated from wearable devices, electronic health records, and clinical assessments are heterogeneous and siloed, limiting the potential for AI-driven insights. This study proposes a four-layer framework combining edge AI and knowledge graphs to enable the integration of rehabilitation data across these diverse settings. Drawing on recent advances in rehabilitation knowledge discovery, data standardization, and knowledge-graph-enhanced personalization, the framework addresses three core challenges: the semantic unification of multi-source physiological and behavioral data, context-aware reasoning for personalized intervention design, and clinical decision support with data provenance. The framework encompasses processes such as wearable data acquisition, provenance-enabled knowledge graph integration, and role-based clinical review. This conceptual architecture lays the foundation for future empirical validation across various rehabilitation populations and care scenarios.

Keywords: *Rehabilitation data integration; Edge artificial intelligence; Knowledge graph; Multi-source data harmonization; Personalized rehabilitation;*

1. Introduction

Rehabilitation services are increasingly delivered across multiple settings-home, school, clinic, and community, yet data generated in each setting remain fragmented, impeding continuity of care and limiting the potential for personalized interventions. Stroke and traumatic brain injury survivors, for example, may receive acute rehabilitation in hospital settings, transition to outpatient therapy, and continue exercises at home with wearable sensors, but data from each phase typically reside in separate systems with incompatible formats and missing contextual information. This fragmentation is not merely a technical inconvenience but a fundamental barrier to understanding what works for which patient, the central question of personalized rehabilitation [16][26].

The rapid advancement of artificial intelligence has opened new possibilities for rehabilitation science, establishing "AI for Science" as a fifth scientific paradigm. However, existing AI applications in rehabilitation have largely focused on isolated tasks such as intelligent sensing or rehabilitation robotics, rather than addressing the systemic challenge of integrating heterogeneous data across the care continuum. A substantial gap persists: the lack of a systematic, AI-driven framework for rehabilitation knowledge discovery and cross-setting data integration [27]. The work of Xu (2025c) on home-school collaborative progress tracking for special children's rehabilitation has demonstrated the practical challenges of cross-setting coordination, where progress tracking across environments required systematic coordination that existing systems did not adequately support [16].

ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

Digital and gamified interventions may improve engagement for some populations, though the effect is population- and intervention-specific. Xu (2025a) reported changes in engagement and language-related measures in a 12-week study of 60 children aged 4–7 with language delays under a tiered gamification protocol, supporting further investigation of engagement-oriented design in that cohort [17]. Similarly, Xu conducted a randomized controlled trial with 120 children with autism spectrum disorder, evaluating the impact of Montessori sensory training on bilingual language expression, demonstrating the importance of controlled designs while also illustrating the challenges of generalizing findings from specific populations to broader rehabilitation contexts [18].

Rehabilitation outcome measurement also requires carefully defined protocols. Fan et al. (2021) published a prospective randomized, double-blind protocol planning to recruit 100 post-stroke participants for active or sham radial extracorporeal shock-wave therapy, providing a methodological template for rigorous rehabilitation research that could inform the design of evaluations for cross-setting data integration systems [5]. A later randomized study by Fan et al. (2025) compared bilateral exoskeleton-assisted and conventional walking training in 25 patients after stroke, demonstrating how neurological, functional, gait, and electromyographic outcomes can be collected together, while also illustrating the need for larger and multicenter validation—a limitation that cross-setting data integration frameworks might address [6].

Wrist-worn devices have been used to analyze sub-activities and physical demands in various populations. Wang et al. (2025) demonstrated the feasibility of wrist-based measurement in construction workers, supporting the potential application of similar devices in rehabilitation contexts. However, this study also illustrates that the same labels, windows, thresholds, and models validated in one population cannot be assumed to transfer to others, such as children with neurodevelopmental conditions or adults after stroke [14]. This observation underscores the importance of provenance-aware data integration that preserves information about the population and context in which models were validated.

Recent research has begun to address data integration challenges from various angles. Studies on rehabilitation knowledge graphs have demonstrated the feasibility of extracting structured disease-symptom relationships from electronic health records. Work on KG-LLM collaboration for rehabilitation assistive device design has shown how knowledge graphs can support personalized conceptual design. Data harmonization efforts have established proof-of-concept for integrating outcome measures across multiple studies. These advances, while significant, remain fragmented and have not been synthesized into a coherent cross-setting data integration framework [26].

The integration of AI into educational and rehabilitation contexts requires systematic training of practitioners. Huang (2025) developed and evaluated a teacher training program in AI technology, demonstrating that competency development is essential for effective AI adoption—a finding with implications for rehabilitation settings where clinicians and therapists must be equipped to interpret AI-generated insights from cross-setting data integration systems [8].

1.1 Problem Statement

Despite growing recognition of the potential value of AI in rehabilitation, significant challenges persist in achieving cross-setting data integration. These challenges can be organized into four interrelated categories.

