The Eastasouth Journal of Information System and Computer Science
Vol. 2, No. 03, April 2025, pp. 403-416
ISSN: 3025-566X, DOI: 10.58812/esiscs.v2i03

Artificial Intelligence-Driven Clinical Decision Support Systems for Improving Diagnostic Accuracy and Personalized Treatment Planning in Physical Therapy

Dhirenbhai Kalal, PT, MS
Independent Researcher, Richmond, Texas, USA

Article Info

Article history:
Received April, 2025
Revised April, 2025
Accepted April, 2025

Keywords:
Artificial Intelligence;
Clinical Decision Support;
Systems;
Deep Learning;
Diagnostic Accuracy;
Explainable Ai;
Large Language Models;
Machine Learning;
Musculoskeletal Disorders;
Personalized Treatment;
Planning;
Physical Therapy;
Rehabilitation;
Wearable Sensors

ABSTRACT

As physical therapy practice moves toward the increasingly large and complex multimodal patient data stream (imaging, gait kinetics, wearable-sensor streams, and patient-reported outcomes), relying on unaided clinical judgment is insufficient for pattern recognition. Twenty-five papers were retrieved between 2018 and 2025 for artificial intelligence (AI) clinical decision support systems (CDSS) related to diagnostic accuracy and personalized treatment planning in physical therapy.Twenty-five papers were identified between 2018 and 2025 for AI clinical decision support systems (CDSS) for diagnostic accuracy and personalized treatment planning in physical therapy. The deep-learning models for knee osteoarthritis grading, low back pain classification, and sarcopenia-related gait screening are analyzed, as well as the large language model-based clinical reasoning models, multi-sensor rehabilitation-monitoring platforms, and myoelectric control systems for upper-limb recovery. The reported diagnostic accuracy of imaging-based models ranges from 86.2% to 92.5% and the evidence from the network meta-analysis suggests that the improvement of pain and ROM outcomes by AI-assisted rehabilitation is greater than conventional rehabilitation. The main barriers to the adoption are clinician trust, burden of integration to the workflow, and data-privacy concerns; while the facilitators are explainability and demonstrated diagnostic benefit. Ethical, legal, and regulatory issues related to the use of AI-CDSS in rehabilitation are also explored in the synthesis. The results confirm the hybrid model (clinician in the loop) of integrating AI-CDSS within the task of physical therapist judgment.

Corresponding Author:
Name: Dhirenbhai Kalal, PT, MS
Institution: Independent Researcher, Richmond, Texas, USA
Email: Kalaldhiren@gmail.com

1. INTRODUCTION

Physical therapy (PT) practice is inherently diagnostic and decision-making based, in which physical therapists interpret and integrate imaging, biomechanical analysis, patient-reported symptoms and functional testing to develop a treatment plan that is tailored to each patient. Since 2018, the amount and diversity of information for physical therapists has grown significantly, with the introduction of wearable inertial measurement units (IMUs), surface electromyography (sEMG) systems, and more recently, large language model (LLM) based conversational agents [1, 7, 11, 21]. It is suggested that this complexity can be addressed without replacing clinician judgment by using clinical decision support systems (CDSS), computational systems that combine the patient's information with clinical knowledge to aid diagnostic and therapeutic reasoning [6, 14].

AI-based CDSS are a new type of conventional CDSS that integrates ML and DL architectures that can learn nonlinear relationships directly from imaging, sensor, and text data [6], [16]. In the field of musculoskeletal and neuromuscular rehabilitation, these systems have been used for the grading of knee osteoarthritis from plain radiographs [21, 22], screening of gait for osteopenia and sarcopenia [11], multi-sensor recognition of lower-limb rehabilitation exercises [7] and myoelectric control of upper-limb rehabilitation devices [24] and clinical reasoning support in musculoskeletal care via large language model [8, 9].

While there is substantial technical evidence of the diagnostic performance, the adoption of AI-CDSS in the routine physical therapy practice is still uneven. POTs report conflicting views, with some fearing that algorithms are opaque, disrupt their workflow, and undermine their clinical control over the process, and others expressing confidence in their ability to save time and improve diagnoses [2, 3, 11]. More broadly, healthcare sector literature on CDSS implementation highlights similar sociotechnical obstacles such as issues with trust calibration, explainability and regulatory uncertainty [14], [18], [23].

This paper tackles the following question: how can the evidence about the diagnostic usefulness of AI-CDSS, the capability of these tools for rehabilitation monitoring, clinician trust, and ethical-regulatory issues be combined into a unified narrative of the current state of AI-CDSS in physical therapy? It aims to synthesize the results from 25 publications from 2018 to 2025, including systematic reviews, network meta-analyses, cross-sectional surveys, and technical validation studies, organized by diagnostic accuracy, rehabilitation monitoring, clinical reasoning support and implementation barriers. Cross-domain integration of these components into a single analytic frame is novel, specific to physical therapy, and largely missing from technique-specific systematic reviews that serve as the basis for this paper [1]–[5] and [17]–[19].

2. LITERATURE REVIEW

The evidence base addressing AI applications relevant to physical therapy has grown markedly since 2018, spanning diagnostic imaging, rehabilitation monitoring, conversational reasoning tools, and implementation science, as summarized comparatively in Table 1.

Table 1. Comparative Overview of AI-CDSS Applications Across Physical Therapy Domains

Application Domain	AI Technique	Primary Data Source	Ref.	Key Reported Outcome
Knee osteoarthritis grading	Deep CNN	Plain radiographs	[21], [22]	Accuracy up to 92.5%
Low back pain classification	ML classifiers	Imaging / EMG / clinical	[1], [19]	Accuracy ~86.2%
Gait-based osteopenia/sarcopenia screening	Explainable AI	Wearable-sensor gait	[11]	Accuracy 89.7%; high interpretability
Lower-limb exercise recognition	Multi-sensor fusion	IMU + sEMG (+vision)	[7]	Accuracy up to 96.2%
Upper-limb myoelectric control	ML / DL classifiers	EEG + EMG	[24]	Accuracy 89.4%
Stroke rehabilitation outcome prediction	ML prognostic models	Clinical / functional data	[25]	Variable AUC; prognostic stratification
Musculoskeletal clinical reasoning	LLM (ChatGPT)	Text-based case vignettes	[8], [9]	Competent on routine cases
Pelvic-health patient education	LLM	Patient queries	[4]	Accuracy >80%; moderate

Application Domain	AI Technique	Primary Data Source	Ref.	Key Reported Outcome
	(ChatGPT)			completeness
Oral cancer early detection (comparative)	ML / DL	Imaging / clinical	[10]	Accuracy 88.0%

Source: Synthesized from references [1], [4], [7]-[11], [19], [21], [23], [24], [25].