First, data heterogeneity presents a formidable obstacle. Rehabilitation data span physiological signals from wearables, clinical assessments, patient-reported outcomes, behavioral observations, and environmental context. Each data type has different formats, temporal resolutions, measurement properties, and semantic meanings, making direct integration without substantial preprocessing and semantic mapping effectively impossible [14][26].

Second, the semantic gap between numerical data and clinical knowledge creates a fundamental barrier to reasoning. Raw physiological signals lack the semantic information required for integration with clinical knowledge about diagnoses, interventions, and outcomes. A wrist-worn accelerometer may record movement patterns that could represent prescribed exercises, ordinary daily activities, caregiver assistance, or sensor displacement but the numerical data alone cannot distinguish among these interpretations without additional contextual information that is rarely captured systematically [16].

ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

Third, provenance deficit undermines clinical interpretability and trust. When data are aggregated across settings, the origin, processing history, and confidence of each derived observation are often lost. This loss of provenance makes it difficult to investigate errors, assess the validity of derived measures, or understand the clinical significance of observed changes. Without provenance, clinicians cannot determine whether a change in a derived measure reflects genuine clinical improvement, differences in measurement conditions, or artifacts of data processing [8][27].

Fourth, interoperability barriers at both technical and semantic levels impede integration. Existing rehabilitation data standards remain insufficiently comprehensive for cross-setting integration, with fragmented approaches to device calibration, outcome measurement, metadata capture, and data exchange. The complexity of rehabilitation data integration demands interdisciplinary collaboration across clinical, technical, and educational domains—a challenge that has been examined in educational informatization contexts where effective integration requires not only technological solutions but also systematic attention to collaborative practices and shared understanding across disciplines [9][15].

1.2 Research Objectives and Significance

The primary objective of this study is to propose a four-layer conceptual framework for cross-setting rehabilitation data integration that combines edge AI for local quality control and inference with knowledge graphs for semantic integration and provenance tracking. This framework is intended to address the gap between isolated AI applications and systematic rehabilitation knowledge discovery by providing an architectural blueprint that can guide implementation and evaluation.

The significance of this study extends beyond its immediate architectural proposal. From a theoretical perspective, it synthesizes insights from rehabilitation science, AI, knowledge engineering, and health informatics to address a challenge that has been recognized across these fields but has not been systematically addressed. From a practical perspective, it provides a foundation for developing systems that can support continuous, cross-setting rehabilitation care. The framework also contributes to the broader effort to establish AI for rehabilitation as a systematic approach to knowledge discovery, complementing existing frameworks' focus on literature-based knowledge extraction with a focus on data integration and reasoning [27][28].

Independent rules prohibit critical tasks on restricted nodes, enforce mandatory diversity, block unapproved degradation, and impose a decision time budget. If the adaptive policy fails to return a valid action, a deterministic fallback uses the most trusted feasible resource. Rule activations are logged so that apparent performance can be separated from safety intervention.

Reproducible reporting discloses hardware, operating system, monitor configuration, workload trace, random seed, attack and fault injection, ambient conditions, and every failed run. Cross platform replication remains necessary before broad generalization.

2. Literature Review

2.1 AI-Driven Rehabilitation Knowledge Discovery

Recent studies have demonstrated the potential of AI for systematic rehabilitation knowledge discovery. The RehabKD framework, integrating LLM-KG-LDACX engine with sentiment analysis, manifold clustering, and complex network analysis, represents a significant advance in this direction. Processing 699 Web of Science articles, RehabKD constructed a three-layer rehabilitation knowledge graph comprising fundamental, technological, and clinical domains. The framework's use of multiple AI models including LLMs for semantic analysis, knowledge graphs for structural representation, and LDACX for topic modeling demonstrates the value of multi-model orchestration for comprehensive knowledge discovery [27].

The RehabKD analysis revealed that AI-rehabilitation research has grown exponentially since 2018, with key topics evolving from algorithm optimization through multi-technology coordination to ethical considerations and deployment. This evolutionary pattern suggests that the field is moving toward more integrated, system-level approaches, though the transition remains incomplete. The RehabKD framework also revealed the dominance of certain research themes and the relative neglect of others, highlighting gaps in the current research landscape [27].

ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

The methodological approach of RehabKD combining sentiment analysis to assess research attitudes, manifold clustering to identify research communities, and complex network analysis to map intellectual structures offers a model for systematic knowledge synthesis that could be extended to other domains. However, it must be acknowledged that RehabKD focuses primarily on literature-based knowledge extraction rather than the integration of clinical and sensor data that is the focus of the present study [26][27].

2.2 Data Harmonization and Standardization

Data harmonization across rehabilitation studies requires systematic handling of heterogeneity in outcome measures, intervention characteristics, and population profiles. The NeuroRehab Common Data Elements initiative, involving 120 experts across 12 domains, represents the most comprehensive effort to establish standardized data elements for neurorehabilitation research. The initiative identified core and supplemental measures for domains including demographics, clinical status, functional outcomes, and participant-reported outcomes, providing a foundation for data integration across studies [4].