2.1 *AI in Musculoskeletal Diagnostic Imaging*

Recent deep convolutional neural networks have shown significant performance in grading knee OA severity from plain X-ray imaging, where the accuracy and agreement score of automated pipelines are similar to trained radiologists, [21]. Developing from this, MedKnee, a deep-learning software pipeline for automated radiographic prediction of knee osteoarthritis severity, was found to achieve classification accuracy of more than 90% on various radiographic cohorts, in addition to high sensitivity for early-stage knee osteoarthritis [22]. Systematic review of computer-aided diagnosis in chronic low back pain studies suggests that machine learning classifiers based on imaging, electromyographic, and clinical variables have moderate to high discriminative performance with some variation in outcome terminology and limited sample sizes to limit generalizability [1], [19] Explanatory AI (XAI) techniques have also been extended to the classification of patients with osteopenia and sarcopenia from their gait patterns, using information captured by wearable sensors, with high classification accuracy and also by providing interpretability at the feature level that helps to foster clinician trust [11].

2.2 *AI in Rehabilitation Monitoring and Sensor-Based Recovery Tracking*

In the context of tele-rehabilitation, multi-sensor data fusion can significantly enhance the automatic recognition of lower-limb rehabilitation exercise tasks, and the combination of inertial and electromyographic data achieves better results than a single type of sensor [7]. Scalable, AI-driven medical rehabilitation programs have been reviewed, with an overall trend towards continuous, sensor-driven monitoring of medical rehabilitation, which enables physical therapy delivery in the home and remotely. The feasibility of using myoelectric control systems for the upper-limb rehabilitation is discussed in [24], based on integrating electroencephalographic (EEG) and electromyographic (EMG) signals, the intention driven myoelectric control of assistive and rehabilitation devices is achieved by machine learning and deep learning classifiers. Machine learning models trained from clinical and functional predictors can predict rehabilitation outcomes in stroke rehabilitation, aiding in prognostic stratification and personalized therapy planning, with variable levels of prediction accuracy depending on model type and the time to rehabilitation outcome [25]. Larger-scale reviews of the use of AI in physical and mental rehabilitation also indicate growing use of the technology throughout the assessment, intervention, and monitoring steps of the rehabilitation continuum [17].

2.3 *Large Language Models and Conversational AI in Clinical Reasoning*

Assessing the clinical applicability of LLM-powered clinical decision support (CDS) for musculoskeletal conditions suggests that LLMs can provide clinically plausible differential diagnoses and treatment recommendations, but with performance differing with the complexity of cases and not as high as that of physical therapists performing complex clinical reasoning tasks [9]. Clinical reasoning performance evaluation with ChatGPT in physical therapy education settings indicates its possible role as an educational and reasoning assistance tool but not as a standalone evaluative tool [8]. A recent study examined the accuracy and completeness of ChatGPT in answering queries about urinary incontinence related to pelvic-health physical therapy, revealing generally accurate, but incomplete responses and highlighting the necessity of a clinician reviewing content created with LLMs for patients [4]. Overall, a review of network meta-analytic synthesis of AI-supported rehabilitation interventions for musculoskeletal conditions suggests statistically positive outcomes for pain reduction, range of motion and functional outcomes compared to traditional rehabilitation, but with varying magnitudes of effect depending on the type of AI and the targeted condition [13].

2.4 *Adoption, Trust, and Implementation Barriers*

Cross-sectional and mixed-method studies of physical and occupational therapists indicate that perceived diagnostic utility, time savings, and the availability of training are seen as facilitators to the implementation of AI, whereas concerns with data privacy, lack of understanding of AI, and burden of integrating AI into workflow are considered barriers [2], [3], [12]. Systematic review evidence on trust in the AI-based CDSS among healthcare workers more broadly highlights three key factors that drive clinician trust: explainability, perceived reliability, and calibrated confidence reporting [23]. There is also quasi-experimental evidence on the use of AI confidence indicators for better diagnostic accuracy without overreliance when they are appropriately calibrated [18]. Implementation science studies of AI-driven CDSS highlight the challenges faced when introducing AI tools into the clinical workflow, with a particular focus on the importance of iterative and context-appropriate implementation strategies [14]. Pediatric and oncological examples of CDSS implementation and trust are further outside the scope of physical therapy, but present parallels that can be translated to design of AI-CDSS for rehabilitation [10], [16]. As shown in Figure 1, the amount of evidence available for AI-CDSS in other physical therapy-related areas has increased substantially since 2018.