However, significant gaps remain in the current data standards. The NeuroRehab CDEs, while comprehensive, do not fully address the diversity of data types generated by wearable sensors and other digital health technologies. Similarly, the standards for intervention reporting do not adequately capture the complexity of rehabilitation interventions that may adapt over time based on patient progress and context [4][15].

Unified modeling language frameworks for telerehabilitation have proposed patient-centered models integrating medical, functional, psychosocial, and organizational data. These models emphasize the importance of representing not just clinical data but also contextual factors such as setting, supervision, and patient preferences. However, these frameworks have been primarily conceptual rather than implemented, and their applicability to cross-setting data integration remains to be demonstrated.

Multi-format data integration is a challenge across domains. Zhang (2026) examined the construction of a multi-format integrated management system for small, medium, and micro-sized enterprises, addressing challenges of heterogeneous data sources and formats that parallel those encountered in cross-setting rehabilitation data integration. The principles of integrating diverse data types while preserving their original meaning and context are relevant to rehabilitation data systems [15].

The effectiveness of intelligent algorithms in resource allocation and matching has been demonstrated across domains. Chen (2025) analyzed the technical application and effectiveness of intelligent algorithms for regional logistics resource matching, providing evidence that algorithmic approaches can improve the efficiency of resource allocation, a principle applicable to the allocation of rehabilitation sensors and clinical resources across settings [2].

2.3 Knowledge Graphs for Rehabilitation

The application of knowledge graphs to rehabilitation has emerged as a promising approach for integrating heterogeneous data and supporting reasoning. Wang and colleagues proposed a knowledge graph-driven framework for personalized rehabilitation design, constructing a rehabilitation product design knowledge graph (RPD-KG) that integrates physiological, behavioral, and product parameter data. The RPD-KG used fuzzy c-means clustering to convert numerical physiological data into semantic labels and enhanced RippleNet with TransD embeddings to capture latent relationships between user data and product recommendations. The approach achieved recommendation accuracy of 0.7949 and AUC of 0.8587, demonstrating feasibility while also revealing opportunities for improvement.

Pan and colleagues developed a context-aware KG-LLM collaborated approach (CACKL) for lower-limb rehabilitation assistive devices. This approach used fine-tuned LLMs to extract fine-grained requirement-function mappings from user descriptions and CoT-guided KG-LLM collaboration to generate design solutions that incorporate actual design knowledge. The approach demonstrated advantages in generating personalized design solutions but also highlighted challenges in integrating clinical knowledge with engineering design [12].

ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

The NeuroRehab Common Data Elements initiative, while primarily focused on research data standardization, also provides a foundation for knowledge graph construction by defining core concepts and their relationships. However, the CDEs are designed for research rather than clinical care, and their translation to knowledge graph representation requires additional work [4].

Resource optimization across multiple settings requires systematic approaches to allocation and scheduling. Liu (2026) proposed a submodular constrained framework for heterogeneous resource slot optimization in recommendation landscapes, offering methods that could be adapted to optimize sensor deployment and data collection across settings in cross-setting rehabilitation systems. The principles of optimizing resource allocation while accounting for cross-space spillover effects have direct relevance to rehabilitation contexts where actions in one setting affect outcomes in others [10].

2.4 Research Gap

The literature review reveals a significant gap that this study aims to address. While there is substantial research on individual components, AI-driven rehabilitation knowledge discovery, data harmonization, knowledge graphs for rehabilitation, there is no integrated framework that coordinates edge AI inference, knowledge graph integration, and cross-setting data provenance for rehabilitation. The RehabKD framework focuses on literature-based knowledge extraction rather than clinical data integration. The RPD-KG and CACKL approaches demonstrate the value of knowledge graphs for personalized design but do not address the cross-setting data integration challenge. Data harmonization efforts provide standards but lack mechanisms for provenance tracking and semantic integration [12][27].

The present study addresses this gap by proposing a four-layer framework that integrates edge AI for local quality control and inference with knowledge graphs for semantic integration and provenance tracking. Building on the foundational contributions of RehabKD, RPD-KG, and other relevant work, the framework is designed to support both clinical decision-making and rehabilitation knowledge discovery across settings [10][15][27].

3. Framework Architecture

3.1 Overview

The proposed framework comprises four interconnected layers that, considered together, provide a comprehensive approach to cross-setting rehabilitation data integration. The four layers are: (1) Wearable Acquisition Layer, (2) Local Quality Control and Inference Layer, (3) Provenance-Aware Knowledge Graph Layer, and (4) Role-Based Clinical Review Layer. Each layer addresses a specific set of challenges identified in the problem analysis and literature review, and the layers interact through defined interfaces that preserve data provenance while enabling semantic integration.