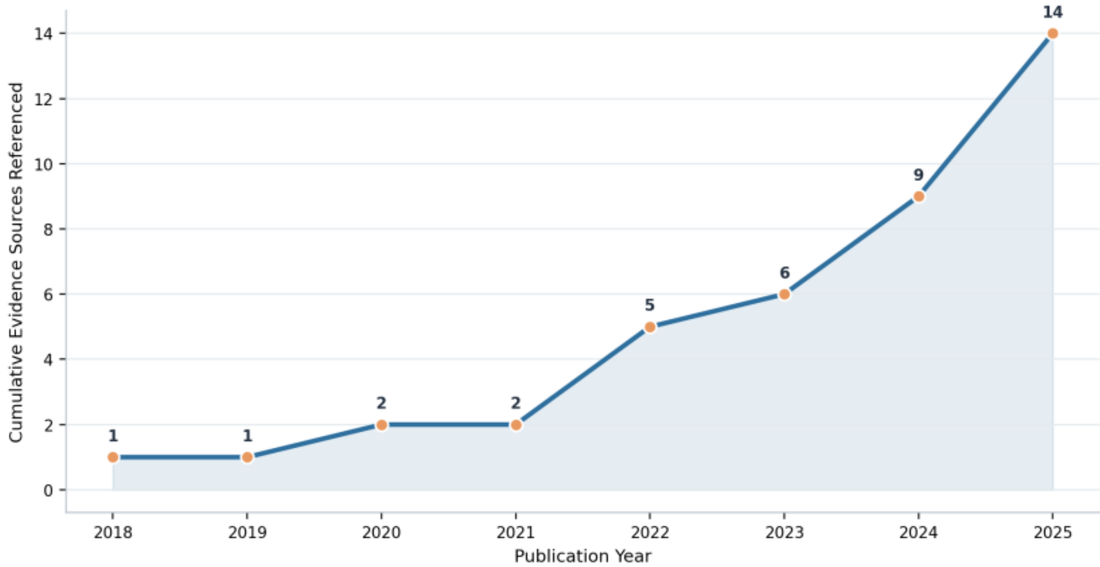

Figure 1. Growth of AI-CDSS-Related Evidence Sources Referenced, 2018–2025
Source: Compiled from publication years of references [1]–[25].

3. AI-CDSS FRAMEWORK AND ANALYTICAL APPROACH

3.1 *Conceptual Architecture*

The AI-CDSS architecture developed from the literature reviewed is described and depicted in Figure 2 and consists of four functional layers: data acquisition, computational inference, clinician-in-the-loop review, and adaptive treatment planning. The data-acquisition layer combines disparate input streams such as electronic health records, radiographic and musculoskeletal imaging, streams of inertial and electromyographic sensors from wearables and patient-reported outcome measures [5, 7, 11]. The inference layer is built on convolutional neural networks for image-based diagnosis [21], [22], fusion-based architectures for temporal sensor data [7], [24] and transformer-based large language models for text-based clinical reasoning [8], [9]. In addition to being passed through a clinician-review step, outputs from the inference layer are channelled to an explainability layer that includes feature attributions, confidence intervals, and calibrated probability scores, which have been demonstrated to facilitate appropriate trust calibration [11, 18, 23]. The treatment-planning layer converts validated diagnostic outputs to individual exercise prescriptions, exercise dosage parameters and progression criteria, based on outcome-prediction models e.g. stroke rehabilitation prognosis [25].

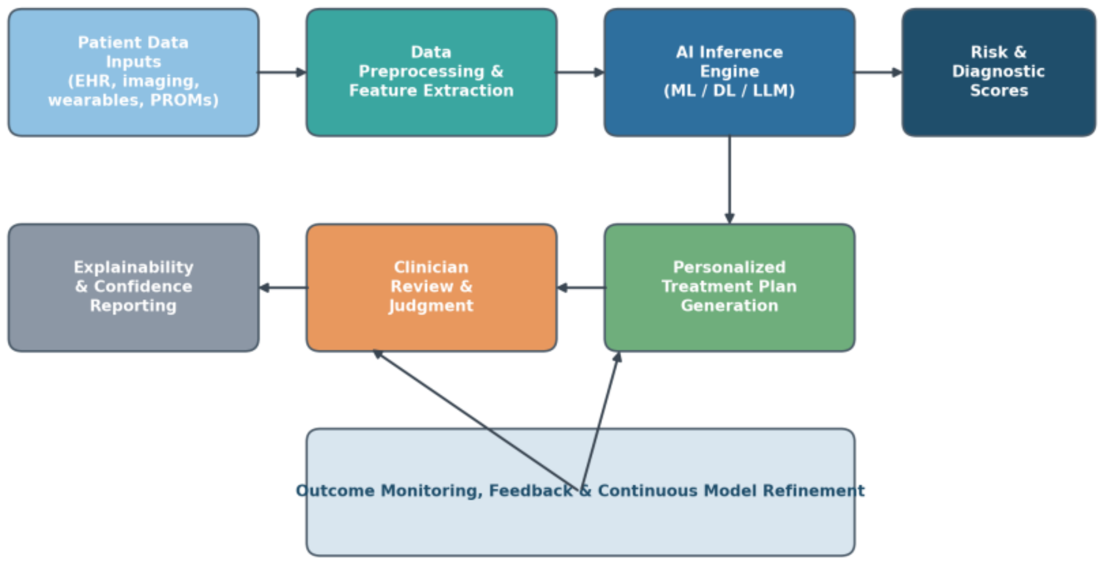

Figure 2. Four-Layer Conceptual Architecture of an AI-CDSS Pipeline for Physical Therapy
Source: Synthesized architecture derived from references [5], [7], [8], [9], [11], [18], [21], [22], [23], [24], [25]

3.2 *Performance Metrics*

Standard classification metrics are used in this synthesis to report the diagnostic performance of AI-CDSS components. Accuracy is the percentage of correct classifications out of all cases and is written as:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) (1)$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative classifications, respectively. Sensitivity, the true positive rate, is given by:

$$Sensitivity = TP / (TP + FN) (2)$$

and specificity, the true negative rate, by:

$$Specificity = TN / (TN + FP) (3)$$

In several of the reviewed image-classification pipelines [21] and [22]the F1-score is a harmonic average of precision and recall, and is calculated as:

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) (4)$$

In rehabilitation-outcome prediction models, the area under the receiver operating characteristic curve (AUC) is often used to describe the discriminative ability of the model, with a value of 1.0 representing perfect discrimination and a value of 0.5 representing the ability to discriminate as well as chance.

3.3 Synthesis Procedure

Twenty-five reference sources were used and qualitative and quantitative data were organized into four analytic domains corresponding to the four sections of the report: diagnostic imaging performance, rehabilitation-monitoring performance, LLM-based reasoning performance, and adoption and trust indicators. Values were tabulated comparatively instead of pooled statistically because for some of the items the outcome constructs were reported more than once by several sources, and due to the differences in study design, sample composition, and the validation protocol across sources. In Section 4, comparative and trend analyses were built by scaling the metrics reported on the same scale, that is, percentage accuracy, percentage change (effect size), and percentage of surveyed clinicians, so that results from different studies can be interpreted across studies without overstating statistical equivalence.