The framework is designed to support a range of rehabilitation applications, including continuous monitoring of exercise performance, tracking of functional outcomes across settings, personalization of intervention recommendations, and clinical review of patient progress. It is intended to be adaptable to different rehabilitation populations (e.g., stroke, traumatic brain injury, spinal cord injury) and care settings (e.g., home, clinic, community) through configuration rather than redesign.

Table 1 provides a comparative summary of existing rehabilitation data integration approaches and their limitations relative to the proposed framework. As shown in the table, existing approaches address different subsets of the integration challenge—RehabKD focuses on literature-based knowledge discovery, RPD-KG and CACKL on personalized design, NeuroRehab CDEs on data standardization, and UML frameworks on conceptual modeling—but none provides a comprehensive solution that simultaneously addresses cross-setting data integration, provenance tracking, and semantic reasoning. The proposed framework is distinguished by its integration of edge AI for local inference with knowledge graph-based semantic integration and provenance tracking.

TABLE 1. COMPARISON OF EXISTING REHABILITATION DATA INTEGRATION APPROACHES AND THEIR LIMITATIONS

Approach	Key Contribution	Data Sources Addressed	Provenance Tracking	Cross-Setting Support	Semantic Integration	Key Limitation

Approach	Key Contribution	Data Sources Addressed	Provenance Tracking	Cross-Setting Support	Semantic Integration	Key Limitation
RehabKD (2025) [27]	Multi-model knowledge graph construction from 699 articles	Literature (Web of Science)	No	No	Yes	Literature-based only; does not integrate clinical or sensor data
RPD-KG	Personalized rehabilitation design via KG-enhanced RippleNet	Physiological, behavioral, product parameters	Partial	No	Yes	Focuses on product design, not cross-setting clinical data integration
CACKL (Pan et al., 2025) [12]	KG-LLM collaboration for assistive device design	User requirements, functional specifications	No	No	Partial	Design-focused; limited clinical validation
NeuroRehab CDEs (Carlozzi et al., 2025) [4]	Standardized data elements across 12 domains	Clinical assessments, outcomes	No	Yes	Partial	Standards only; no implementation mechanism
UML Telerehabilitation Framework	Patient-centered model integrating medical and biopsychosocial data	Medical, functional, psychosocial, organizational	No	Yes	Yes	Conceptual; not implemented or validated
Proposed Framework	Four-layer edge AI + knowledge graph integration	Wearable, clinical, patient-reported, environmental	Yes (PROV-O)	Yes	Yes	Conceptual; requires empirical validation

3.2 Wearable Acquisition Layer

The Wearable Acquisition Layer collects multi-modal data from diverse sources, including wearable devices (accelerometry, gyroscopes, photoplethysmography, pressure sensors), electronic health records (clinical assessments, diagnoses, medications), patient-reported outcomes (questionnaires, diaries), and device-interaction logs. Each raw observation is stored with minimum metadata including device model, firmware version, sensor type, collection setting, collection time, and participant identifier.

The requirement for comprehensive metadata collection emerged from the recognition that data from different settings and devices are not directly comparable without explicit contextual information. A movement pattern measured by a wrist-worn accelerometer may have different meanings depending on whether the wearer is engaged in prescribed exercises at a clinic, performing daily activities at home, or receiving assistance from a caregiver. Without metadata capturing these contextual factors, integration across settings is likely to produce misleading results [14][16].

The acquisition layer also addresses the challenge of data heterogeneity by providing a common interface for diverse data sources. This interface includes data ingestion functions that handle different data formats and provide semantic mapping to a common data model, and metadata validation functions that ensure required metadata are present and valid. The principles of multi-format data integration examined by Zhang (2026) in enterprise contexts offer useful analogies for addressing these challenges in rehabilitation settings [15].

Multi-agent coordination frameworks offer insights for managing distributed rehabilitation devices. Luo et al. (2023) developed a multi-agent deep reinforcement learning approach for fleet rebalancing in shared e-mobility systems, demonstrating how distributed agents can coordinate resource allocation, methodologically relevant for coordinating wearable devices and edge nodes across rehabilitation settings [29].

3.3 Local Quality Control and Inference Layer

The Local Quality Control and Inference Layer performs operations that are time-sensitive and context-specific at the edge, near the data source. This design choice is motivated by the recognition that rehabilitation data are often generated in settings with limited connectivity and that timely feedback is essential for effective intervention. The operations performed by this layer include

ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

timestamp validation to identify gaps or overlaps in data collection; saturation detection to identify periods where sensor readings exceed valid ranges; non-wear detection to identify periods when devices were not being worn; implausible value filtering to remove readings inconsistent with physiological constraints; placement change detection to identify when devices were moved to different body locations; and confidence calibration to provide uncertainty estimates for derived measures.