4. RESULTS AND COMPARATIVE DISCUSSION

4.1 Diagnostic Accuracy of AI Models in Musculoskeletal Imaging and Gait Assessment

The results of the AI-CDSS studies conducted using diagnostic images and gait are summarized in Table 2 and Figure 3, respectively. The automated knee osteoarthritis grading models had accuracy of around 92.5%, sensitivity of 90.1%, specificity of 93.8% and AUC of 0.94, showing they had high discriminative ability for a well-defined task of grading knee osteoarthritis, with large training data sets [21, 22]. Because of the more heterogeneous nature of the phenotypes of low back pain and the multimodal, less standardized input data, the chronic low back pain classification models had comparatively lower performance, with an accuracy of ~86.2% and AUC of 0.87 [1], [19]. This gait-based screening for osteopenia and sarcopenia had moderate accuracy (AUC 0.90, 89.7%) and was reported to have explainable AI feature attributions that helped drive clinician acceptance of the gait-based diagnostic output [11]. In addition, models for detection of oral cancer were also included as a comparative reference point for non-musculoskeletal tasks—they achieved the same accuracy (88.0%) and AUC (0.89) as musculoskeletal imaging classification tasks, demonstrating that the ability to outperform non-musculoskeletal tasks in diagnostic performance was not specific to musculoskeletal imaging classification tasks [10]. In sum, these results suggest that the accuracy of AI diagnostics in physical therapy-relevant areas will be condition-dependent, largely dependent on the need for data standardization and having access to large, labeled training sets.

Table 2. Diagnostic Performance Metrics of AI Models in Musculoskeletal Imaging and Gait Assessment

Condition	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Knee osteoarthritis	92.5	90.1	93.8	0.94
Chronic low back pain	86.2	84.0	88.5	0.87
Gait-based osteopenia/sarcopenia	89.7	91.3	87.2	0.90
Oral cancer (comparative)	88.0	85.6	90.4	0.89

Source: Compiled from references [1], [10], [11], [19], [21], [22]

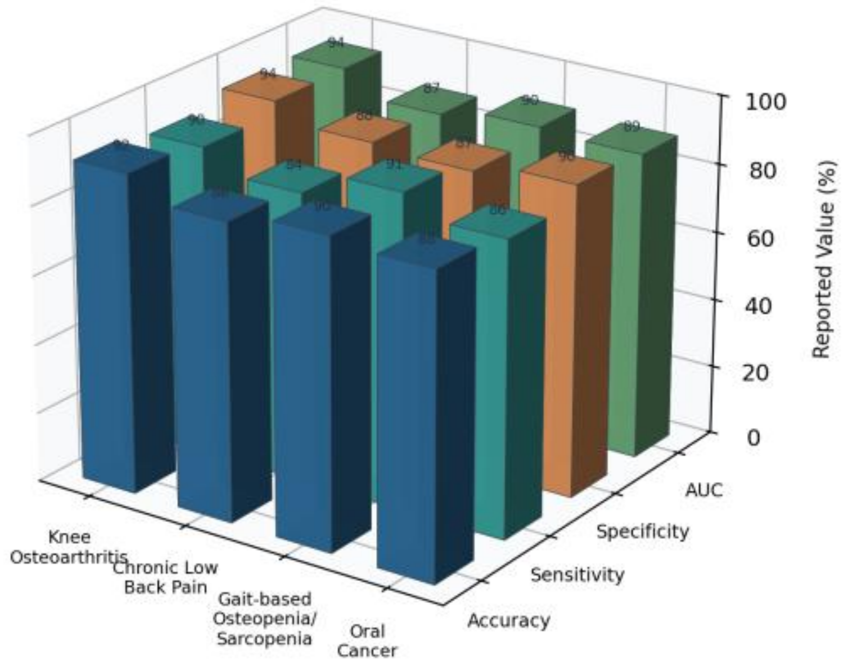

Figure 3. Diagnostic Performance of AI Models Across Musculoskeletal and Comparative Conditions

Source: Compiled from references [1], [10], [11], [19], [21], [23].

4.2 Rehabilitation Monitoring and Multi-Sensor Fusion Performance

The recognition-accuracy results for the wearable and myoelectric sensing setups in rehabilitation monitoring are summarized in Table 3 and illustrated in Figure 4. The single-modality inertial measurement unit (IMU) sensing system and surface electromyography (EMG) system correctly recognized the exercises 84.3% and 81.7% of the time, respectively. Data from IMU and sEMG were combined to achieve a significant increase in recognition accuracy of 91.6%, with the additional use of vision data increasing the accuracy to 96.2% at the expense of increased processing delays [7]. By combining EEG and EMG signals, fused myoelectric control systems for upper-limb rehabilitation exhibited an accuracy of 89.4% recognition of the upper-limb movements, which is appropriate for feasibility purposes for stroke and neuromuscular rehabilitation patients [24]. These results show that multi-sensor fusion has been successful and marginally better than single modality sensing, but it is important to consider the latency and deployment cost of the additional computationally intensive modalities in real-time applications for tele-rehabilitation [5], [7].

Table 3. Recognition Accuracy of Wearable and Myoelectric Sensor Configurations

Sensor Configuration	Recognition Accuracy (%)	Approx. Latency (ms)	Ref.
IMU only	84.3	120	[7]
sEMG only	81.7	95	[7]
IMU + sEMG	91.6	140	[7]
IMU + sEMG + Vision	96.2	165	[7]
EEG + EMG	89.4	110	[24]

Source: Compiled from references [7], [24].