Edge AI for rehabilitation must operate under energy and latency constraints. Hao (2026) examined energy-efficient multi-core task scheduling for real-time edge AI systems, demonstrating that energy and timing must be treated as coupled variables, a consideration critical for wearable rehabilitation devices that must balance inference quality with battery life [19]. The scheduler's own computation must be considered as part of the real-time cost. Hao (2026) developed a low-overhead scheduling approach for real-time AI workloads, demonstrating that scheduling decisions must account for the computational cost of the scheduler itself, particularly relevant when security alerts, health indicators, and shared risk relationships enlarge the decision state [20].

Dynamic task prioritization is essential for balancing competing demands in edge AI systems. Hao (2026) examined dynamic task prioritization for edge AI in smart cities, balancing latency and energy efficiency, considerations that are equally relevant in rehabilitation settings where time-sensitive alerts must be prioritized over routine data processing [21]. The layer also performs local inference for tasks that can be reliably performed at the edge. Examples include activity classification to identify exercise type; movement quality assessment to evaluate adherence to prescribed exercises; and early warning generation to identify patterns suggesting deterioration or risk. The outputs of local inference are stored with confidence scores and calibration indicators that reflect the reliability of the inference given available data.

Real-time adaptive scheduling is essential for handling variable data loads and processing demands. Huang (2025) developed a real-time adaptive dispatch algorithm for dynamic vehicle routing with time-varying demand, offering methodological insights for scheduling edge inference tasks in resource-constrained rehabilitation settings [7].

Fault tolerance is essential for edge AI systems in critical applications. Hao (2025) developed fault-tolerant real-time scheduling for edge AI in critical infrastructure, embedding primary-backup execution and degradation strategies—mechanisms that could inform the design of robust rehabilitation data systems that continue to function under partial failures [22]. Task migration in edge systems can disturb cache locality and introduce timing jitter. Hao (2025) examined task affinity-aware scheduling for multi-core edge devices, demonstrating that migration decisions must consider affinity relationships, a significant constraint for rehabilitation data systems where moving a task away from a suspicious or failing node is not free [23].

Dynamic risk assessment and prioritization frameworks have been developed in diverse domains. Ren (2025) applied reinforcement learning for prioritizing anti-money laundering case reviews based on dynamic risk assessment, illustrating how evolving evidence can inform prioritization, a principle that extends to rehabilitation contexts where patients with deteriorating status should be prioritized for clinical review [24]. Real-time threat identification frameworks from other domains offer methodological insights for rehabilitation data verification. Ren (2025) developed a real-time threat identification system for financial API attacks under a federated learning framework, demonstrating distributed evidence processing that could inform the verification of rehabilitation data integrity across settings [25].

3.4 Provenance-Aware Knowledge Graph Layer

The Provenance-Aware Knowledge Graph Layer is the core semantic integration component of the framework. It represents relationships among patients, observations, interventions, and outcomes using a knowledge graph that distinguishes among different types of evidence. The knowledge graph distinguishes "observed," "derived," "reported," and "clinically confirmed" facts, ensuring that data from different sources and levels of validation are not inappropriately equated.

Observed facts correspond to raw data collected from sensors or direct observation. Derived facts correspond to inferences made from observed data, such as activity classifications or quality assessments. Reported facts correspond to information provided by patients, caregivers, or other informants. Clinically confirmed facts correspond to observations or assessments that have been reviewed and validated by a clinician. This distinction among evidence types is essential for clinical decision-making, as a model-generated classification should not be treated as equivalent to a clinician-confirmed assessment.

ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

Provenance is represented using the W3C PROV-O concepts of Entity, Activity, and Agent, together with relations such as wasGeneratedBy, wasAttributedTo, and wasDerivedFrom. For each derived observation, the provenance record captures the source signal, preprocessing activity, model version, device, setting, and responsible agent. This provenance information enables investigation of errors by identifying the source and processing chain for any observation; assessment of validity by determining whether derived measures are based on appropriate data and methods; and clinical interpretation by providing context for observed changes.

The knowledge graph construction process builds on established methodologies such as the RPD-KG approach of Wang and colleagues. Following this approach, physiological signal features are extracted and converted to semantic labels using fuzzy c-means clustering. The graph integrates wearable data, clinical assessments, intervention records, and outcomes. However, the present framework extends the RPD-KG approach by adding provenance tracking and distinguishing among evidence types.

3.5 Role-Based Clinical Review Layer

The Role-Based Clinical Review Layer provides different views for different stakeholders, supporting clinical decision-making while preserving appropriate boundaries of access and responsibility. The layer recognizes that different stakeholders have different information needs and different levels of authority to act on information.

For participants, the view provides feedback on exercise performance, progress tracking, and personalized recommendations. For caregivers, the view provides information about task completion, participation, and observed issues. For educators, the view provides information about school-based activities and performance. For clinicians, the view provides comprehensive data on progress, trends, and quality flags, with the ability to review and confirm observations. For researchers, the view provides access to de-identified data and derived measures for analysis.