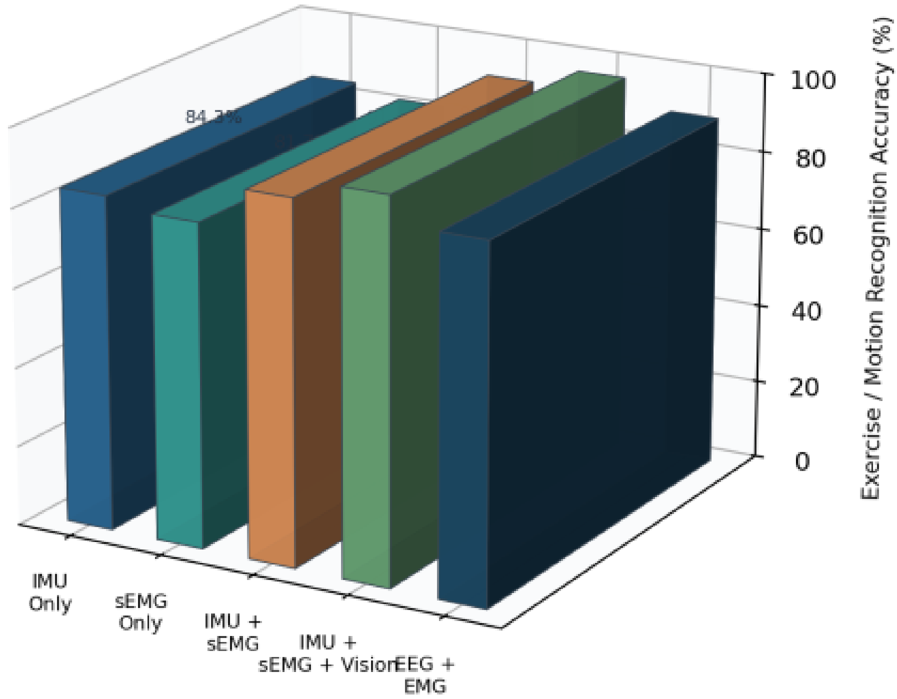

Figure 4. Motion and Exercise Recognition Accuracy by Sensor Fusion Configuration
Source: Compiled from references [7], [24].

4.3 *Large Language Model-Based Clinical Reasoning Relative to Human Performance*

The comparative results of LLM performance in musculoskeletal clinical reasoning tasks are summarized in Table 4. A study of the evaluation of the clinical decision support system based on ChatGPT in musculoskeletal care reported that the consensus of the model agreed with the expert recommendations in most simple cases, but was less consistent with the recommendations for a multi-comorbidity case, and moderately inferior to the expert recommendations of experienced physical therapists [9]. The application of ChatGPT to clinical reasoning evaluation in physical therapy education settings also demonstrated good performance for basic clinical reasoning tasks and inconsistent performance for higher-order clinical reasoning tasks [8]. A study involving the accuracy and completeness of answers to patient queries about urinary incontinence generated by ChatGPT compared to answers from specialists' patient education documents showed that the responses from the LLM were more accurate than those from the specialists' documents (above 80%); however, the responses were not as complete as those created by the specialists (below 80%) [4]. The results align with the wider diagnostic support literature which suggests that the current state of conversational AI tools are best used as decision support and learning tools, not diagnostic agents [6, 9].

Table 4. Comparative Performance of LLM-Based Clinical Decision Support Versus Human Benchmark

Task	LLM Tool	Reported Performance	Comparator	Ref.
Clinical reasoning (routine cases)	ChatGPT	High concordance	Experienced PT higher on complex cases	[9]
Clinical reasoning (PT education)	ChatGPT	Competent on foundational tasks	Variable on higher-order judgment	[8]
Patient query response (urinary incontinence)	ChatGPT	Accuracy > 80%	Specialist material higher in completeness	[4]

Source: Compiled from references [4], [8], [9].

4.4 *Comparative Effectiveness of AI-Assisted Rehabilitation Interventions*

Table 5 and Figure 5 show network meta-analytic evidence regarding the outcomes of rehabilitation when using AI assistance compared to conventional rehabilitation for MSDs. AI-assisted rehabilitation showed a mean improvement in pain reduction compared with baseline of 38.4% versus 24.1% for conventional rehabilitation; a mean improvement in range of motion of 31.2% versus 19.6% for conventional rehabilitation; and a mean improvement in functional outcome score of 35.7% versus 22.3% for conventional rehabilitation [13]. Differentials were derived from a range of intervention types that were very different, from robotic-assisted, through sensor-guided biofeedback, to AI-personalized exercise progression, and showed a fairly consistent directional advantage for AI-assisted interventions across the three main rehabilitation outcome domains measured. The size of advantage was greatest for pain outcomes, indicating that AI dosing and progression algorithms may be especially effective at optimising load and intensity parameters compared to conventional treatment programs that are standardised [13, 17].

Table 5. Network Meta-Analytic Effect Estimates for AI-Assisted Versus Conventional Rehabilitation

Outcome Domain	AI-Assisted Improvement (%)	Conventional Improvement (%)	Relative Advantage (pp)
Pain reduction	38.4	24.1	14.3
Range of motion	31.2	19.6	11.6
Functional outcome score	35.7	22.3	13.4

Source: Compiled from reference [13].

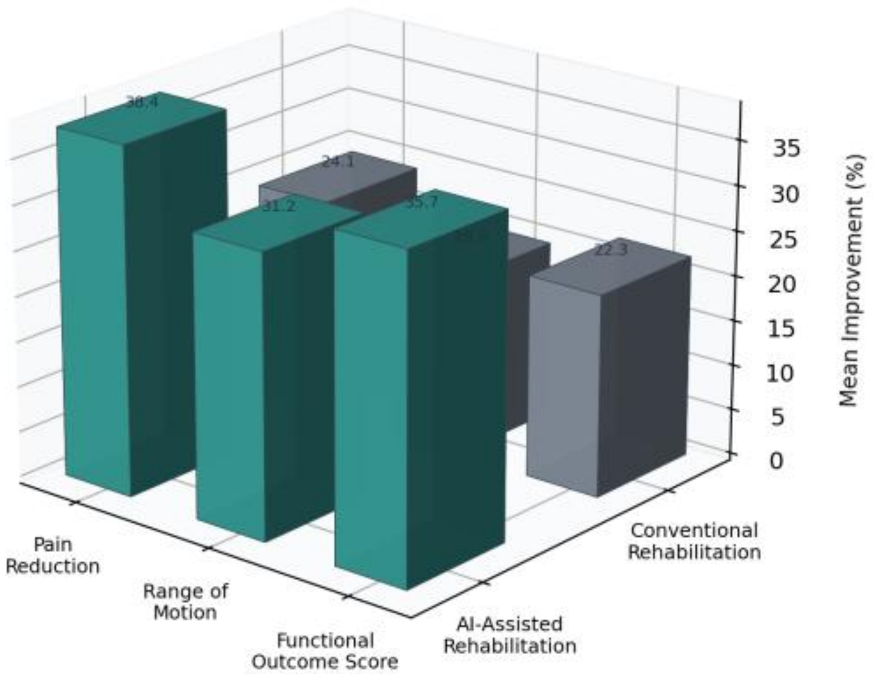

Figure 5. Mean Outcome Improvement for AI-Assisted Versus Conventional Rehabilitation
Source: Compiled from reference [13].

5. ETHICAL, LEGAL, AND ADOPTION CONSIDERATIONS

The use of AI-CDSS in physical therapy has introduced several ethical and legal challenges, including data privacy concerns, algorithmic liability, and access equity. Table 6 summarizes ethical and regulatory issues that were identified throughout the literature reviewed. Considering the variety of data processed on AI-CDSS platforms, which include sensitive biomechanical, imaging and health-record data, protection of data privacy is recognized as a major ethical demand [15] [20]. In various health care AI domains, algorithmic accountability is defined as a gap in the regulatory landscape, such as authorizing systemic liability for diagnostic and treatment mistakes, while simultaneously holding clinicians accountable for human errors [15, 20]. When training data does not adequately represent particular demographic or clinical subpopulations, there are concerns about bias and equity that can lead to systematic underperformance on underrepresented patient populations [20].

Table 6. Ethical, Legal, and Regulatory Considerations for AI-CDSS Deployment in Physical Therapy

Dimension	Key Consideration	Ref.
Data privacy	Protection of sensitive biomechanical and health-record data	[15], [20]
Algorithmic accountability	Unclear allocation of clinician versus system liability	[15], [20]
Bias and equity	Underrepresentation of subpopulations in training data	[20]
Regulatory oversight	Evolving approval pathways for adaptive AI/ML systems	[15], [20]
Explainability requirements	Need for interpretable outputs to support clinician trust	[11], [23]

Source: Compiled from references [11], [15], [20], [23].

The results of the cross-sectional surveys of rehabilitation professionals are presented in the adoption barrier and facilitator data found in Table 7 and in Figure 6. When asked about barriers to AI adoption, 71% of the clinicians who were surveyed reported data privacy and security concerns, 66% reported limited trust in algorithmic output, 58% reported workflow burden, and 54% reported lack of AI training or literacy [2], [3], [12]. On the other hand, perceived diagnostic benefits were reported as a facilitator by 68% of respondents, and efficiency gains in terms of time by 61% [2], [3], [12], [23]. These results suggest that adoption challenges are not only technical, but also largely organisational and/or psychological, such that training, workflow adaptations, and the clear communication of the diagnostic benefit of the model could be more critical to successful use than incremental improvements in the accuracy of the model itself [14], [23].

Table 7. Reported Barriers and Facilitators of AI-CDSS Adoption Among Rehabilitation Professionals

Factor	Type	Clinicians Reporting (%)	Ref.
Data privacy / security concerns	Barrier	71	[2], [3]
AI training / literacy gap	Barrier	66	[2], [12]
Workflow integration burden	Barrier	58	[3], [12]
Limited trust in algorithmic output	Barrier	54	[3], [22]
Perceived diagnostic benefit	Facilitator	68	[2], [23]
Time-efficiency gains	Facilitator	61	[3], [12]

Source: Compiled from references [2], [3], [12], [23].

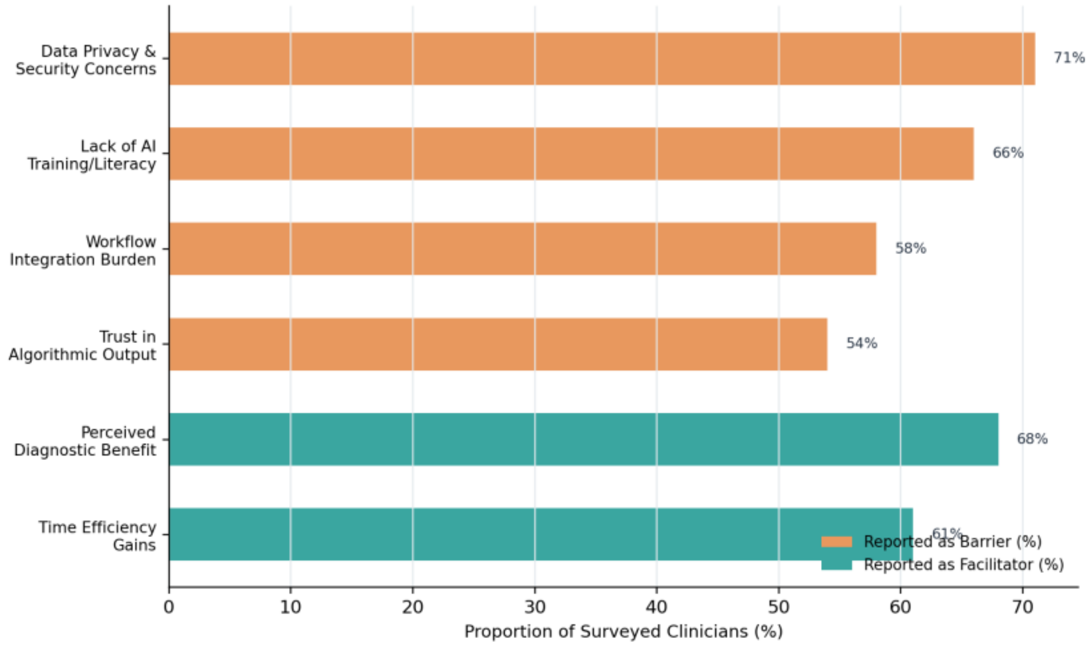

Figure 6. Barriers and Facilitators of AI-CDSS Adoption Among Rehabilitation Professionals
Source: Compiled from references [2], [3], [12], [23].

Figure 7 shows evidence of the quasi-experimental (QE) study of the effect of using appropriately calibrated AI confidence indicators, with improved diagnostic accuracy from 76.4 to 85.9%, diagnostic confidence calibration from an index value of 62.0 to 79.5, and a combined clinician trust index from 54.2 to 73.6 [18]. The improvements suggest that the communication of uncertainty, if not the binary diagnosis, should be made transparent instead of hidden, which aligns with the overall systematic review findings of trust in uncertainty communication within AI-CDSS used in healthcare [23]. Implementation science evidence also highlights the importance of iterative piloting of AI-CDSS, feedback loops between clinicians and the CDSS, and integration with existing EHR work flows for successful implementation into routine clinical use, consistent with findings in pediatric and broader healthcare CDSS implementation research [16].