Cross-setting rehabilitation data systems must maintain stability and resilience under varying loads and conditions. Liu (2026) developed a predictive multi-dimensional safeguards architecture for billion-scale concurrent platforms, demonstrating principles of cascading resilience that could inform the design of robust, fault-tolerant rehabilitation data integration systems [11].

The layer also implements access controls based on role and purpose, ensuring that users have access only to information that is appropriate for their role. This design addresses both privacy and clinical governance requirements by ensuring that sensitive data are disclosed only to those who need them. The principles of personalized learning have parallels in personalized rehabilitation. Huang (2025) explored the application of AI to personalized learning platforms in higher education, demonstrating how AI can adapt content and pacing to individual learner characteristics, work that offers methodological insights for developing personalized rehabilitation recommendation systems that adapt interventions based on individual patient data and progress [13].

The four layers interact through defined interfaces that preserve data provenance while enabling semantic integration. The Wearable Acquisition Layer provides raw observations with metadata to the Local Quality Control and Inference Layer, which provides quality-controlled data and local inferences to the Provenance-Aware Knowledge Graph Layer. The Knowledge Graph Layer integrates data from multiple sources and provides query and reasoning services to the Role-Based Clinical Review Layer.

These interactions are designed to support both online processing (real-time) and offline processing (retrospective). Online processing supports immediate feedback and early warning, while offline processing supports comprehensive analysis and knowledge discovery. The separation between online and offline processing addresses the temporal constraints of rehabilitation settings where some decisions require immediate action while others can be made with more comprehensive information.

4. Implementation Considerations

4.1 Knowledge Graph Construction

Building on the RPD-KG methodology, the knowledge graph construction process involves several steps. First, physiological signal features are extracted and converted to semantic labels using fuzzy c-means clustering. This step transforms numerical data

ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

into meaningful categories that can be represented in the knowledge graph. Second, relationships among entities are identified and represented using the PROV-O ontology, ensuring that provenance information is captured at the time of relationship creation.

The construction process also involves defining the schema for the knowledge graph. The schema includes entities such as Patient, Observation, Intervention, Outcome, Setting, and Device; relationships such as hasObservation, hasIntervention, hasOutcome, wasObservedIn, wasMeasuredBy; and attributes such as timestamp, confidence, and provenance.

A significant challenge in knowledge graph construction is the integration of data from different sources that may use different terminology. This challenge is addressed by using standard ontologies where available (such as the NeuroRehab CDEs) and mapping local terms to standard concepts. However, complete semantic harmonization is rarely possible, and it must be acknowledged that some degree of uncertainty in integration is unavoidable. The framework addresses this uncertainty by capturing provenance information that allows the basis for integration to be assessed [4].

4.2 Personalized Reasoning

Following the enhanced RippleNet approach, personalized reasoning is implemented using TransD embeddings to capture latent relationships between user data and intervention recommendations. This approach leverages the knowledge graph structure to provide recommendations that are informed by both explicit knowledge (as represented in the graph) and implicit relationships (as captured by embeddings).

The reasoning process considers multiple factors including patient characteristics (diagnosis, functional status, prior interventions), setting (home, clinic, community), available resources (devices, support), and preferences. The output is a set of personalized recommendations with associated rationale and confidence scores.

Structure-aware representation of tasks and heterogeneous cores can improve scheduling decisions. Hao (2026) developed a structure-aware deep reinforcement learning approach for latency-minimal scheduling on heterogeneous cores, representing tasks and resources relationally rather than as unrelated state entries, an approach that is particularly valuable when nodes share communication paths, software images, or attack surfaces [28]. Quality-of-service-aware resource allocation can incorporate predictive performance and application-specific costs. Hao et al. (2026) demonstrated this in a telemedicine context using a heart-failure prediction task, showing that resource decisions should reflect asymmetric application consequences rather than processor use alone, a principle applicable to rehabilitation systems where the cost of misclassification varies by clinical context [26].

However, it must be acknowledged that the reasoning process is only as good as the knowledge graph from which it derives. Gaps or errors in the knowledge graph will propagate to the reasoning output, and the models used for reasoning may not generalize well to new patients or settings. These limitations suggest that the reasoning system should be treated as a decision support tool rather than a replacement for clinical judgment.

4.3 Integration with Clinical Workflows

The framework is designed to integrate with existing clinical workflows rather than requiring new workflows. This design choice is motivated by the recognition that clinical adoption is more likely when technology supports existing practices rather than requiring transformation. The framework supports common clinical tasks such as progress review, treatment planning, and documentation.

Integration points include electronic health records (for data exchange and documentation), telehealth systems (for remote monitoring and consultation), and clinical decision support systems (for providing recommendations). The framework uses standard data exchange formats and APIs to facilitate integration.