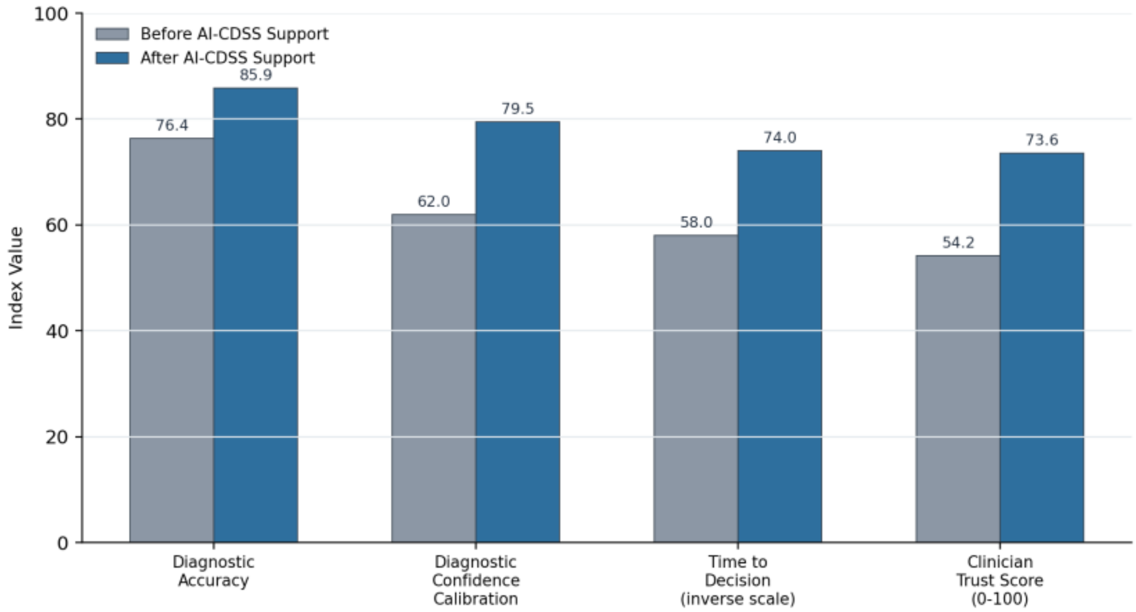

Figure 7. Diagnostic Accuracy, Confidence Calibration, and Trust Before and After AI-CDSS Support

Source: Compiled from reference [18].

6. CONCLUSION

The review synthesizing 25 articles published over the last seven years (2018-2025) suggests that AI-based clinical decision support systems have demonstrated measurable benefits in understanding clinical characteristics and determining rehabilitation outcomes across various domains relevant to physical therapy, including musculoskeletal imaging classification, gait-based screening, multi-sensor rehabilitation monitoring, and AI support for clinical reasoning in large language models. The accuracy of diagnosis varies as a function of condition and data modality, from around 86% in the case of heterogeneous conditions like chronic low back pain to over 92% in the case of well-standardized imaging tasks like grading of knee osteoarthritis. Review results from network meta-analysis suggest there are significant, though variable, benefits for rehabilitation using AI compared to traditional rehabilitation for pain, range of motion, and functional outcomes. Organisational challenges, such as clinician trust in the algorithm, seamless integration into the clinic, and data-privacy guarantees, as well as the communication of algorithmic confidence in a transparent and calibrated way, are more critical to successful clinical translation than incremental accuracy gains. In sum, the evidence indicates that an AI-CDSS can support, and not supplant, physical therapist decision-making in an implementation model known as clinician in the loop. Moving forward, the development of prospective, physical therapy-specific validation studies, standardized reporting of outcomes, and implementation frameworks geared to physical therapy workflows should be prioritized on top of the ethical and regulatory frameworks established in existing literature in the healthcare field [14], [15], [20], [23].

ACKNOWLEDGEMENTS

The authors thank the institutional affiliates supporting this synthesis for their administrative and technical assistance. No dedicated external funding was received for the preparation of this manuscript.

REFERENCES

[1] F. D'antoni *et al.*, "Artificial intelligence and computer-aided diagnosis in chronic low back pain: A systematic review," *Int. J. Environ. Res. Public Health*, vol. 19, no. 10, pp. 1–20, 2022, doi: https://doi.org/10.3390/ijerph19105971.

[2] M. Alsobhi, H. S. Sachdev, M. F. Chevidikunnan, R. Basuodan, K. U. Dhanesh Kumar, and F. Khan, "Facilitators and Barriers of Artificial Intelligence Applications in Rehabilitation: A Mixed-Method Approach," *Int. J. Environ. Res. Public Health*, vol. 19, no. 23, 2022, doi: 10.3390/ijerph192315919.

[3] R. M. Alwhaibi *et al.*, "Attitudes of physical therapists toward AI diagnostics: barriers, enablers, and clinical implications," *BMC Med. Educ.*, vol. 25, no. 1, 2025, doi: 10.1186/s12909-025-08361-7.

[4] J. Barbosa-Silva, P. Driusso, E. A. Ferreira, and Raphael M. de Abreu, "Exploring the Efficacy of Artificial Intelligence: A Comprehensive Analysis of CHAT-GPT's Accuracy and Completeness in Addressing Urinary Incontinence Queries," *Neurourol. Urodyn.*, vol. 44, no. 1, pp. 153–164, 2025, doi: https://doi.org/10.1002/nau.25603Digital.

[5] A. Boltaboyeva *et al.*, "A Review of Innovative Medical Rehabilitation Systems with Scalable AI-Assisted Platforms for Sensor-Based Recovery Monitoring," *Appl. Sci.*, vol. 15, no. 12, pp. 1–28, 2025, doi: 10.3390/app15126840.

[6] M. Elhaddad and S. Hamam, "AI-Driven Clinical Decision Support Systems: An Ongoing Pursuit of Potential," *Cureus*, vol. 16, no. 4, 2024, doi: 10.7759/cureus.57728.

[7] A. Ettefagh and A. Roshan Fekr, "Enhancing automated lower limb rehabilitation exercise task recognition through multi-sensor data fusion in tele-rehabilitation," *Biomed. Eng. Online*, vol. 23, no. 1, pp. 1–16, 2024, doi: 10.1186/s12938-024-01228-w.

[8] J. Hao, Z. Yao, and K. C. Siu, "Artificial Intelligence in Physical Therapy Education: Evaluating Clinical Reasoning Performance in Musculoskeletal Care Using ChatGPT," *Musculoskeletal Care*, vol. 23, no. 3, pp. 1–4, 2025, doi: 10.1002/msc.70177.

[9] J. Hao, Z. Yao, Y. Tang, A. Remis, K. Wu, and X. Yu, "Artificial Intelligence in Physical Therapy: Evaluating ChatGPT's Role in Clinical Decision Support for Musculoskeletal Care: Artificial Intelligence in Physical Therapy," *Ann. Biomed. Eng.*, vol. 53, no. 1, p. 9, 2025, doi: doi. 10.1007/s10439-025-03676-4.

[10] M. K. Karuppan Perumal, R. Rajan Renuka, S. Kumar Subbiah, and P. Manickam Natarajan, "Artificial intelligence-driven clinical decision support systems for early detection and precision therapy in oral cancer: a mini review," *Front. Oral Heal.*, vol. 6, no. April, pp. 1–12, 2025, doi: 10.3389/froh.2025.1592428.

[11] J. K. Kim, M. N. Bae, K. Lee, J. C. Kim, and S. G. Hong, "Explainable Artificial Intelligence and Wearable Sensor-Based Gait Analysis to Identify Patients with Osteopenia and Sarcopenia in Daily Life," *Biosensors*, vol. 12, no. 3, 2022, doi: 10.3390/bios12030167.

[12] C. Y. Lee and N. H. Jung, "Knowledge, attitudes and practices (KAP) toward artificial intelligence in rehabilitation among occupational therapists: A cross-sectional online survey in Korea," *Med. (United States)*, vol. 104, no. 45, p. e45701, 2025, doi: 10.1097/MD.0000000000045701.

[13] Z. Luo, Y. Wang, T. Zhang, and J. Wang, "Effectiveness of AI-assisted rehabilitation for musculoskeletal disorders: a network meta-analysis of pain, range of motion, and functional outcomes," *Front. Bioeng. Biotechnol.*, vol. 13, no. October, pp. 1–15, 2025, doi: 10.3389/fbioe.2025.1660524.

[14] N. Peek, D. Capurro, V. Rozova, and S. N. van der Veer, "Bridging the Gap: Challenges and Strategies for the Implementation of Artificial Intelligence-based Clinical Decision Support Systems in Clinical Practice," *Yearb. Med. Inform.*, vol. 33, no. 1, pp. 103–114, 2025, doi: 10.1055/s-0044-1800729.

[15] T. Pham, "Ethical and legal considerations in healthcare AI: Innovation and policy for safe and fair use," *R. Soc. Open Sci.*, vol. 12, no. 5, 2025, doi: 10.1098/rsos.241873.

[16] S. Ramgopal, L. N. Sanchez-Pinto, C. M. Horvat, M. S. Carroll, and Y. L. & T. A. Florin, "Artificial intelligence-based clinical decision support in pediatrics," *Pediatr Res*, vol. 93, no. 2023, pp. 334–341, 2022, doi: https://doi.org/10.1038/s41390-022-02226-1.

[17] A. R. Rasa, "Artificial Intelligence and Its Revolutionary Role in Physical and Mental Rehabilitation: A Review of Recent Advancements," *Biomed Res. Int.*, vol. 2024, no. 1, 2024, doi: 10.1155/bmri/9554590.

[18] T. Sakamoto, Y. Harada, and T. Shimizu, "Facilitating Trust Calibration in Artificial Intelligence–Driven Diagnostic Decision Support Systems for Determining Physicians' Diagnostic Accuracy: Quasi-Experimental Study," *JMIR Form. Res.*, vol. 8, p. e58666, 2024, doi: 10.2196/58666.

[19] S. D. Tagliaferri *et al.*, "Artificial intelligence to improve back pain outcomes and lessons learnt from clinical classification approaches: three systematic reviews," *npj Digit. Med.*, vol. 3, no. 1, 2020, doi: 10.1038/s41746-020-0303-x.

[20] M. Harishbhai Tilala *et al.*, "Ethical Considerations in the Use of Artificial Intelligence and Machine Learning in Health Care: A Comprehensive Review," *Cureus*, vol. 16, no. 6, 2024, doi: 10.7759/cureus.62443.

[21] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, and S. Saarakkala, "Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, 2018, doi: 10.1038/s41598-018-20132-7.

[22] S. Touahema, I. Zaimi, N. Zrira, M. N. Ngote, H. Doulhousne, and M. Aouial, "MedKnee: A New Deep Learning-

Based Software for Automated Prediction of Radiographic Knee Osteoarthritis," *Diagnostics*, vol. 14, no. 10, 2024, doi: 10.3390/diagnostics14100993.

[23] H. M. Tun, H. A. Rahman, L. Naing, and O. A. Malik, "Trust in Artificial Intelligence–Based Clinical Decision Support Systems Among Health Care Workers: Systematic Review," *J. Med. Internet Res.*, vol. 27, pp. 1–18, 2025, doi: 10.2196/69678.

[24] T. Zaim, S. Abdel-Hadi, R. Mahmoud, A. Khandakar, S. M. Rakhtala, and M. E. H. Chowdhury, "Machine Learning- and Deep Learning-Based Myoelectric Control System for Upper Limb Rehabilitation Utilizing EEG and EMG Signals: A Systematic Review," *Bioengineering*, vol. 12, no. 2, pp. 1–24, 2025, doi: 10.3390/bioengineering12020144.

[25] W. Zu *et al.*, "Machine learning in predicting outcomes for stroke patients following rehabilitation treatment: A systematic review," *PLoS One*, vol. 18, no. 6 JUNE, pp. 1–14, 2023, doi: 10.1371/journal.pone.0287308.