4.4 Evaluation Framework

ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

A staged evaluation framework separates engineering performance from clinical effectiveness. This separation is essential because engineering metrics such as latency and accuracy do not necessarily predict clinical utility. The evaluation framework includes four stages, as summarized in Table 2.

Stage A involves bench and simulation testing, where technical outcomes such as latency, energy consumption, memory use, and behavior under data loss and corruption are assessed. Sustainability optimization across multiple sites and systems is a challenge shared across domains. Nie (2025) addressed sustainability optimization in North American cross-border logistics networks, offering methodological approaches to optimizing performance across distributed settings that could inform the design of sustainable cross-setting rehabilitation data systems [1].

Stage B involves retrospective validation, where the system is evaluated on previously collected data to assess sensitivity, specificity, and calibration. Stage C involves prospective feasibility studies, where usability, acceptability, and safety are assessed with human participants. Stage D involves controlled clinical evaluation, where clinical effectiveness is assessed through trials with appropriate comparison groups.

The sustainability of digital transformation in rehabilitation requires attention to long-term mechanisms. Zhang (2026) examined the long-term mechanisms of digital transformation for small and medium-sized enterprises, highlighting factors that sustain adoption over time, insights that are relevant to the sustained implementation of cross-setting rehabilitation data integration systems [3].

TABLE 2. STAGED EVALUATION FRAMEWORK FOR CROSS-SETTING REHABILITATION DATA INTEGRATION SYSTEMS

Stage	Objective	Key Metrics	Data Source	Success Criterion	Primary Limitation
A: Bench & Simulation	Assess technical feasibility	End-to-end latency, Energy per inference, Memory usage, Deadline-miss rate, Provenance completeness	Recorded/simulated sensor streams	Meets prespecified technical thresholds	Does not assess clinical utility or usability
B: Retrospective Validation	Validate model performance on existing data	Sensitivity, Specificity, F1 score, Calibration error, Abstention coverage	Previously collected rehabilitation data	Exceeds baseline performance; calibration within acceptable range	Data may not represent target population or setting
C: Prospective Feasibility	Assess usability and safety in real-world settings	Recruitment rate, Device wear time, Signal quality adequacy, Model abstention rate, User satisfaction, Protocol deviations	Human participants in target settings	Feasibility metrics meet prespecified targets; safety events within acceptable limits	Sample size limited; not powered for efficacy
D: Controlled Clinical Evaluation	Assess clinical effectiveness	Primary clinical outcome, Quality of life, Healthcare utilization, Adverse events	Randomized controlled trial or rigorous quasi-experimental design	Clinically meaningful and statistically significant improvement	Resource-intensive; requires regulatory approval

However, it must be acknowledged that staging evaluation in this way is time-consuming and resource-intensive. Not all systems will proceed through all stages, and for some applications the clinical effectiveness stage may not be appropriate. The framework should be adapted to the specific application context and regulatory requirements.

5. Conclusion

This chapter does not simply rehearse the preceding findings but seeks to elevate the theoretical height and practical significance of the conclusions drawn. The four-layer framework proposed in this study offers a way forward from the fragmented,

ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

technology-centric approaches that have characterized rehabilitation data integration to date. By recognizing the interdependence of data acquisition, quality control, semantic integration, and clinical review—and the mediating role of provenance in bridging these dimensions—the framework provides a more adequate foundation for understanding and facilitating cross-setting rehabilitation data integration. From a theoretical standpoint, the framework extends existing approaches to rehabilitation knowledge discovery by explicitly addressing the cross-setting data integration challenge that has been recognized but not systematically addressed in prior work. The identification of provenance tracking as a core design principle challenges the implicit assumption that data from different settings can be integrated without explicit contextual information, suggesting that the semantic interpretation of rehabilitation data is fundamentally context-dependent. The framework's emphasis on distinguishing among evidence types, observed, derived, reported, and clinically confirmed, offers a more nuanced approach to representing clinical knowledge than the flat representations common in current systems. The practical implications are equally significant, offering guidance for system developers on architectural design, for clinicians on data interpretation, and for researchers on evaluation design. Several limitations of the study warrant acknowledgment and point toward promising directions for future research. The framework remains conceptual and requires empirical validation across diverse rehabilitation populations and settings. The specific implementation of each layer, particularly the knowledge graph construction and reasoning components, requires further specification and testing. The evaluation framework, while comprehensive in scope, requires adaptation to specific application contexts and regulatory requirements. Considering the above factors, this leads us to further thinking about several directions: how can the framework be adapted to different rehabilitation populations and care settings? What are the optimal approaches to knowledge graph construction and maintenance in clinical practice? How can privacy and consent be managed in cross-setting data sharing? How can the framework support continuous learning and adaptation as new evidence emerges? Ultimately, the cross-setting rehabilitation data integration framework represents not a definitive solution but a structured approach to thinking about the complex challenge of integrating rehabilitation data across settings, a way of understanding the multi-dimensional nature of rehabilitation knowledge and a foundation upon which future research can build to refine and extend our understanding of this critical process.

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] *Nie, X. (2025). Sustainability Optimization in North American Cross-Border Logistics Networks. Journal of World Economy, 4(6), 66-73.*

[2] *Chen, Y. (2025). Analysis of the technical application and effectiveness of intelligent algorithms empowering regional logistics resource matching. Engineering Frontiers, 1(3).*

[3] *Zhang, N. (2026). Research on the long-term mechanism of digital transformation for small and medium-sized enterprises. Journal of Intelligence and Engineering Technology, 1(2), 19-28.*

[4] *Carlozzi, N. E., Mendoza-Puccini, M. C., Marden, S., Backus, D., Bambrick, L., Baum, C., ... & Heinemann, A. W. (2025). Common data elements for rehabilitation research in neurologic disorders (NeuroRehab CDEs). Archives of physical medicine and rehabilitation, 106(7), 981-988.*

ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

[5] *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.*

[6] *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).*

[7] *Huang, S. (2025). Real-time adaptive dispatch algorithm for dynamic vehicle routing with time-varying demand. Academic Journal of Computing & Information Science, 8(9), 108-118.*

[8] *Huang, H. (2025). Development and evaluation of a teacher training program in artificial intelligence technology. Journal of Advanced Research in Education, 4(1), 23-31.*

[9] *Huang, H. (2025). Interdisciplinary collaboration in educational informatization innovation: Importance and practice. Research and Advances in Education, 4(1), 38-46.*

[10] *Liu, Y. (2026). Heterogeneous resource slot optimization in multi-dimensional recommendation landscapes: A submodular constrained framework with cross-space spillover effects. Journal of Progress in Engineering and Physical Science, 5(1), 26-31.*

[11] *Liu, Y. (2026). Cascading resilience through predictive multi-dimensional safeguards: System stability architecture for billion-scale concurrent platforms. Innovation in Science and Technology, 5(1), 35-45.*

[12] *Pan, X., Zhuang, W., Wen, S., Yu, W., Bao, J., & Li, X. (2025). A context-aware KG-LLM collaborated conceptual design approach for personalized products: A case in lower limbs rehabilitation assistive devices. Advanced Engineering Informatics, 66, 103422.*

[13] *Huang, H. (2025). AI meets higher education: Applying artificial intelligence to personalized learning platforms. Innovation in Science and Technology, 4(2), 43-50.*

[14] *Wang, J., Seo, J., Gong, Y., Siu, M. F. F., 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), 516-531.*

[15] *Zhang, N. (2026). Research on the construction of a multi-format integrated management system for small, medium and micro-sized enterprises. Frontiers in Management Science, 5(2), 9-18.*

[16] *Xu, T. (2025). A study on the design and application of a home-school collaborative progress tracking system for special children's rehabilitation. Research and Advances in Education, 4(8), 24-29.*

[17] *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.*

[18] *Xu, T. (2025). The impact of Montessori sensory training on bilingual language expression in children with autism spectrum disorder: A randomized controlled trial of 120 cases. Journal of Innovations in Medical Research, 4(6), 29-33.*

[19] *Hao, Z. (2026). Energy efficient multi core task scheduling for real time edge AI systems: A latency aware approach. International Journal of Advance in Applied Science Research, 5(3), 1-14.*

[20] *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.*

[21] *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.*

[22] *Hao, Z. (2025). Fault-tolerant real-time scheduling for edge AI in US critical infrastructure. Engineering Frontiers, 1(4).*

[23] *Hao, Z. (2025). Task affinity-aware scheduling for multi-core edge devices in autonomous vehicles. Engineering Frontiers, 1(2).*

[24] *Ren, L. (2025). Reinforcement learning for prioritizing anti-money laundering case reviews based on dynamic risk assessment. Journal of Economic Theory and Business Management, 2(5), 1-6.*

[25] *Ren, L. (2025). Real-time threat identification systems for financial API attacks under federated learning framework. Academic Journal of Business & Management, 7(10), 65-71.*

[26] *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.*

[27] *Zhang, W., Xing, Z., & Chen, S. (2025). RehabKD With LLM-KG-LDACX: Orchestrating AI Models for Rehabilitation Knowledge Discovery. IEEE Access, 13, 215003-215030.*

ISSN 3107-7657 (Print) ISSN 3107-7657 (Online)

[28] *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.*

[29] *Luo, M., Du, B., Zhang, W., Song, T., Li, K., Zhu, H., ... & Wen, H. (2023). Fleet rebalancing for expanding shared e-mobility systems: A multi-agent deep reinforcement learning approach. IEEE Transactions on Intelligent Transportation Systems, 24(4), 3868-3881.*